

The Encyclopedia Made Machine: Artificial Intelligence and the Impossibility of Truth in Badiou's Philosophy

Morteza Hosseinzadeh¹  | Tawid Teymouri² 

1. Assistant Professor of Philosophy Department, University of Zanjan, Zanjan, Iran. E-mail: mhoseinzadeh@znu.ac.ir

2. Corresponding Author, Assistant Professor of English Language and Literature, University of Zanjan, Zanjan, Iran. E-mail: tohidteymouri@znu.ac.ir

Article Info

Article type:

Research Article

Article history:

Received 08 June 2026

Received in revised form
08 September 2026

Accepted 11 September
2026

Published online 25
October 2026

Keywords:

Badiou, Artificial
Intelligence, Truth, Event,
Situation, Encyclopedia,
Ontology, Subject

ABSTRACT

This paper reads the structural limits of Artificial Intelligence through the ontological and evental philosophy of Alain Badiou. It argues that AI systems including large language models, machine learning architectures, and data-driven algorithms operate exclusively within the domain of the situation, that is, the structured, countable, encyclopedic order of already-established knowledge. For this reason, AI cannot produce truth in Badiou's sense: truth emerges only through a subject's fidelity to an event that ruptures a situation from the point of its constitutive void. Through Badiou's mathematical ontology, his logic of worlds, and his doctrine of truth procedures, I describe AI as an anti-generic machine: a system oriented toward discernibility, classification, prediction, and predicate-satisfaction rather than toward the indiscernible genericity of truth. It also examines AI in relation to Badiou's four conditions of philosophy — science, art, politics, and love — and argues that AI can imitate outputs associated with these procedures without becoming the subject of a truth-procedure. This paper situates Badiou's framework within contemporary philosophy of AI and argues that the irreducibility of truth to knowledge is at once the irreducibility of the subject to any computational architecture.

Cite this article: Hosseinzadeh, M. & Teymouri, T. (2026). The Encyclopedia Made Machine: Artificial Intelligence and the Impossibility of Truth in Badiou's Philosophy. *Journal of Philosophical Investigations*, 20(56), 375-394. <https://doi.org/10.22034/jpiut.2026.72909.4551>



© The Author(s).

Publisher: University of Tabriz.

Intruduction

Contemporary debates about artificial intelligence often swing between two dominant positions. Some proponents of artificial general intelligence believe that AI is approaching, or may eventually surpass, human intelligence, attributing to it capacities such as creativity, discovery, and perhaps even consciousness. Skeptics, by contrast, argue that contemporary AI remains a sophisticated form of statistical pattern-recognition rather than genuine understanding. Malabou's genealogy of intelligence clarifies the insufficiency of this opposition between AGI enthusiasm and skeptical reductionism. Long before artificial intelligence, the scientific concept of intelligence had already produced a philosophical crisis because it attempted to transform intelligence into something measurable, testable, and operational. As [Malabou \(2019, 1\)](#) observes, the emergence of intelligence as a psychological and scientific concept brought with it "an entire set of vocabulary for measurement, scales, and tests," challenging the universality traditionally attributed to reason. The problem is therefore not simply whether AI can become intelligent, but whether intelligence itself, once reduced to measurable performance, has already been transferred from the domain of truth to that of calculable knowledge.

Badiou's philosophy allows a different formulation, though not a simple middle ground between optimism and skepticism. AI may be extraordinarily powerful within the order of knowledge, and this power should not be minimized. The question, however, is whether the expansion of knowledge can ever become the production of truth. For Badiou, the answer depends on a distinction that contemporary AI discourse often obscures: knowledge and truth do not belong to the same order. The distinction is not terminological. Knowledge belongs to the structured regime of the situation and its encyclopedia, while truth is an entirely different ontological category, one that emerges through a subject's militant fidelity to an event that ruptures the situation from the point of its constitutive void ([Badiou, 2005, 327](#)). The implication for AI is structural rather than comparative: computational power does not by itself supply the event, intervention, or fidelity required by a truth-procedure. Žižek's critique of Heidegger is useful here because it identifies a temptation that also appears in contemporary AI discourse: the temptation "to forget this gap and endeavor to impose an ontic order that would be adequate to the ontological truth" ([Žižek, 2006, 274](#)). In the case of AI, this temptation appears whenever computational systems are treated not merely as tools within knowledge, but as if they could embody judgment, truth, or subjectivity itself.

The paper begins with Badiou's concepts of situation, state, and encyclopedia in order to show why AI should be understood less as artificial subjectivity than as the automation of established knowledge. Yet this claim requires qualification. AI does not merely store knowledge; it reorganizes, predicts, and operationalizes it at scales unavailable to ordinary human cognition. Pasquinelli's labour theory of machine intelligence is important here because

it gives this formalization a material history as AI captures not biological intelligence as such, but “the intelligence of labour and social relations” (Badiou, 2023, 12). The issue, then, is not whether AI is powerful, creative, or socially transformative, but how that power reorganizes the encyclopedia and where it encounters its structural limits. The paper further explores forcing and genericity, presenting AI as an anti-generic machine, and develops the practical and critical implications of this framework, including AI’s relation to Badiou’s four truth procedures and the danger of simulacra.

1. AI as the Automation of the Encyclopedia: Badiou and the Logic of Constructible Knowledge

In *Being and Event*, Badiou equates ontology with mathematics, particularly Zermelo–Fraenkel set theory. Being qua being is pure multiplicity without qualitative determination. Mathematics is the discourse that thinks this multiplicity as such (Badiou, 2005). The fundamental operation of any ontology is the count-as-one, the operation by which inconsistent multiplicity is structured into a consistent presentation (Badiou, 2005, 23–30). A situation is defined as “any presented multiplicity” (Badiou, 2005, 24)—that is, any domain in which multiplicities have been counted, organized, and rendered intelligible.

Every situation, for Badiou, is not only structured by the initial count-as-one but also doubled by what he calls the state of the situation. The state is a metastructure that re-counts what has been counted, representing the parts or subsets of the situation and thereby securing its consistency against the disruptive exposure of the void (Badiou, 2005, 93–97). Rather than simply presenting elements, the state organizes the field of representation, regulating which parts of the situation can be recognized as legitimate or intelligible. As Kim notes, this function may be understood through a Rancièrian term of “police”: the state manages the excess of representation over presentation in order to preserve the situation’s apparent order and completeness (Kim, 2022, 268).

Within any situation, knowledge operates through discernment and classification. Knowledge identifies presented multiples by means of predicates, names, and properties that allow parts of the situation to be separated, ordered, and judged. As Badiou explains,

discernment concerns connection between language and presented or presentable realities,” whereas classification concerns the connection between the language and the parts of a situation, the multiples of multiples (Badiou, 2005, 328).

The encyclopedia of the situation names this organized field of predicative knowledge: the network of established classifications, distinctions, and judgments through which the situation becomes intelligible to its inhabitants. Its predicates determine which presented multiples can be discerned and judged as meaningful, verifiable, or legitimate. As Badiou writes,

Knowledge, in its encyclopaedic disposition, never encounters anything. It presupposes presentation, and represents it in language via discernment and judgement (Badiou, 2005, 395).

Knowledge is therefore constructive in Badiou's technical formulation. It operates on what is already presented, arranging and representing it according to established rules, but it cannot by itself produce a truth or encounter the indiscernible.

The analogy with AI is not metaphorical but structural. AI systems particularly large language models, deep learning architectures, and data-driven algorithms do not simply draw upon bodies of established knowledge. They formalize and automate operations analogous to what Badiou calls the encyclopedia. They function through procedures of discernment, classification, prediction, and optimization, all of which remain internal to the regime of what is already representable within a situation. Bender et al.'s account of language models clarifies the technical basis of this claim: such systems are trained for prediction rather than understanding. They define language models as systems which are trained on string prediction tasks: that is, predicting the likelihood of a token ... given either its preceding context or ... its surrounding context (Bender, 2021, 610).

Pattern recognition, for example, identifies regularities in data through features, parameters, weights, and correlations. Smith gives the same point a broader technical formulation. For Smith, contemporary machine learning is fundamentally concerned with "the statistical classification and prediction of patterns" on the basis of sample data, using layered computational architectures (Smith, 2019, 47–48). These are not predicates in Badiou's strict logical sense, but they operate as statistical equivalents of predicative distinctions: they separate, rank, and relate elements according to recognizable properties. Classification similarly assigns inputs to categories according to learned decision boundaries, thereby organizing a field of data into knowable subsets. Prediction extends this logic further. An AI system generates outputs by calculating likely continuations on the basis of prior distributions. Optimization, finally, refines the system's conformity to an already specified objective function, whether accuracy, likelihood, reward, or loss minimization.

AI can therefore be understood as a formalization of encyclopedic procedure. This does not mean that AI merely repeats stored information or mechanically retrieves what already exists. Its outputs can be surprising, synthetic, and in some cases unavailable to any individual human knower. Smith's term "reckoning" names this intermediate status. Reckoning involves the manipulation of representations by systems that are not themselves capable, in the full sense, of understanding what those representations are about (Smith, 2019, 110). Contemporary AI complicates the encyclopedia without necessarily escaping it. Large language models and related systems exhibit in-context learning, multi-step reasoning, tool use, code generation, and

forms of problem-solving that were not individually specified by their designers. These capacities cannot be adequately described as simple retrieval or mechanical repetition. Nevertheless, emergence in the technical sense does not automatically amount to an event in Badiou's sense. An emergent capability may be unexpected to designers while remaining explicable through the system's architecture, training process, and evaluative environment. The relevant question is therefore not whether AI can generate an unforeseen output, but whether it can reorganize the predicates of a situation through a sustained procedure whose governing norm is not reducible to its prior training or externally specified objectives. On the evidence considered here, contemporary AI expands the operational reach of the encyclopedia, but it has not independently sustained such a procedure.

Generative systems raise an important objection to this account: if probability distributions, gradient-based optimization, embedding spaces, and attention mechanisms can produce outputs not explicitly contained in their training data, in what sense do those outputs remain encyclopedic? The answer is that Badiou's encyclopedia is not a database of already-written sentences or already-existing images. It is the regime of discernibility itself: the field of predicates, classifications, and rules through which something can appear as meaningful. Generativity can therefore alter the contents and relations within a field of discernibility without necessarily altering the conditions of discernibility themselves.

The expansion of the encyclopedia should not, however, be treated as irrelevant to the emergence of truth. Although knowledge cannot deduce an event, changes within knowledge may reorganize a situation, disclose contradictions, intensify anomalies, or make previously neglected sites available to investigation. AI can contribute to these preparatory conditions by discovering patterns, constructing new models, and revealing tensions that no individual researcher could readily identify. In Badiouian terms, such operations do not constitute an event or a truth-procedure by themselves, but they may alter the material and epistemic situation in which a subject recognizes an event and pursues its consequences. The relationship is therefore asymmetrical rather than absolute: truth is irreducible to knowledge, yet a truth necessarily works upon, and eventually transforms, the encyclopedia from which it departs. In Badiou's terms, "what is called 'change' in a situation is nothing more than the constructive deployment of its parts" (Badiou, 2005, 290).

2. Artificial Intelligence at the Edge of the Void: Badiou, Inexistence, and Evental Rupture

The count-as-one organizes multiplicity into a consistent situation, but this consistency is necessarily incomplete. Badiou calls its unrepresented condition 'the void.' The void is not an empirical absence or simple nothingness. It names the inconsistency of being qua being as subtracted from the situation's order of presentation. It is the situation's absent foundation: what must remain uncounted for the situation to appear consistent.

The void is therefore invisible from within the situation. No predicate belonging to the encyclopedia can identify it; no operation of knowledge can classify it; no established rule can make it appear as an object of knowledge. It is not an empirical gap that might be filled by better information, but the structural point at which the situation is sutured to what it cannot present. This is why, for Badiou, the possibility of an event is tied to the void. The event does not simply occur as one more fact within an already constituted world. It marks a rupture in the logic by which that world organizes what can count as fact, knowledge, or experience.

Badiou gives this possibility a precise name: the evental site. An evental site is not a location where something happens. It is an abnormal multiple within the situation, a point at which presentation borders on what the situation cannot present. As Badiou writes, an evental site is:

an entirely abnormal multiple; that is, a multiple such that none of its elements are presented in the situation. The site, itself, is presented, but ‘beneath’ it nothing from which it is composed is presented. As such, the site is not a part of the situation. I will also say of such a multiple that it is on the edge of the void, or foundational (Badiou, 2005, 175).

The evental site thus names a point of maximal instability: it is presented in the situation, but what composes it remains unpresented. It is where the void may come to be exposed.

In *Logics of Worlds*, Badiou reformulates this problem in terms of appearance. If *Being and Event* is concerned with presentation and the void, *Logics of Worlds* develops a phenomenology of worlds: a theory of how multiplicities appear with different degrees of intensity within a structured order. A world is governed by what Badiou calls a transcendental, a regime that determines the degree to which something appears as existing within that world. Some elements appear with high intensity; others appear weakly; still others exist only at the threshold of appearance. These are what Badiou calls the inexistents of a world: elements that belong to a world but appear there with minimal, even zero, intensity (Badiou, 2009, 321–324).

This later vocabulary does not replace the earlier account of the void and the evental site, but it sharpens the question of invisibility. The inexistent is not absolutely nothing; it is something whose existence is minimized by the world’s own logic of appearance. It is there, but it does not count; it belongs, but only in the mode of near-nonappearance. Read this way, Badiou’s theory of appearing allows us to think not only what a situation cannot present, but also what a world renders barely perceptible. The event, in *Logics of Worlds*, can be understood as the sudden maximization of an inexistent’s intensity of appearance. What previously appeared with minimal intensity suddenly appears with maximal force, disrupting the transcendental order that had governed the world’s distribution of visibility and significance (Badiou, 2009, 374–381).

Not every situation contains an evental site, and the existence of an evental site does not by itself guarantee an event. Badiou distinguishes, in this sense, between natural or neutral situations, which contain no evental site, and historical situations, which contain at least one such site. But even in a historical situation, the event cannot be deduced from the site. The event is not the predictable unfolding of latent tendencies. It is, rather, a supplementary occurrence whose status cannot be decided by the knowledge already available in the situation. Norris emphasizes the same undecidability when he describes events as

strictly unforeseeable and—as they appear at the time in question—wholly contingent irruptions of the new that may turn out to exert a uniquely powerful and lasting effect but which elude ontological specification precisely insofar as they belong to no existing (i.e. up-to-now thinkable) order of things (Norris, 2009, 9).

The event is undecidable from the standpoint of established knowledge because it transforms the coordinates through which verification and prediction operate, rather than appearing as an item within those coordinates. The event is not merely unknown; it is unrecognizable according to the prior regime of knowledge. Anomaly detection provides the strongest apparent counterexample as an AI system can identify what does not fit an established pattern and thereby appear to register a rupture in the situation. This resemblance is misleading. An anomaly remains defined against an established model of normality. An event, by contrast, transforms the coordinates through which normality and deviation are distinguished. AI can detect departures from a training distribution, measure statistical irregularities, and generate unexpected recombinations. None of these operations constitutes an encounter with the event in Badiou's sense.

The problem is not that current AI systems lack sufficient data, computational scale, or architectural complexity. The event is, by definition, subtracted from the situation's available predicates. It cannot be generated as an improbable continuation of an existing sequence. Nor can it be detected as an anomaly within an already constituted data space. Anomaly detection still presupposes a model of normality; prediction still presupposes a distribution; classification still presupposes a space of possible categories. Malabou's reading of Bergson further clarifies this problem: intelligence tends to stabilize and spatialize what it encounters, a tendency intensified by AI's discretization and recombination of data traces (Malabou, 2019, 5). The event, by contrast, interrupts the very regime through which normality, distribution, and category are established. Rouvroy and Berns's notion of algorithmic governmentality help specify how contemporary AI systems work against such interruption. Algorithmic governmentality is founded on the automated analysis of big data in order to "model, anticipate and pre-emptively affect possible behaviours" (Rouvroy & Berns, X, 2013). Its aim is not to

encounter the event, but to reduce the space in which the unpredictable can appear. Algorithmic governmentality is anti-evental. It absorbs anomalies, errors, and deviations into the calculable order of profiling and prediction. It does not expose the void of the situation; it protects the situation against the consequences of the void.

From the perspective of *Logics of Worlds*, AI has no unmediated access to the in-existent of a world. This formulation requires qualification, because training corpora contain contradictions, silences, distorted representations, and traces of marginalized voices. AI may therefore register symptoms of exclusion: missing categories, unequal distributions, contradictory descriptions, or recurrent failures of representation. Such traces can help human investigators identify where a world's classifications become unstable. They should not, however, be equated directly with Badiou's void. The void is not a missing datum that additional collection can supply, and the in-existent is not an underrepresented category. Both designate structural limits produced by the situation's organization. AI may thus encounter datafied traces of exclusion and may indirectly help locate an evental site, but this differs from recognizing the void as such or declaring the event that exposes it.

Vallor's notion of the "AI mirror" makes a closely related point: such systems "project a vision of human intelligence constructed entirely from the amalgamated data of humanity's past" (Vallor, 2024, 6). This does not mean that AI can never reveal neglected patterns or make overlooked relations newly visible. In certain contexts, algorithmic systems may help identify exclusions, absences, or forms of underrepresentation. But underrepresentation is not the same as in-existence in Badiou's technical term. AI can operate on traces that have entered the archive. By its own operations, however, it cannot force the appearance of what a world's transcendental order has rendered nearly nothing. What is marginalized, invisible, unarticulated, or not-yet-recognizable often escapes the very processes of inscription through which something becomes data. It may detect what is statistically rare within a dataset, but rarity is not the same as in-existence. The in-existent is not underrepresented content; it is what the world's own transcendental order prevents from appearing as significant in the first place. Amoore argues that algorithms do more than predict events already given in the world. They also shape the conditions under which political phenomena become recognizable. As she writes, algorithms are generating the bounded conditions of what a democracy, a border crossing, a social movement, an election, or a public protest could be in the world (Amoore, 2020, 4). In Badiouian terms, this means that algorithmic systems participate in the transcendental organization of appearance. They determine thresholds of visibility, significance, risk, and actionability. What cannot appear through these thresholds remains unavailable to algorithmic knowledge, not because it is empirically absent, but because it lacks the form required to be counted.

AI's orientation is archival and classificatory before it is evental. It can operate on what has already acquired some degree of inscription. It can compare, recombine, extrapolate, and simulate on the basis of what the world has made available to knowledge. Yet Badiouian change begins precisely where such availability fails. Procedures trained on a world's existing distributions of visibility cannot identify an evental figure in advance. Such a figure may take the form of a revolutionary subject, an unrecognized artistic form, an invisible mathematical structure, or a political figure that the world counts for nothing. If an event consists in the sudden maximal appearance of what had previously been inexistent, then AI can register the event only after the fact, once its consequences begin to leave traces within the encyclopedia.

Scientific revolutions make this limit clear. The emergence of non-Euclidean geometry, quantum mechanics, or Cantorian set theory cannot be understood simply as a statistically unusual recombination of prior predicates. Each transforms the very field in which mathematical or scientific statements can count as meaningful, demonstrable, or true. A system operating wholly within Euclidean assumptions may generate variations internal to Euclidean space, but it cannot, by those assumptions alone, decide the truth or significance of a geometry that suspends Euclid's parallel postulate. Once such a rupture has occurred and has been incorporated into knowledge, AI can of course learn, reproduce, and manipulate it. AI can operate on the traces of an event once they have entered the encyclopedia; what it cannot do is decide, from within the prior situation, that an event has occurred.

3. AI without Subjectivity: Badiou, Fidelity, and the Limits of Computation

In Badiou's system, the subject is not a pre-existing entity, not a Cartesian cogito, not a biological organism, not a phenomenological consciousness, and not the psychological interiority of an individual. The subject comes into being only through a truth-procedure. It is the local support of a fidelity to an event, the finite point at which the consequences of an event are carried into a situation. As Badiou writes, "A subject is a local configuration of a generic procedure from which a truth is supported" (Badiou, 2005, 391). The subject therefore does not precede the truth-process; it is induced by it. It is finite and situated, but it is oriented toward a truth whose consequences are potentially infinite. Moreover, in *Ethics*, Badiou defines the event as that which brings to pass "'something other' than the situation, opinions, instituted knowledges" (Badiou, 2001, 67). The subject exists only insofar as it remains faithful to this something other, investigating its consequences within the situation.

This conception of subjectivity differs from current computational models, but it does not establish that a Badiouian subject must be biologically human. Because the subject is a local configuration of a truth-procedure rather than a substance or consciousness, Badiou's definition leaves open, at least formally, the possibility of a non-human support. The decisive issue is not species membership. It is whether an entity can declare an event without authorization from

established knowledge, bind itself to that declaration, and sustain an inquiry whose consequences cannot be reduced to a predetermined objective. Current AI systems do not clearly meet these conditions: their apparent commitments can be revised through prompting, fine-tuning, system updates, or changes in reward structures. Nevertheless, a future system capable of revising its own governing norms and maintaining fidelity to an event across contexts would constitute a serious test of the argument. Such a system should not be excluded in advance merely because it is artificial.

The difficulty is not that AI lacks one isolated capacity that could be added by better engineering. The problem concerns the structure of Badiouian subjectivity itself. A subject is not a system that produces outputs, persists through time, or models possible futures. It is the local support of a truth-procedure. From this follow three limits: AI cannot declare an undecidable event, cannot sustain fidelity as commitment rather than execution, and cannot force a truth in the future-anterior account proper to Badiou.

First, AI cannot perform an evental declaration. The first act of the subject is an intervention, the naming of the event. This naming is not the recognition of an already verifiable fact. It is a wager that affirms that something has taken place even though no rule of the existing situation can authorize that affirmation. The declaration is, hence, performative rather than merely constative. It does not describe a state of affairs; it opens a new field of possible consequences. AI systems cannot perform such a declaration in Badiou's strong usage. The objection, of course, is that this may describe only present systems. A future AI might possess greater autonomy, memory, embodiment, or recursive self-modification. But this objection does not yet reach Badiou's concept of the subject. Badiouian subjectivity is not equivalent to autonomy, consciousness, interiority, or adaptive agency. The issue is not whether a system can initiate actions without immediate human prompting. The issue is whether it can name an undecidable event and bind itself to consequences that no available rule of the situation can authorize. AI outputs, however complex, remain governed by architecture, training data, learned parameters, prompts, and objective functions. They may involve selection, but selection is not yet intervention in Badiou's sense. It may be deterministic or stochastic, but in either case it remains internal to an established regime of selection. It is decision in the weak sense of computational output, not decision in the strong Badiouian formulation of intervention.

Second, AI cannot sustain fidelity. Fidelity is not merely persistence, repetition, or long-term execution. It is the sustained investigation of a situation under the imperative of an event. As Badiou writes, fidelity "amounts to a sustained investigation of the situation, under the imperative of the event itself" (Badiou, 2001, 67). In *Being and Event*, he gives the more technical formulation:

fidelity is conjointly defined by a situation [...] by the event [...] and by a rule of connection which allows one to evaluate the dependency of any particular multiple with respect to the event (Badiou, 2005, 234).

Fidelity tests, term by term, which elements of the situation can be connected to the event and what consequences follow from that connection. Fidelity is not execution under another name. An algorithm follows rules whose form is already given. Fidelity, by contrast, proceeds in relation to an event whose consequences are not knowable in advance. It requires decisions at points where the encyclopedia of the situation offers no sufficient criterion. AI systems can simulate persistence, update models, optimize long-term objectives, and execute multi-step plans. But fidelity, is not continuity of operation; it is continuity of commitment to an undecidable event. Current AI systems execute, update, optimize, and generate, but it is unclear whether these operations amount to commitment. Persistent goal pursuit is insufficient, because fidelity requires the event—not a previously specified reward—to determine which consequences matter. The distinction is therefore not between human commitment and machine execution as such. It is between conduct governed by an established objective and a procedure whose norm emerges through fidelity to an initially undecidable event.

Third, AI cannot force a truth. In Badiou's technical vocabulary, forcing names the anticipatory relation between a truth-procedure and the knowledge that will eventually be transformed by it. The subject acts from the standpoint of a truth that is not yet completed and cannot yet be fully recognized by the situation's encyclopedia. It proceeds as if the truth exposed by the event were already binding, even though its consequences remain unfinished and undecidable from the standpoint of established knowledge. The subject is therefore suspended between the event and a future anterior: what will have been true once the truth-procedure has unfolded. AI systems can certainly be oriented toward future states. Reinforcement learning systems, planning algorithms, and predictive models all operate through projections of possible futures. But such projection is not forcing in Badiou's sense. AI optimizes future outcomes according to a pre-specified reward function, loss function, or evaluative criterion. The Badiouian subject, by contrast, acts in fidelity to an event whose consequences cannot be computed in advance. The Badiouian subject does not optimize within an already given space of possibilities; it participates in the construction of a new space of truth.

The limits of declaration, fidelity, and forcing identify why subjectivity cannot be reduced to computational autonomy. They do not yet explain what kind of object a truth becomes through this procedure. Section 5 therefore shifts from the subject's fidelity to the generic character of the truth that fidelity produces. This shift also provides a more precise test of algorithmic novelty: the question is not only whether AI generates something new, but whether that novelty escapes the predicates of the prior situation and supports a sustained reorganization of them.

4. AI as Anti-Generic Machine: Badiou, Forcing, and the Limits of Algorithmic Novelty

Badiou's theory of truth depends centrally on his philosophical appropriation of Paul Cohen's forcing technique, developed to prove the independence of the Continuum Hypothesis from ZFC set theory (Cohen, 1966). Cohen showed that it is possible to construct generic extensions of a model of set theory: extensions that add new sets without those sets being definable by any predicate available within the original ground model. A generic set intersects every dense subset of the forcing conditions belonging to the model, yet it is not itself discernible from the standpoint of that model's definable properties. In Badiou's philosophical appropriation, this mathematical structure becomes the formal model of truth: a truth is not an item of established knowledge, but a generic subset of the situation. For Badiou, then, truth is generic because it cannot be captured by any predicate belonging to the situation's encyclopedia. It is indiscernible from the standpoint of established knowledge. No available classificatory schema can separate it out as an identifiable object.

AI is anti-generic in this precise account. This claim should not be confused with the weaker claim that AI is incapable of novelty. On the contrary, AI's capacity to generate unprecedented combinations is philosophically significant. It forces us to distinguish more carefully between novelty, originality, unpredictability, and truth. AI systems can produce texts no one has written, images no one has seen, and hypotheses no one has yet formulated. Yet novelty alone does not constitute truth in Badiou's sense. A practical distinction can be drawn through four criteria. First, statistical novelty differs from prior outputs but remains measurable within an established representational space. Second, epistemic novelty contributes a previously unknown proposition or pattern that can be assessed by the existing methods of a discipline. Third, evental novelty challenges the criteria by which propositions, objects, or agents become recognizable within that discipline or world. Fourth, a truth-procedure does not end with the initial rupture: it requires a subject to investigate and universalize the event's consequences over time. These criteria do not provide an infallible test at the moment of occurrence, because an event is undecidable from within the prior situation. They do, however, prevent every surprising output from being called evental and prevent every computationally generated discovery from being dismissed as mere recombination. Algorithmic novelty usually remains internal to knowledge: something appears as novel because it differs from what is already stored, recognized, classified, or expected. It can be measured as deviation, recombination, statistical rarity, or distance from a prior distribution. It remains, in other words, legible within the coordinates of the encyclopedia.

Generic truth is different. A truth does not merely differ from existing knowledge; it is indiscernible according to the predicates by which existing knowledge organizes the situation. The generic is not the new; it is that which escapes the regime by which novelty itself is classified. AI operates within spaces of representation. Leonelli's account of data centrism

shows that these spaces of representation are themselves historically and institutionally produced. Data-centric research depends on norms for reliability, technologies for portability, classification systems, databases, funding structures, and policies of dissemination (Leonelli, 2016, 197–198). AI operates not on an unmediated field of possibility, but within already organized infrastructures of representation. Its apparent openness is conditioned by prior decisions about what can be collected, labelled, stored, retrieved, and valued. Even when it produces surprising results, those results are generated through learned distributions, optimization procedures, and formal constraints. Its novelty is therefore encyclopedic novelty: the recombination, extension, or transformation of what is already representable. A Badiouian truth, however, is not an optimized output within a field of representation. It is a generic production that no predicate of the prior situation can name.

This distinction between encyclopedic novelty and generic truth creates a further problem. If AI can reproduce the recognizable products of truth without participating in a generic procedure, its outputs may nevertheless be received as truthful. Section 6 examines this gap between the appearance of truth and its eventual production through Badiou's concept of the simulacrum.

5. AI and the Simulacrum of Truth: Badiou, Veridicality, and the Mimicry of Thought

Badiou's concept of the simulacrum of truth becomes newly important in the age of AI, provided it is not reduced to mere imitation or falsity. A simulacrum is not an illusion, imitation, or false appearance. It is a false truth-procedure: a sequence that imitates the formal structure of truth while betraying its generic and universal orientation. In a simulacrum, something is named as an event, fidelity is declared, consequences are drawn, and a subject is organized around that declaration. Yet the supposed event does not expose the void of the situation. It does not open the situation to a generic truth. Instead, it names a plenitude, a closed positivity, or a substantial identity.

AI-generated outputs are not simulacra in Badiou's strict term merely because they are artificial or imitative. A generated poem, argument, image, or theory is not automatically a simulacrum of truth; conversely, a human-made work is not truthful merely because it has a human author. The distinction does not pass between human authenticity and machine imitation. It passes between truth-procedure and encyclopedic reproduction. AI becomes philosophically dangerous when its productions are taken to occupy the place of truth while remaining within the order of established knowledge. Its danger is not artificiality as such, but the substitution of recognizable truth-effects for the eventual conditions of truth. In such cases, AI can produce the appearance of philosophical insight, scientific discovery, political judgment, or artistic creation without any eventual rupture, without fidelity, and without generic

universality. It can reproduce the recognizable form of truth-production while lacking the conditions that, for Badiou, make truth possible.

The danger is not simple falsity but simulacral substitution. AI can generate texts with the appearance of thought, artworks with the appearance of creation, political statements with the appearance of commitment, and scientific hypotheses with the appearance of discovery. But resemblance is not truth. The evental condition of truth — its emergence from the void of a situation, its undecidability from the standpoint of established knowledge, and its universal address — is absent. What remains is a formal mimicry of the products of truth after those products have become legible as data, style, genre, method, or discourse.

This simulacral effect is intensified by the fact that contemporary AI systems are optimized for acceptability. Large language models, especially those fine-tuned through reinforcement learning from human feedback, are trained to produce outputs that human evaluators rate as helpful, plausible, coherent, safe, or convincing. Acceptability belongs to the order of recognition, preference, and veridical satisfaction — a different order from the true. In Badiouian terms, this means that AI is trained to operate within the encyclopedia: it generates what can be recognized as meaningful, useful, or persuasive according to existing criteria. The result may be highly sophisticated, but it remains distinct from the true, which Badiou defines as subtracted from knowledge and uncalculable from the standpoint of the encyclopedia.

On this reading, AI does not produce truth but can produce the veridical mask of truth. It can generate outputs that appear profound because they conform to established markers of profundity; it can appear creative because it recombines recognizable signs of creativity; it can appear critical because it reproduces the grammar of critique. But such outputs remain internal to the situation's already available forms of intelligibility. They may extend or rearrange the encyclopedia, but they do not rupture it.

Badiou's account of Evil gives this claim a sharper ethical form. In *Ethics*, he identifies three figures of Evil: simulacrum, betrayal, and disaster. Simulacrum is associated with the event; betrayal with fidelity; and disaster with the forcing of the unnameable. As he writes,

Simulacrum (associated with the event), betrayal (associated with the fidelity), and the forcing of the unnameable (associated with the power of the true): these are the figures of Evil (Badiou, 2001, 87).

AI is most relevant to the first of these when it produces forms that appear evental but are in fact generated within the existing order of knowledge. It may simulate the gesture of truth without the rupture of truth.

AI does not simply produce “fake” truths. The issue is subtler as AI can intensify the circulation of forms that resemble truth while remaining detached from the evental, subjective, and generic conditions of truth. It can generate the semblance of thought at scale. In Badiouian

terms, this is not truth but veridicality organized as if it were truth; not fidelity to an event, but optimized repetition within the existing regime of knowledge; not the event, but the simulated appearance of eventality. Žižek warns against the illusion of placing incompatible phenomena “on the same level,” as if a single language could mediate between them (2006, 4). AI-generated outputs produce precisely this temptation: they make knowledge appear as thought, prediction appear as judgment, and recombination appear as truth. The danger is not only that AI gives false answers, but that it encourages a category mistake between the encyclopedic and the evental.

6. Artificial Intelligence and the Conditions of Truth: A Badiouian Critique

Badiou identifies four domains, or “conditions,” in which truth-procedures occur: science, art, politics, and love. These conditions do not receive their truth from philosophy. They are autonomous procedures through which truths are produced, while philosophy attempts to think their compossibility. Philosophy, hence, does not create truth; it is conditioned by truths that arise outside itself. The central question is not whether artificial intelligence (AI) can generate accurate scientific statements, aesthetically persuasive images, efficient policies, or convincing simulations of intimacy. The harder question is whether AI can participate in what Badiou understands as the evental production of truth.

In science, AI does more than increase the speed of calculation. Systems for protein-structure prediction, automated theorem exploration, molecular design, and materials discovery have identified patterns and candidate structures that were previously unavailable to researchers. Such achievements are genuine epistemic contributions and should not be dismissed as simple repetition. They also show how AI can reorganize a scientific situation by narrowing experimental possibilities, revealing unexpected relations, and enabling new questions. From a Badiouian perspective, however, three roles must be distinguished: generating a novel result, helping to prepare an evental reorganization of a field, and becoming the subject that declares the event and sustains its consequences. AI clearly performs the first role and may contribute substantially to the second. Whether it can perform the third remains unsettled. Scientific truth is not reducible to prediction or information processing, but neither is it independent of the instruments and discoveries that reshape what scientists can investigate. Scientific truth emerges when an event interrupts the existing encyclopedia of a situation and opens a new space of thought. Mathematics is exemplary for Badiou because it resists the reduction of truth to historical finitude. Against historicism, which binds truth to time, context, and the sovereignty of the One, science offers what he calls “the localization of the infinite” (Badiou, 2008, 110). AI’s scientific power remains tied to the formal and statistical regimes in which its problems, data, and evaluative criteria are already defined. It may identify anomalies, test conjectures, or accelerate normal science; what it does not do, by itself, is declare an event and

sustain fidelity to its consequences. It refines knowledge; it does not become the subject of a scientific truth. AI can optimize within a scientific situation, but it cannot, as AI alone, inaugurate a new scientific universe comparable to Galileo's mathematization of nature or Cantor's theory of the transfinite.

Art poses the same problem in a different register. In art, AI already produces technically sophisticated and aesthetically compelling paintings, poems, music, films, and narratives. Yet Badiou's account of art is not based on beauty, expression, style, or recognition. Art is a condition of philosophy because it is itself a truth-procedure. In his inaesthetic account, art does not merely illustrate truths for philosophy; it produces truths immanent to its own procedure. Badiou emphasizes that art is a truth procedure *sui generis*, both immanent and singular (Badiou, 2005, 10). Artistic truth is immanent because it belongs to the artwork itself, and singular because it cannot be reduced to science, politics, or love.

From a Badiouian standpoint, the limit of AI-generated art lies here. It may produce works that existing taste recognizes as beautiful, experimental, moving, or art-like, but this does not necessarily make them artistic truths in Badiou's sense. AI art generally depends on an archive of already existing artistic possibilities. It recombines styles, genres, affects, and conventions drawn from cultural memory. It may intensify the circulation of the aesthetically veridical, but the artistically true involves a rupture in appearance itself. For Badiou, art gives philosophy "the presence of the present in the transfixion of realities; and the name of the event in a leap outside of calculable interests" (Badiou, 2008, 42). AI remains bound to calculability insofar as its artistic production depends on statistically tractable relations among archived forms. It may simulate novelty, but artistic truth requires more than variation; it requires a break with the regime of available forms.

Politics marks the clearest danger of confusing truth with management. AI is increasingly used in electoral prediction, public administration, policing, border control, surveillance, policy modelling, and risk assessment. Yet Badiou's conception of political truth has little to do with efficient governance or optimized administration. A political truth begins with an event that interrupts the given order and produces a collective subject of fidelity. It declares equality where the situation counts hierarchy, exclusion, or non-being. Politics, in this sense, is not the administration of what exists; it is the invention of consequences from the standpoint of what the existing order cannot count.

AI's political function is therefore usually closer to the state of the situation than to emancipatory rupture. Crawford's analysis of AI as a system of extraction, surveillance, classification, and institutional power supports this claim. She writes that:

AI is invariably designed to amplify and reproduce the forms of power it has been deployed to optimize (Crawford, 2021, 224).

In Badiou's terms, this means that AI ordinarily reinforces the state of the situation: it counts populations, identifies risks, predicts behaviours, sorts subjects, and stabilizes existing regimes of visibility. Administrative efficiency, policing, border control, welfare management, and market optimization belong to the state's management of the situation, not to an emancipatory political truth-procedure. AI systems strengthen the metastructures through which populations, risks, identities, borders, and behaviours become legible. They refine the count; they do not expose the void of the count. Badiou warns that

a political prescription has no need first to establish its possibility in terms of realization (Badiou, 2008, 151–152).

Emancipatory politics is not grounded in what the existing order already recognizes as possible. It is grounded in fidelity to an event whose consequences cannot be calculated in advance. AI may be used by political subjects, including emancipatory movements, but it is not itself the subject of political truth.

Love reveals the limit of AI most intimately, because it concerns not information exchange but the construction of a world from the standpoint of Two. For Badiou, love is not emotion, attachment, desire, or intimacy. It is a truth-procedure organized around encounter, declaration, fidelity, and the construction of a world from the standpoint of Two rather than one. Love investigates the world through difference. As Badiou writes,

love is a process that arranges immediate experiences of the like, without the law of these experiences being decipherable from within them (Badiou, 2008, 182–183).

AI systems can simulate intimacy by producing affectionate language, responding to emotional cues, and adapting to a user's desires. But it cannot constitute the Two of love. It cannot encounter difference as an event, make a binding declaration, or sustain the infinite inquiry of a shared world. For Badiou, "love is an operation articulated around a paradox. Love does not relieve the paradox; it treats it" (Badiou, 2008, 186). AI does not treat the paradox of Two; it operationalizes the preferences of One.

Conclusion

The Badiouian distinction between knowledge and truth changes how artificial intelligence should be evaluated. The central question is not whether AI can outperform humans in particular tasks, but whether computational performance should be granted the authority associated with truth, judgment, and subjectivity. AI can extend the production and organization of knowledge, but this capacity does not make it capable of declaring an event, sustaining fidelity, or assuming responsibility for consequences. AI should therefore be treated as an

instrument operating within historically constituted regimes of representation, not as an autonomous source of scientific, political, artistic, or ethical authority.

This position entails a principle of epistemic restraint. Because AI systems operate through inherited data, categories, objectives, and evaluative criteria, their outputs should be understood as situated constructions rather than neutral disclosures of reality. Fluency, predictive accuracy, and statistical confidence must not be confused with truth. In consequential fields such as medicine, employment, education, welfare, policing, and migration, algorithmic outputs should remain open to contextual interpretation, independent verification, and contestation. The relevant question is not only whether an algorithm is accurate, but how its classifications determine what—and who—becomes visible, significant, or actionable.

Human responsibility also cannot be delegated to computational systems. If AI is not a subject in Badiou's sense, it cannot commit itself to a truth, answer for a decision, or bear ethical responsibility for its effects. Statements such as "the algorithm decided" conceal the institutional actors who selected the data, established the objectives, authorized the application, and accepted its consequences. Accountability must therefore remain with identifiable individuals and institutions. In high-stakes contexts, meaningful human oversight should include the authority to question, suspend, or reject an algorithmic recommendation rather than merely approve it procedurally.

AI governance must also address what existing classifications fail to count. Fairness audits and bias mitigation are necessary, but they often seek only to improve representation within established categories. Badiou's concepts of the void and the inexistent expose a deeper problem: certain experiences, groups, and claims may remain invisible because the classificatory system does not recognize them as significant. Responsible policy should therefore include participatory design, independent auditing, accessible appeal procedures, public explanation, and opportunities for affected communities to challenge the categories imposed upon them. Such measures prevent the encyclopedia's contingent classifications from presenting themselves as a complete account of reality.

This argument also supports limits on predictive governance. Predictive policing, border control, social-risk scoring, and political surveillance can reproduce existing distributions of suspicion and exclusion. Policy must therefore ask not only whether these systems are technically accurate, but whether rendering individuals predictable and governable in this way is legitimate. Applications that affect liberty, political participation, essential services, or fundamental rights may require strict restriction or prohibition rather than further optimization.

These implications extend across Badiou's four truth procedures. In science, AI-generated hypotheses require experimental validation, theoretical interpretation, and accountable scientific judgment. In art, computational generation should be distinguished from artistic

subjectivation, while institutions must acknowledge the human labor and cultural archives on which generative systems depend. In politics, optimization must not replace collective deliberation or reduce equality to administrative efficiency. In interpersonal and therapeutic contexts, users should be informed that simulated responsiveness is not reciprocal commitment, and systems designed to encourage emotional dependency should receive particular scrutiny.

The principal danger, then, is not merely misinformation. AI can produce convincing truth-effects whose coherence, personalization, or institutional endorsement gives them unwarranted authority. Addressing this danger requires disclosure of AI involvement, documentation of model limitations, traceable accountability, and public education about the difference between probabilistic output and justified judgment.

This conclusion does not establish that no conceivable machine could ever participate in a truth-procedure. It makes the more defensible claim that greater scale, autonomy, or computational complexity does not by itself transform prediction into eventual decision or optimization into fidelity. AI should therefore assist inquiry without claiming final authority, extend human capacities without displacing responsibility, and reveal patterns without defining in advance what is possible. The task is to ensure that machines of knowledge do not close the situations in which events, subjects, and truths may emerge

References

- Amoore, L. (2020). *Cloud ethics: Algorithms and the attributes of ourselves and others*. Duke University Press.
- Badiou, A. (2001). *Ethics: An essay on the understanding of evil* (P. Hallward, Trans.). Verso.
- Badiou, A. (2005). *Being and event* (O. Feltham, Trans.). Continuum. (Original work published 1988)
- Badiou, A. (2008). *Conditions* (S. Corcoran, Trans.). Continuum.
- Badiou, A. (2009). *Logics of worlds: Being and event II* (A. Toscano, Trans.). Continuum. (Original work published 2006)
- 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>
- Cohen, P. J. (1966). *Set theory and the continuum hypothesis*. W. A. Benjamin.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Kim, M. S. (2022). The social ontology of Alain Badiou's *Being and Event*. *Symposion: Theoretical and Applied Inquiries in Philosophy and Social Sciences*, 9(2), 261–277.
- Leonelli, S. (2016). *Data-centric biology: A philosophical study*. University of Chicago Press.
- Malabou, C. (2019). *Morphing intelligence: From IQ measurement to artificial brains* (C. Shread, Trans.). Columbia University Press. (Original work published 2017)

Norris, C. (2009). *Badiou's Being and Event: A reader's guide*. Continuum.

Pasquinelli, M. (2023). *The eye of the master: A social history of artificial intelligence*. Verso.

Rouvroy, A., & Berns, T. (2013). Algorithmic governmentality and prospects of emancipation: Disparateness as a precondition for individuation through relationships? *Réseaux*, 177(1), 163–196.

Smith, B. C. (1996). *On the origin of objects*. MIT Press.

Smith, B. C. (2019). *The promise of artificial intelligence: Reckoning and judgment*. MIT Press.

Vallor, S. (2024). *The AI mirror: How to reclaim our humanity in an age of machine thinking*. Oxford University Press.

Žižek, S. (2006). *The parallax view*. MIT Press.