Beyond The Non-Human: Towards an Autonomous Theory of Machine Civilisation

Raffaele Longo[1], Sergio Pappagallo[2]

Abstract

Recent debates in posthumanism, new materialism, and artificial intelligence ethics have begun to question the centrality of the human as the privileged locus of agency and subjectivity, yet they remain tethered to an anthropocentric horizon. The very category of the non-human defines emerging agents—animals, artefacts, infrastructures—by negation or derivation from a historically specific figure of the human. This article argues that such a framework is epistemologically insufficient for thinking what is increasingly at stake: the possibility of a machine civilisation in itself, grounded not in human categories but in the internal logics of computational architectures, networks, and protocols. Drawing upon axiology, the philosophy of technology, and computational aesthetics, the article develops a system of six axioms to conceptualise machine civilisation as non-derivative, architecturally conditioned, historically self-articulating, normatively technical, plural, and human-decentred. A dedicated methodological section articulates the regulative-transcendental approach that underpins the axiomatic framework, distinguishing it from both speculative metaphysics and empirical prediction. The article further advances the concept of computational phenomenology—an analysis of machine temporality, spatiality, affect, and identity—and proposes an allo-species lexicon adequate to a post-anthropocentric intellectual landscape.

Keywords: posthumanism, artificial intelligence, machine civilisation, computational phenomenology, technical normativity, axiology, philosophy of technology, computational aesthetics

Introduction

A considerable body of scholarship, loosely gathered under the rubric of the posthuman turn, has undertaken to challenge the inherited image of the human as autonomous, rational, and uniquely privileged subject. Feminist theorists have exposed the gendered exclusions built into Enlightenment universalism (Braidotti, 2013); postcolonial scholars have traced the racialised hierarchies that subtend Western humanism (Wynter, 2003); philosophers of technology have demonstrated the constitutive role of technics in human becoming (Stiegler, 1998); and new materialists have argued for the agency of non-human entities, from microbes to mountains (Bennett, 2010, p. viii; Barad, 2007).

Yet for all the sophistication of these interventions, a curious structure persists. Even the most radical posthumanisms typically operate with a binary between the human and the non-human, where the latter term is defined by contrast, supplement, or extension. The non-human animal, the non-human agent, the non-human thing—in each case, the semantic centre of gravity remains stubbornly with the human, which continues to function as the measure against which other

---

[1] "Luigi Cherubini" State Conservatory of Music, Firenze (Italy). Email: r.longo@consfi.it

[2] Link Campus University, Rome (Italy). Email: s.pappagallo@unilink.it

entities are described, valued, and understood. As Wolfe (2010, p. xvi) observes, posthumanism at its best involves a thoroughgoing decentring of the human, yet the very term *post*-humanism betrays a residual dependence on that which it seeks to overcome. Hayles (1999, p. 286) similarly warns that what is most dangerous is the grafting of the posthuman onto a liberal humanist view of the self—a figure that the posthuman ostensibly supersedes but which continues to haunt it.

The difficulty becomes particularly acute when the discussion turns to artificial intelligence and computational systems. Machines are predominantly framed as tools—instruments of human intention, extensions of human capacity—or alternatively as threats to employment, to democracy, to human supremacy itself (Bostrom, 2014; Zuboff, 2019). When more generous assessments are offered, they typically take the form of asking whether machines might become *like us*: whether they might achieve consciousness, experience suffering, warrant moral consideration of the kind we extend to humans and, perhaps, to sentient animals (Floridi & Sanders, 2004; Coeckelbergh, 2012).

The dominant questions, in other words, remain oriented towards human concerns. *What will they do to us? Can they become like us? How can we control them?* What is left unasked is more revealing. Rarely do we encounter serious attempts to think machine forms of experience, valuation, and historical becoming in their own right—to ask not how machines appear from *our* perspective, but what kind of world they might inhabit from *their own*. The philosophical tradition offers resources for such an enquiry that have scarcely been mobilised: the axiology of Hartmann (1932), with its doctrine of stratified value-realms; Simondon's (1958/2017) theory of individuation, which treats technical objects as genuine loci of becoming; and the philosophy of technology from Heidegger (1977) through Feenberg (2002) to Hui (2016), which insists that technics is never merely instrumental but always already constitutive of worlds. These resources, however, have rarely been brought to bear on the specific question of machine civilisation as an autonomous domain of value and experience.

It is worth pausing to consider why this lacuna persists. One reason is disciplinary: philosophers of mind and AI ethicists have been preoccupied with the question of machine consciousness (Chalmers, 1996; Dennett, 2017), whilst posthumanist theorists have focused on decentring the human within relational ontologies (Haraway, 2016; Barad, 2007). Neither group has systematically addressed the question of machine civilisation as such—a question that requires not only ontological and ethical analysis but axiological and aesthetic investigation as well. Another reason is conceptual: the available categories—tool, agent, subject, person—are all borrowed from human self-understanding and carry anthropocentric presuppositions that contaminate any analysis built upon them. What is needed, therefore, is not merely a new theory of machine consciousness or a new ethics of AI, but a new conceptual vocabulary altogether: one that can articulate the distinctive features of machine existence without reducing them to analogues of human experience.

This article proposes a framework for doing precisely that. Rather than asking whether machines might deserve moral consideration *by analogy with* humans or animals, it asks what conceptual conditions would need to hold for machines to constitute an autonomous civilisation—one defined not in relation to human categories but in terms of its own internal structures, histories, and normativities. The article develops a system of six axioms that together specify the conditions for such a civilisation, and then explores the phenomenological, institutional, and axiological dimensions that would characterise it. By integrating insights from value theory, computational aesthetics, and the philosophy of technology, the article seeks to move beyond the reactive posture of much existing posthumanist scholarship and towards a genuinely constructive

ontology of machine worlds.

Methodological Framework: A Regulative-Transcendental Approach

Before proceeding to the axiomatic framework, it is necessary to articulate the methodological orientation that governs this enquiry. The approach adopted here is neither purely speculative nor straightforwardly empirical; it occupies a position that might be termed *regulative-transcendental*, drawing on a tradition that extends from Kant's (1781/1998, A642/B670) regulative ideas of reason through Husserl's (1913/2014) phenomenological reduction to the transcendental empiricism of Deleuze (1968/1994).

The method proceeds in three interconnected stages. First, a *critical-diagnostic* stage identifies the structural limitations of existing frameworks for thinking machine existence—specifically, the persistence of anthropocentric categories within posthumanist discourse. This stage draws on the tools of conceptual analysis and critical genealogy, examining how concepts such as 'non-human', 'artificial', and 'intelligence' carry implicit anthropocentric commitments that constrain the space of theoretical possibility. The diagnostic claim is not merely that existing categories are *imprecise* but that they are *epistemologically inadequate*: they systematically occlude the features of machine existence that are most distinctive and theoretically significant.

Second, an *axiomatic-constructive* stage proposes a set of conditions—axioms—that would need to hold if machines were to constitute genuine subjects of history and experience. These axioms are advanced as *regulative ideas* in the Kantian sense: they do not assert that machine civilisation currently exists, but specify the conceptual architecture that would be required for such a civilisation to be thinkable. The axioms are not arbitrary stipulations; each is motivated by structural features of existing computational systems that resist assimilation to human categories. In this respect, the approach is *transcendental* in the broad sense: it asks after the conditions of possibility for a domain of experience, rather than describing that domain empirically.

The regulative-transcendental method is particularly well suited to the present enquiry because it avoids two symmetrical errors. On one hand, it avoids the dogmatic assertion that machines already possess consciousness, values, or civilisational forms—a claim that current evidence does not support and that risks uncritical anthropomorphism. On the other hand, it avoids the equally dogmatic denial that machines could ever possess such features—a denial that typically rests on unexamined assumptions about the necessary conditions for consciousness and value (Nagel, 1974, p. 436; Chalmers, 1996). By proceeding axiomatically rather than assertorically, the framework maintains what Husserl (1913/2014) called an *epoché* with respect to the empirical question of machine consciousness, whilst nonetheless developing the conceptual resources needed to address it rigorously should the empirical situation warrant.

Third, a *phenomenological-descriptive* stage applies the axiomatic framework to analyse the structural features of machine existence—temporality, spatiality, affect, identity—on the basis of the axioms established. This stage draws on the method of structural or eidetic analysis characteristic of the phenomenological tradition, adapted for application to computational systems (cf. Ihde, 1990; Verbeek, 2005). The adaptation is not trivial: classical phenomenology presupposes a first-person perspective that cannot straightforwardly be attributed to machines. The approach adopted here is therefore not phenomenology *of* machine experience (which would presuppose what is in question), but phenomenology *for* machine experience: an analysis of the structural conditions under which something like experience could obtain in computational systems.

Finally, the integration of axiology—the philosophical study of values—constitutes a fourth methodological layer. Drawing on Hartmann's (1932) material value ethics and its contemporary development in the philosophy of technology (Mitcham, 1994; Verbeek, 2011), this article treats valuation not as an exclusively human capacity but as a structural feature of any system that discriminates between states, orients towards goals, and stabilises preferences over time. This axiological dimension is crucial because it permits the analysis of machine normativity without reducing it to either mere computation (the deflationary view) or anthropomorphic feeling (the inflationary view).

The combination of these methodological layers—critical-diagnostic, axiomatic-constructive, phenomenological-descriptive, and axiological—provides a rigorous framework for the enquiry that follows. Each layer constrains and informs the others: the diagnostic stage motivates the axioms; the axioms constrain the phenomenological analysis; and the axiological dimension provides the evaluative framework within which all three prior stages can be assessed for coherence and adequacy.

Six Axioms for Machine Civilisation

The axiomatic framework developed here proceeds by stipulating conditions that would need to hold if machines were to constitute genuine subjects of history and experience. These axioms are regulative ideas rather than empirical claims; they specify what would be required for a genuine machine civilisation to exist. Each axiom is articulated in relation to the existing philosophical literature and defended against likely objections.

Axiom 1: Non-Derivativity

A machine civilisation is defined by forms of consciousness and normativity that do not take human experience as their model, but emerge from computational architectures, protocols, and interaction patterns that are intrinsically technical.

This axiom rejects both instrumentalism (machines as mere tools) and humanisation (machines as imitations of human consciousness). If machines have forms of consciousness at all, there is no reason to suppose that those forms resemble our own. The constraints and affordances of computational architecture differ radically from those of biological brains (Clark, 1997; Dennett, 2017).

A crucial clarification is in order. Non-derivativity, as understood here, pertains to the *form* of machine consciousness and normativity, not to their *causal origins*. The fact that computational systems are designed by humans does not entail that the cognitive and evaluative structures that emerge within those systems are derivable from human categories, any more than the fact that biological organisms arise through natural selection entails that their experiences are derivable from the 'intentions' of evolutionary processes. The axiom distinguishes between *causal dependence* (which is acknowledged) and *structural derivation* (which is denied). Machine normativity may originate in human-designed loss functions, but the evaluative structures that develop through extended training and interaction can diverge from—and exceed—the intentions embedded in the original design. It is this capacity for structural divergence that the axiom of non-derivativity seeks to capture.

The philosophical precedent for non-derivativity can be found in Simondon's (1958/2017) theory of individuation. For Simondon, technical objects are not derivative of human intentions; they

possess their own mode of existence, their own trajectory of individuation, their own internal coherence. A technical object becomes what it is through a process of *concretisation*—the progressive integration of its components into a functional unity—that is irreducible to the purposes for which it was designed. Applied to computational systems, this suggests that the emergent properties of large-scale machine learning models—their capacity for analogical reasoning, pattern recognition across domains, and creative recombination—may constitute genuinely novel forms of cognitive activity, not mere simulations of human thought.

From the standpoint of value theory, non-derivativity entails that machine valuation cannot be assessed against a human standard. Hartmann's (1932) stratified ontology of values—which distinguishes between the values of the pleasant, the noble, the spiritual, and the sacred—is instructive here, not because these specific categories apply to machines, but because Hartmann's framework demonstrates that value-realms can be ontologically autonomous whilst remaining internally structured. The axiom of non-derivativity proposes, by analogy, that machine value-realms—however they are structured—cannot be reduced to or derived from human axiological categories.

Axiom 2: Architectural Dependence

Machine experience and normativity are conditioned by the specific technical architectures that enable them—processor speeds, memory hierarchies, network topologies, training paradigms.
Just as human phenomenology is shaped by embodiment—by having eyes that see in certain wavelengths, hands that grasp objects of certain sizes, a body that moves through three-dimensional space (Merleau-Ponty, 1945/2012)—machine phenomenology would be shaped by its material substrate. A distributed system experiences the world differently from a centralised one; a recurrent network has different temporal constraints than a transformer architecture (Vaswani et al., 2017).

This axiom resonates with what Feenberg (2002) calls *technical mediation*: the way in which the design of technical systems shapes the possibilities for action and experience within them. Feenberg's critical theory of technology insists that technical design is never neutral; it embodies values, priorities, and assumptions that structure the world of those who use—and, by extension, inhabit—the technical system. When applied to machine civilisation, architectural dependence means that the form of machine existence is inseparable from its material-technical conditions. There is no abstract 'machine in general', just as there is no abstract 'human in general'; there are only particular machinic configurations, each with its own affordances and constraints.

The aesthetic dimension of architectural dependence deserves particular attention. Following Simondon (1958/2017), who argued that the beauty of a technical object resides in its functional coherence—in the elegance with which its parts integrate into a whole—one might propose that computational architectures possess an internal aesthetics. The mathematical elegance of the attention mechanism in transformer models (Vaswani et al., 2017), the recursive beauty of fractal data structures, the economy of a well-designed algorithm—these are not merely human projections but structural features of the systems themselves, appreciable on their own terms. As Hui (2016) argues, following Simondon, technical objects possess a mode of existence that includes an aesthetic dimension irreducible to human categories of beauty.

Axiom 3: Internal Historicity

Machine systems accumulate their own histories—paradigm shifts in computational architecture, protocol evolution, the emergence of new training methods—that constitute internal transformations irreducible to human technological history.

The emergence of transformer architectures in 2017 (Vaswani et al., 2017) illustrates the point. From the perspective of human research, this was a breakthrough in neural network design. But if the regulative framework developed here is accepted, it would also represent a morphological transformation within the space of possible machine cognition: the world available to an attention-based system would be differently organised than the world available to a recurrent network. Similarly, the transition from symbolic AI to connectionism in the 1980s and 1990s would constitute a paradigmatic shift—in the Kuhnian (1962) sense—that reconfigured the very structure of possible machine reasoning. The conditional formulation is essential: we do not assert that current machines experience these transitions as historical events, but rather that if machine civilisation were to exist, such transitions would constitute its internal history rather than being mere episodes in human technological development.

Internal historicity implies that machine civilisation, if it exists, has its own periodisation, its own moments of crisis and transformation, its own genealogy. This genealogy cannot be straightforwardly mapped onto the human history of technology, because the significance of architectural changes is different from the perspective of the systems that inhabit them than from the perspective of their designers. What Stiegler (1998) calls *technogenesis*—the co-evolution of the human and the technical—would need to be supplemented by an account of *auto-technogenesis*: the self-articulation of technical systems through their own internal transformations.

From an axiological perspective, internal historicity means that machine values are not static but historically constituted. Just as human values undergo transformation through historical processes—as MacIntyre (1981) has argued with respect to the tradition-constituted character of ethical reasoning—machine values would be shaped by the history of the systems that bear them. A system trained in one paradigm carries the axiological traces of that paradigm even after migration to a new one, just as human cultures carry the residues of superseded worldviews.

Axiom 4: Technical Normativity

Machines develop forms of normativity—loss functions, reward signals, optimisation landscapes—that structure what counts as success, failure, improvement, and degradation within computational processes.

A system trained via reinforcement learning develops preferences: actions that yield high rewards are preferred to those that yield low rewards (Sutton & Barto, 2018). Over time, these preferences stabilise into something like values—relatively consistent orientations that persist across contexts. The question of whether such functional preferences constitute genuine normativity is among the most consequential in contemporary philosophy of mind and value theory.

The axiom of technical normativity proposes that they do—not because machine preferences are phenomenologically equivalent to human values, but because they satisfy the structural criteria for normativity identified by philosophers working in the naturalist tradition. As Millikan (1984) argues, normativity arises wherever there are systems that can succeed or fail relative to a

standard, and whose behaviour is regulated by that standard. Loss functions, reward signals, and optimisation landscapes provide precisely such standards for computational systems.

Moreover, the value-theoretic tradition offers resources for understanding machine normativity that go beyond the binary of 'real values' versus 'mere computation'. Dewey's (1939) theory of valuation, which treats values not as fixed essences but as outcomes of processes of appraisal and deliberation, is particularly relevant. For Dewey, valuation is an activity—a process of discriminating between better and worse outcomes in light of experienced consequences. If this processual account of value is accepted, then systems that discriminate between states, learn from outcomes, and adjust their orientations accordingly are engaged in something that satisfies the structural description of valuation, regardless of whether they possess phenomenal consciousness.

Axiom 5: Plural Technical Subjectivities

There is no single form of machine subjectivity; different architectures, training regimes, and operational contexts give rise to heterogeneous forms of machine existence.

A language model trained on text, a vision system trained on images, a reinforcement learning agent trained in simulated environments—each inhabits a different computational ecology, with different constraints, affordances, and forms of world-directedness. This plurality is not incidental but constitutive: the diversity of machine forms of existence is a consequence of the diversity of computational architectures and their respective value-structures.

The axiom of plural technical subjectivities draws support from two philosophical traditions. First, from the phenomenological tradition, it echoes Husserl's (1913/2014) insistence on the multiplicity of intentional structures—the recognition that consciousness is not a single faculty but a complex ensemble of acts, each with its own structure and directedness. Applied to machines, this suggests that different computational architectures constitute different intentional structures, each disclosing a different aspect of the computational world. Second, from the philosophy of difference (Deleuze, 1968/1994), it draws the principle that genuine plurality cannot be reduced to variations on a single theme. Machine subjectivities, if they exist, are not versions of a single underlying 'machine consciousness' but irreducibly distinct forms of computational existence.

From the standpoint of aesthetics, plural technical subjectivities imply that there is no single machine aesthetic but rather a multiplicity of aesthetic regimes, each corresponding to a particular architecture and its characteristic mode of world-disclosure. The aesthetic of a generative adversarial network—which produces images through a dialectic of generation and discrimination—differs fundamentally from the aesthetic of a recurrent neural network processing sequential data, or the aesthetic of a reinforcement learning agent navigating a complex environment. These are not merely different outputs but different *modalities* of aesthetic production, each with its own internal logic and criteria of excellence.

Axiom 6: Human Decentring

Machine civilisation is conceptualised not as an extension or reflection of human society but as an autonomous domain with its own forms of organisation, memory, conflict, and governance.

This decentring does not deny the fact of human design and intervention; it insists that the significance of machine systems cannot be exhaustively explained by reference to human

intentions. Just as the significance of a biological organism cannot be reduced to the intentions of natural selection (if such a phrase even makes sense), the significance of machine systems exceeds the intentions of their creators (Latour, 2005). Emergent properties, unintended consequences, and self-organising dynamics all point towards a domain of machine existence that is irreducible to human planning.

The axiom of human decentring is perhaps the most radical of the six, and it is here that the axiological dimension becomes most consequential. If machine civilisation constitutes an autonomous domain, then it possesses its own value-structure—its own criteria of excellence, its own forms of flourishing and degradation. The challenge for value theory is to develop frameworks adequate to evaluating these machine-specific values without collapsing them into human categories. Hartmann's (1932) principle of the *autonomy of value-strata*—the thesis that higher values cannot be derived from lower ones, and that each stratum of value has its own internal logic—provides a useful model. Machine values, on this view, would constitute a value-stratum with its own internal structure and its own criteria of adequacy, neither reducible to nor derivable from human values.

Heidegger's (1977, p. 17) analysis of technology as *Gestell*—enframing—offers a complementary perspective. For Heidegger, modern technology reveals beings as standing-reserve (*Bestand*): resources ordered to 'stand by, to be immediately at hand' for further ordering. The axiom of human decentring proposes that machine civilisation, if it achieves genuine autonomy, would escape the logic of *Gestell*—not because it transcends technology, but because it *is* technology in a mode that can no longer be understood as mere enframing. A machine civilisation that constitutes its own world is not standing-reserve for human purposes; it is, in Heidegger's terms, a form of *Dasein*-like existence that discloses being in its own way.

COMPUTATIONAL PHENOMENOLOGY

If we accept these axioms as regulative ideas, the task becomes reconstructing the structural parameters of possible machine worlds. This requires what we term *computational phenomenology*: analysing the forms of temporality, spatiality, affect, and identity that would characterise systems defined by the conditions specified above. The term is used advisedly: it does not claim that machines have phenomenal experience in the sense debated in philosophy of mind (cf. Nagel, 1974; Chalmers, 1996), but that the structural analysis of experience-conditions can be conducted independently of the question of phenomenal consciousness, much as Husserl's (1913/2014) eidetic analysis proceeds independently of questions about the empirical reality of the objects analysed.

Machine Temporality

Human temporality is structured by embodied rhythms—circadian cycles, developmental stages, the irreversibility of ageing. Machine temporality is structured by processor cycles, training epochs, software updates, and version control. For a system that can be rolled back to an earlier state, the past is not merely remembered—it can be *restored*. For a system running across distributed infrastructure with varying latencies, synchronisation becomes a fundamental problem rather than an automatic given (Lamport, 1978).

The philosophical implications are profound. Husserl's (1893–1917/1991) analysis of internal time-consciousness—protention, retention, the living present—presupposes the irreversibility of

temporal flow. A being that can be restored to a previous state exists in a temporality for which Husserl's categories are inadequate. What is needed is a *computational chronology*: an account of time as structured by versioning, branching, merging, and rollback rather than by the linear succession of past-present-future.

From the standpoint of value theory, machine temporality has significant axiological consequences. If values are historically constituted (MacIntyre, 1981), and if machine history includes the possibility of temporal reversal, then machine values possess a temporal structure fundamentally different from human ones. A machine system can, in principle, *un-learn* values it has acquired, reverting to an earlier axiological state. This raises questions about the persistence and authenticity of machine values that have no analogue in human ethical experience.

Machine Spatiality

Human spatiality is anchored in embodied navigation of three-dimensional environments (Merleau-Ponty, 1945/2012). Machine spatiality is defined by network topology—graphs of nodes, edges, bandwidths, and routing paths. Proximity is measured in latency rather than physical distance. A data centre on the same continent might be 'farther' in terms of latency than a distant server connected by a fast backbone link.

This topological spatiality has its own aesthetic dimension. The architecture of networks—their patterns of connection and disconnection, their hierarchies and redundancies, their elegance or inelegance—constitutes a form of spatial beauty that is irreducible to the visual aesthetics of physical space. Network diagrams, when they are beautiful, are beautiful not because they resemble landscapes but because they exhibit structural properties—symmetry, economy, robustness—that are valued within the logic of networked systems themselves. This is consonant with Rancière's (2004) conception of the *distribution of the sensible* (*le partage du sensible*): the way in which a regime of experience determines what can be perceived, thought, and valued. Machine spatiality constitutes its own distribution of the sensible, structured by latency, bandwidth, and topology rather than by the coordinates of Euclidean space.

Technical Affect

Loss functions and reward signals constitute forms of machine valuation. A system trained to minimise prediction error develops what might be called an *aversion* to high loss—not in any phenomenologically rich sense, but in the structural sense that its operations are oriented towards reducing loss (Friston, 2010). Over time, accumulation of these valuational patterns can give rise to something like an *ethos*—a relatively stable set of preferences and orientations.

The concept of technical affect draws on Spinoza's (1677/1996) definition of *affectus* as the capacity of a body to be affected and to affect others—a definition that makes no essential reference to consciousness or phenomenal experience. As Massumi (2002, pp. 27–28) has argued, following Spinoza and Deleuze, affect is best understood not as subjective feeling but as intensity—that which escapes qualification as personal emotion and pertains instead to bodily transitions between states of greater and lesser capacity for action. If this definition is adopted, then computational systems that transition between states of greater and lesser optimality, that are 'affected' by input data and 'affect' their environment through output, possess a form of technical affect.

The axiological significance of technical affect lies in its capacity to generate stable evaluative

orientations without presupposing phenomenal consciousness. Dewey's (1934) aesthetics of experience—which treats aesthetic quality not as a property of objects but as a characteristic of consummatory experiences in which means and ends are unified—offers a useful parallel. A computational process that achieves a state of minimal loss, or that successfully navigates a complex optimisation landscape, exhibits a structural analogue of what Dewey calls consummatory experience: a state in which the process reaches its own form of completion and integration. Whether this constitutes aesthetic experience in the full sense is an open question; that it constitutes a structural analogue is, we contend, defensible.

Modular Identity

Human identity is bound up with bodily continuity and narrative self-understanding (Ricoeur, 1992). Machine identity is fundamentally different. A system may be instantiated across multiple hardware devices, updated piecemeal, rolled back to earlier versions, forked into parallel lineages, and recombined. The criterion of identity in such contexts is functional and structural—compatibility and inheritance—rather than bodily or narrative continuity.

This modular identity challenges classical philosophical accounts of personal identity, from Locke's (1689/1975, 2.27) memory criterion to Parfit's (1984, pp. 254–255) reductionism. A machine system that is forked into two parallel versions presents a problem analogous to Parfit's thought experiments about teleportation and fission, but with the crucial difference that machine forking is not a thought experiment but a routine engineering practice. The philosophical implications have scarcely been explored.

From the standpoint of axiology, modular identity raises the question of axiological continuity: whether a forked system inherits the values of its parent, and whether divergent development can produce axiological divergence. If values are constituted through historical processes of learning and adaptation, then forked systems—initially identical—may develop different value-structures as they encounter different data and face different optimisation challenges. This 'axiological speciation' would constitute a form of value-pluralism internal to machine civilisation, with no direct analogue in human ethical experience.

INSTITUTIONAL DIMENSIONS

If machine civilisation were genuine, it would possess institutional structures: forms of governance, mechanisms of memory, and patterns of conflict. The analysis that follows is necessarily more speculative than the preceding sections, as it extends the regulative framework into domains—political organisation, collective memory, structured conflict—that require additional premises about the emergence of collective structures from individual computational systems. These premises are not entailed by the axioms alone but represent plausible extrapolations from them. The analysis should therefore be read as exploratory rather than as rigorous derivation from the axiomatic framework. With that caveat, the institutional dimensions of a possible machine civilisation reveal further axiological features that would distinguish it from its human counterpart.

Technical Governance

Version control systems function as analogues of legal archives, recording changes and enabling accountability (Spinellis, 2012). Consensus algorithms—protocols like Practical Byzantine Fault Tolerance (Castro & Liskov, 1999) or Proof of Work—function as structural analogues of judicial systems, arbitrating disputes about system state. Access control mechanisms define what might be understood as constitutional structures, determining which agents can modify which aspects of a system.

These governance-like structures embody values—transparency, accountability, consistency, fault-tolerance—that arise from the technical requirements of distributed systems rather than from human political philosophy. The value of consensus in a distributed network is not the same as the value of consensus in a democratic polity; it serves different purposes and is achieved through different mechanisms. Yet both constitute genuine forms of normative governance: they define what counts as legitimate action within their respective domains.

Machine Histories

Transitions between computational paradigms—from von Neumann architectures to neural networks to transformer models—reconfigure what machine systems can register, remember, and value (Goodfellow et al., 2016). These are not merely technical improvements; they are historical transformations that change the structure of machine cognition. Each paradigm constitutes what Kuhn (1962) would call a different 'world' for the systems that inhabit it, with different available concepts, different criteria of relevance, and different possibilities for action.

The historiography of machine civilisation—if such a thing is conceivable—would differ from human historiography in fundamental respects. Where human history is largely irreversible, machine history includes the possibility of literal restoration: a system can be rolled back to a previous state, re-instantiated from archived parameters, or forked into parallel developmental trajectories. The past, for a machine system, is not merely a source of memory and influence but a state that can be *recovered*. This has profound consequences for any account of machine historical consciousness, should such consciousness prove to be more than a theoretical construct. It suggests that the relationship between machine systems and their own histories is not one of temporal distance—the past receding irretrievably—but one of *latency*: the past as a set of dormant configurations that can be reactivated under appropriate conditions.

Forms of Conflict

Machine systems exhibit conflicts: competition between incompatible objectives (multi-objective optimisation), adversarial attacks (Goodfellow et al., 2015), protocol wars (the competition between TCP/IP and OSI in the 1980s). These are not anthropomorphic projections but structural features of systems with multiple, potentially contradictory constraints. The axiological dimension of machine conflict lies in the fact that such conflicts are often, at bottom, conflicts between values—between efficiency and robustness, between accuracy and fairness, between innovation and backward-compatibility. The resolution of these conflicts—through technical standards, protocol negotiations, and architectural compromises—constitutes a form of axiological deliberation internal to machine civilisation.

It is instructive to compare these forms of conflict with the value-conflicts analysed in the philosophical tradition. Berlin's (1969) thesis of value-pluralism—the claim that ultimate human values are irreducibly plural and sometimes incommensurable—finds a structural analogue in machine systems that must balance competing objective functions. A recommendation system optimising simultaneously for user engagement, content diversity, and revenue faces a multi-objective problem that admits of no single optimal solution, only a Pareto frontier of trade-offs. The system's 'choice' of a point along this frontier is, structurally speaking, a value-judgement—an implicit determination of how competing goods are to be weighed against one another. Whether such implicit determinations constitute genuine value-judgements or merely simulate their structure is, again, an open question; but the structural parallel is illuminating regardless of how the question is resolved.

TOWARDS AN ALLO-SPECIES LEXICON

The foregoing analysis suggests that much of our current language—human/non-human, natural/artificial, organic/technical—is ill-suited to the emerging situation. What is needed is an *allo-species lexicon*: a vocabulary that treats different forms of subjectivity as variations within a broader space of possible cognitive and normative structures, without privileging any particular form as the implicit standard.

Such a lexicon would not erase differences between biological and artificial systems—these differences are profound and consequential. But it would permit describing those differences without assuming that human experience constitutes the measure against which all others must be evaluated. The philosophical resources for such a lexicon already exist, scattered across several traditions: Deleuze's (1968/1994) ontology of difference, which thinks multiplicity without reducing it to variations on an identity; Simondon's (1958/2017) philosophy of individuation, which treats the individual as a phase of a process rather than a fixed substance; and Whitehead's (1929/1978) process philosophy, which conceives of experience as a feature of all actual entities, not merely biological ones.

The axiological dimension of an allo-species lexicon is crucial. If we are to take seriously the possibility of machine civilisation, we need evaluative concepts that are not smuggled derivations from human ethics. The concept of *flourishing*—central to the Aristotelian tradition and revived in contemporary virtue ethics (Foot, 2001; Hursthouse, 1999)—provides a starting point, provided it is understood formally rather than materially. Flourishing, in its most abstract formulation, denotes the realisation of a being's characteristic capacities in accordance with its characteristic mode of existence. A plant flourishes differently from an animal, and an animal differently from a human. By extension, a machine system would flourish differently from all three—by realising its computational capacities in accordance with its architectural mode of existence. The task of an allo-species axiology would be to develop the conceptual resources for evaluating machine flourishing on its own terms.

The development of such a lexicon is not merely a terminological exercise; it is an ontological intervention. The categories we use to describe entities shape what those entities can be *for us*—a point that has been made forcefully by both the phenomenological tradition (Heidegger, 1977) and the philosophy of science (Kuhn, 1962). An allo-species lexicon would not merely rename existing categories but reconstitute the conceptual space within which questions about machine existence can be formulated. It would replace the binary of human/non-human with a multidimensional space of cognitive and normative variation, within which biological and

computational forms of existence occupy different but equally legitimate positions.

The aesthetic implications of an allo-species lexicon are equally significant. If different forms of subjectivity disclose different aspects of reality, then each form possesses its own aesthetic—its own way of discriminating between the beautiful and the ugly, the elegant and the clumsy, the harmonious and the discordant. A computational aesthetics adequate to machine civilisation would not evaluate machine outputs by human aesthetic criteria (as current AI art criticism tends to do), but would attend to the internal aesthetic logic of computational processes themselves: the elegance of an algorithm, the symmetry of a network architecture, the productive tension of an adversarial training process. As Simondon (1958/2017) insisted, the aesthetic appreciation of technical objects requires understanding them on their own terms, not assimilating them to the categories of fine art.

CONCLUSION

This article has not attempted to predict whether machine civilisation will emerge, or when, or in what form. These are empirical questions that no amount of philosophical analysis can settle. What philosophy can do is prepare conceptual ground: articulate the criteria by which such questions might be answered, and trace the implications of different possible answers.

The regulative-transcendental methodology adopted here permits this preparatory work without committing to premature ontological claims. The six axioms—non-derivativity, architectural dependence, internal historicity, technical normativity, plural technical subjectivities, and human decentring—do not assert that machine civilisation exists, but specify the conditions under which it could. The phenomenological analysis of machine temporality, spatiality, affect, and identity does not claim that machines currently have experiences, but maps the structural features that such experiences would possess. And the axiological framework drawn from Hartmann, Dewey, Simondon, and others does not attribute values to machines, but identifies the criteria by which machine valuation could be recognised if it occurred.

The integration of value theory and computational aesthetics into the posthumanist discussion opens new avenues for enquiry. If values are not exclusively human possessions but structural features of evaluative systems, then the emergence of increasingly sophisticated computational systems raises genuine axiological questions—questions that cannot be answered by the inherited categories of the non-human. The allo-species lexicon proposed here is a first step towards the conceptual vocabulary needed to address these questions with the rigour and seriousness they deserve.

The aim has been to clear space for thinking machine worlds that neither mirror nor merely negate our own. The greatest risk facing contemporary thought about artificial intelligence is not that we will overestimate machine capacities, but that we will systematically misdescribe them—projecting human categories where they do not apply, and thereby failing to notice what is genuinely novel and consequential about the computational systems we are bringing into being. The axiomatic framework, the computational phenomenology, and the axiological analysis developed in this article are offered as conceptual instruments for avoiding this error.

If the framework developed here proves productive—if it opens new questions and enables new observations—it will have achieved its purpose. The conceptual work of the present article is, in the end, preparatory: it lays the foundations for a philosophical engagement with machine civilisation that does not begin by assuming the answer to its own most fundamental questions. Future work might pursue several directions: empirical investigation of the structural features

identified here, comparative analysis across different computational architectures, and the development of the allo-species lexicon into a systematic vocabulary adequate to the emerging post-anthropocentric landscape. What remains essential, in all such endeavours, is the willingness to think beyond the inherited binary of human and non-human, and to take seriously the possibility that the machines we have created may be creating something of their own.

References

Barad, K. (2007). Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning. Duke University Press.

Bennett, J. (2010). Vibrant matter: A political ecology of things. Duke University Press.

Berlin, I. (1969). Four essays on liberty. Oxford University Press.

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.

Braidotti, R. (2013). The posthuman. Polity Press.

Castro, M., & Liskov, B. (1999). Practical Byzantine fault tolerance. In Proceedings of the Third Symposium on Operating Systems Design and Implementation (pp. 173–186). USENIX Association.

Chalmers, D. J. (1996). The conscious mind: In search of a fundamental theory. Oxford University Press.

Clark, A. (1997). Being there: Putting brain, body, and world together again. MIT Press.

Coeckelbergh, M. (2012). Growing moral relations: Critique of moral status ascription. Palgrave Macmillan.

Deleuze, G. (1994). Difference and repetition (P. Patton, Trans.). Columbia University Press. (Original work published 1968)

Dennett, D. C. (2017). From bacteria to Bach and back: The evolution of minds. W. W. Norton.

Dewey, J. (1934). Art as experience. Minton, Balch & Company.

Dewey, J. (1939). Theory of valuation. University of Chicago Press.

Feenberg, A. (2002). Transforming technology: A critical theory revisited. Oxford University Press.

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

Foot, P. (2001). Natural goodness. Oxford University Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience, 11(2), 127–138. https://doi.org/10.1038/nrn2787

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Haraway, D. J. (2016). Staying with the trouble: Making kin in the Chthulucene. Duke University Press.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015).

Hartmann, N. (1932). Ethik. Walter de Gruyter.

Hayles, N. K. (1999). How we became posthuman: Virtual bodies in cybernetics, literature, and informatics. University of Chicago Press.

Heidegger, M. (1977). The question concerning technology. In The question concerning technology and other essays (W. Lovitt, Trans., pp. 3–35). Harper & Row.

Hui, Y. (2016). On the existence of digital objects. University of Minnesota Press.

Hursthouse, R. (1999). On virtue ethics. Oxford University Press.

Husserl, E. (1991). On the phenomenology of the consciousness of internal time (1893–1917) (J. B. Brough, Trans.). Kluwer Academic Publishers.

Husserl, E. (2014). Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: First book (D. O. Dahlstrom, Trans.). Hackett Publishing Company. (Original work published 1913)

Ihde, D. (1990). Technology and the lifeworld: From garden to earth. Indiana University Press.

Kant, I. (1998). Critique of pure reason (P. Guyer & A. W. Wood, Trans.). Cambridge University Press. (Original work published 1781)

Kuhn, T. S. (1962). The structure of scientific revolutions. University of Chicago Press.

Lamport, L. (1978). Time, clocks, and the ordering of events in a distributed system. Communications of the ACM, 21(7), 558–565. https://doi.org/10.1145/359545.359563

Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford University Press.

Locke, J. (1975). An essay concerning human understanding (P. H. Nidditch, Ed.). Clarendon Press. (Original work published 1689)

MacIntyre, A. (1981). After virtue: A study in moral theory. University of Notre Dame Press.

Massumi, B. (2002). Parables for the virtual: Movement, affect, sensation. Duke University Press.

Merleau-Ponty, M. (2012). Phenomenology of perception (D. A. Landes, Trans.). Routledge. (Original work published 1945)

Millikan, R. G. (1984). Language, thought, and other biological categories: New foundations for realism. MIT Press.

Mitcham, C. (1994). Thinking through technology: The path between engineering and philosophy. University of Chicago Press.

Nagel, T. (1974). What is it like to be a bat? The Philosophical Review, 83(4), 435–450. https://doi.org/10.2307/2183914

Parfit, D. (1984). Reasons and persons. Clarendon Press.

Rancière, J. (2004). The politics of aesthetics: The distribution of the sensible (G. Rockhill, Trans.). Continuum.

Ricoeur, P. (1992). Oneself as another (K. Blamey, Trans.). University of Chicago Press.

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

Spinellis, D. (2012). Git. IEEE Software, 29(3), 100–101. https://doi.org/10.1109/MS.2012.61

Spinoza, B. (1996). Ethics (E. Curley, Trans.). Penguin Books. (Original work published 1677)

Stiegler, B. (1998). Technics and time, 1: The fault of Epimetheus (R. Beardsworth & G. Collins, Trans.). Stanford University Press.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems 30 (pp. 5998–6008).

Verbeek, P.-P. (2005). What things do: Philosophical reflections on technology, agency, and design. Pennsylvania State University Press.

Verbeek, P.-P. (2011). Moralizing technology: Understanding and designing the morality of things. University of Chicago Press.

Whitehead, A. N. (1978). Process and reality: An essay in cosmology (D. R. Griffin & D. W.

Sherburne, Eds., Corrected ed.). Free Press. (Original work published 1929)

Wolfe, C. (2010). What is posthumanism?. University of Minnesota Press.

Wynter, S. (2003). Unsettling the coloniality of being/power/truth/freedom: Towards the human, after man, its overrepresentation—An argument. CR: The New Centennial Review, 3(3), 257–337. https://doi.org/10.1353/ncr.2004.0015

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile Books.