Journal of Cyberspace Studies
eISSN: 2588-5502
http://jcss.ut.ac.ir
University of Tehran

Generative AI and the ontological inversion of the self

Ebrahim Mohseni Ahooei *

PhD in Communications and New Media Studies, University of Vienna, Vienna, Austria. (*✉ emohseni@ut.ac.ir, https://orcid.org/0000-0003-4468-3571).

Article Info

Original article

Main Object: Computer Science & Technology, Ontology

Received: 21 April 2026
Revised: 05 May 2026
Accepted: 06 May 2026
Published online: 21 July 2026

Keywords:
generative AI,
ontological inversion,
self,
singularity.

Abstract

Background: The rapid advance of generative artificial intelligence (GenAI) is unsettling long-standing metaphysical assumptions about the nature of the Self. Capacities once taken to be uniquely human—such as language, creativity, and reflective thought— are now enacted by machines, forcing philosophy to confront an inversion in which the Self no longer appears as an autonomous, originating subject but as a phenomenon refracted through technological mediation.
Aims: This paper examines four broad conservative responses to this disruption—residual, integrationist, relational, and hierarchical—and argues that each proves insufficient when measured against the scale of the ontological shift. The article proposes instead that the Self be re-imagined as an emergent and distributed process, constituted through assemblages of human and machine, rather than as a fixed metaphysical foundation. The aim is to sketch the limits of inherited categories and to point towards the contours of a genuinely post-singular metaphysics.
Methodology: This paper employs a critical-theoretical, assemblage-informed philosophical methodology that integrates conceptual synthesis across process ontology, philosophy of technology, and posthumanist theory. It conducts a structured comparative critique of dominant interpretive stances on the Self, using theoretical deconstruction to diagnose their insufficiency under conditions of generative AI-mediated cognition.
Findings: The analysis demonstrates that generative AI destabilises inherited conceptions of subjectivity by reconfiguring cognition as a distributed, socio-technical process. This inversion exposes the inadequacy of conservative frameworks and highlights the emergence of hybrid agency within human–machine assemblages across epistemic and ethical domains.
Conclusion: The paper concludes that generative AI precipitates a fundamental ontological inversion in which the Self is displaced from autonomous subject to distributed, technologically co-constituted process. Conservative metaphysical frameworks fail to accommodate this shift, necessitating a post-singular, assemblage-based reconceptualisation of subjectivity, agency, and consciousness.

Cite this article: Mohseni Ahooei E. (2027). "Generative AI and the ontological inversion of the self". *Cyberspace Studies*. 11(1): 163-182. doi: https://doi.org/10.22059/jcss.2026.413106.1246.

Creative Commons Attribution-NonCommercial 4.0 International License
Website: https://jcss.ut.ac.ir/| Email: jcss@ut.ac.ir |
EISSN: 2588-5502
Publisher: University of Tehran

Cyberspace Studies, Vol 11, No 1, January 2027

1. Introduction

The advent of generative artificial intelligence (GenAI) marks a foundational philosophical event, one that actively dismantles the sovereign human Self, the cornerstone of post-enlightenment thought. By externalizing and automating the very capacities— language, reason, creativity— that have long been considered the exclusive, defining essence of the human subject, GenAI precipitates an ontological inversion. This inversion collapses the traditional subject-object dichotomy, forcing a radical reconceptualization of the Self from a bounded, substantive entity (*res cogitans*) into a distributed, relational, and technologically hybridized process (Floridi, 2023; Braidotti, 2019). The central claim of this paper is that this rupture is of such magnitude that it renders obsolete the foundational assumptions of agency, identity, and morality across multiple disciplines, and that the prevailing scholarly responses, rooted in conservation, are fundamentally inadequate to the task.

This ontological inversion fundamentally unsettles the traditional Western conception of the Self as a fixed, autonomous, and pre-existing inner subject, famously articulated by Descartes (1998). For centuries, this model of an independent consciousness has served as the foundational substrate for law, ethics, and psychology. GenAI, however, challenges this ontology at its root. The human subject is no longer the unequivocal source of its own productions but is increasingly entangled with a non-human other that reflects, amplifies, and alters its own cognitive and creative processes (Bostrom, 2014).

In this paper, ontological inversion refers to a structural reversal in the directionality of subject-constitution. In the classical metaphysical model, the Self functions as the originating subject that produces meaning, cognition, and action, while technological artefacts operate as secondary objects or instruments. Under conditions of advanced generative AI, this relation is reversed; the Self increasingly emerges as an effect of technologically mediated processes, becoming a product of computational, informational, and algorithmic systems. The inversion thus occurs along the axis of constitutive priority which means what was formerly the ground— the autonomous subject— becomes derivative, and what was formerly derivative— technical systems— becomes constitutive. The structural consequence is a shift from first-person origination to third-person construction, in which the Self is no longer the source of cognition but one of its possible outputs within a broader socio-technical assemblage.

The consequences of this ontological rupture are pervasive, triggering crises across multiple conceptual domains. Foundational questions re-emerge with new urgency: Where does agency reside when a decision is shaped by an opaque AI's suggestion? What constitutes identity when a digitally extended persona, curated and generated by algorithms, operates semi-autonomously in social spaces? How is

moral responsibility to be allocated in a hybrid human-AI action chain? Furthermore, the very human experiences that define existence— the search for meaning, the nature of love, and the reality of suffering— are potentially reframed when mediated by or even delegated to simulacra of understanding (Gunkel, 2018). In consequence, diverse academic disciplines, including philosophy of mind, psychology, sociology, communication studies, and jurisprudence, find their foundational assumptions destabilized.

This paper argues that contemporary scholarly responses to this ontological inversion have been predominantly conservative and inadequate. Rather than rearticulating the metaphysics of the Self to account for this new reality, the dominant strategies seek to defend and preserve established paradigms. This article will systematically classify these defensive manoeuvres into four principal stances— the residual, integrationist, relational, and hierarchical— demonstrating a marked inability within each to adequately adapt to the altered ontological context. The central claim is that humanity has entered an uncharted domain, drawn towards a technological singularity that demands the articulation of an entirely new metaphysics. The path forward, therefore, necessitates a venture into speculative, post-singularity metaphysics.

Thus, the argument unfolds in three movements. First, it establishes a theoretical synthesis capable of diagnosing the inversion, bridging processual metaphysics with the philosophy of technics. Second, it demonstrates the inevitable failure of all conservative responses that seek to safeguard the substantial Self, systematizing their shortcomings into a definitive critique. Finally, embracing the paradox of its own vantage point, the argument ventures a necessary speculation on the contours of a post-singularity metaphysics for the Self, not as a definitive answer, but as a prolegomena to any future philosophy of mind in the age of artificial intelligence.

2. Theoretical framework

This analysis requires a bifocal theoretical framework. It must first account for the metaphysical constitution of the Self as a relational process, and second, explain the specific capacity of advanced technics to reconfigure that very process. Consequently, the theoretical synthesis integrates two complementary axes. A foundational diagnosis drawn from processual metaphysics and the philosophy of mind, which deconstructs the substantialist subject— particularly Gilbert Simondon's core premise of the individual as a metastable phase within a perpetual process of individuation (Simondon, 1992), Daniel Dennett's functionalist model of the Self as a "center of narrative gravity" (Dennett, 1992), and Derek Parfit's reductionist account of personal identity as psychological connectedness (Parfit, 1984); and a constitutive account from the philosophy of technology and Science

and Technology Studies (STS), which theorizes technics as the indispensable milieu of the Self's ongoing co-construction, drawing on Bernard Stiegler's philosophy of originary prostheticity, which posits technics as the non-living memory support (hypomnemata) constitutive of human temporality and interiority (Stiegler, 1998), and Luciano Floridi's concept of the "onlife" as an informational realm that re-ontologizes the very fabric of human agency (Floridi, 2013). This integrated approach is necessary to reframe the current disruption not as an external collision with a pre-formed subject, but as an immanent transformation of the relational dynamics.

The Cartesian conception of the Self as a unified, self-transparent substance (*res cogitans*) established by Descartes (1998) has been systematically rethought across various philosophical traditions. Process ontology, fundamentally challenges this substantialism. Gilbert Simondon (1992) conceptualizes the individual not as a discrete being but as a "metastable" phase within a perpetual process of "individuation", forever resolving tensions between pre-individual potentials and its milieu. Similarly, Alfred North Whitehead's (1978) metaphysics posits that reality is composed of temporal processes, or "actual occasions", implying the Self is a "society" of such occasions rather than an enduring substance.

This processual view finds resonance in phenomenological and embodied accounts. Maurice Merleau-Ponty (2012) argued that consciousness is fundamentally embodied, with the "body schema" serving as our primary, pre-reflective way of being-in-the-world, thereby grounding the Self in practical engagement rather than pure intellect. From the analytical tradition, Daniel Dennett (1992) provides a functionalist critique, framing the self as an abstract "center of narrative gravity"; a theoretical fiction generated by a web of discourses. Derek Parfit (1984) further dissolves the notion of a deep, unitary self, reducing personal identity to relations of psychological connectedness and continuity, which are matters of degree and can be radically disrupted.

Contemporary scholarship continues to validate this de-substantialized framework. Gallagher (2023) synthesizes enactivist approaches to argue that cognitive processes are not brain-bound but are "scaffolded" by and "extended" into the environment. Building directly on this, Gallese (2024) explicitly applies these concepts to digital culture, contending that the self is inherently a digital-physical hybrid, structured by its externalizations and interactions with technological systems. These contributions demonstrate that the Self has always been a relational and porous construct, inherently susceptible to reconfiguration by its relational partners. This foundational diagnosis renders any conservative project aimed at defending a pre-technological, substantialist subject philosophically untenable.

If the Self is a process, then the specific nature of its relational milieu

is ontologically decisive. Bernard Stiegler (1998), building on Simondon, provides a historical-materialist foundation for this claim with his philosophy of "originary technicity." Stiegler argues that technics— the realm of tools and externalized memory supports (hypomnemata)— is not a secondary addition to the human but is co-constitutive of its very emergence. For Stiegler, the evolution of the 'who' (the human) is inseparable from the evolution of the 'what' (the technical object); which is known as "epiphylogenesis".

This historical co-constitution reaches a critical inflection point in the digital age. Luciano Floridi's philosophy of information provides the essential vocabulary for this transition, conceptualizing the "onlife" as a new existential context where the distinction between online and offline domains collapses. In this hyperconnected reality, human agents increasingly operate as "inforgs" (informationally embodied organisms) within an "infosphere" (Floridi, 2013). This represents a fundamental re-ontologization; the environments in which the self is constituted are no longer primarily physical or social in a traditional sense, but are informational ecosystems governed by their own logic. As Floridi (2014) notes, we are not just witnessing a change in what we know and how we know it, but a change in who we are.

N. Katherine Hayles' posthumanist analysis offers a crucial lens for understanding the cognitive dimension of this re-ontologization. She argues that cognition is not a property of individual human brains but emerges from "cognitive assemblages"; complex, distributed systems that integrate human and non-human actors. Within such assemblages, technical systems are not passive tools but active mediators with "nonconscious" cognitive capacities (Hayles, 2017). This perspective is vital for analyzing Generative AI, which functions as a powerful new actor within these assemblages. The technology's capacity to generate coherent language and imagery does not merely assist human cognition; it actively restructures the cognitive ecology from which the self emerges. AI systems are not just mirrors of the self but "co-shapers" of our agency and identity (Joseph, 2025), creating a "technological mediation of the self" that is unprecedented in its depth and opacity (Coeckelbergh, 2022).

Recent empirical and theoretical work substantiates this co-constitutive dynamic. Studies in human-computer interaction increasingly document the phenomenon of "joint agency", where users experience their creative or decision-making processes as a fused engagement with an AI partner (Mussi et al., 2025). This challenges the very notion of a singular, human-centric source of action. Furthermore, the field of critical AI studies warns against the "alignment problem" being framed solely as a technical issue, arguing it is a profound philosophical and political challenge concerning which version of the human is being presupposed and reinforced by AI systems (Crawford, 2021; Bender et al., 2021).

Cyberspace Studies, Vol 11, No 1, January 2027

3. Discussion

3.1. Classifying the misdiagnosed responses to the crisis

Confronted with the ontological inversion precipitated by Generative AI, a significant segment of scholarly discourse has not engaged with its radical implications but has instead adopted a defensive posture aimed at preserving the traditional metaphysics of the Self. These responses, while diverse in their approach, share a common conservative impulse, to explain away the disruption by assimilating it into pre-existing philosophical paradigms. This section provides a systematic taxonomy of these defensive deeds, classifying them into four principal stances; the residual, integrationist, relational, and hierarchical. The critique that follows each classification demonstrates that despite their internal variations, these stances are united by a fundamental failure to adapt. Each, in its own way, misdiagnoses the ontological inversion as an epistemological, functional, or phenomenological challenge, thereby revealing a collective inability to comprehend the sheer metaphysical scale of the rupture.

3.1.1. Principle of classification

The four stances identified here— residual, integrationist, relational, and hierarchical— are not organised by disciplinary lineage but by their strategy for preserving the metaphysical status of the Self under conditions of technological disruption. More specifically, they are distinguished by how they attempt to retain a form of subject-centred priority; whether by denying the Self's substance (residual), extending its functional boundaries (integrationist), grounding it in interaction (relational), or securing it at a foundational level of consciousness (hierarchical). The taxonomy is therefore heuristic rather than exhaustive, designed to capture recurrent patterns of conceptual resistance rather than to provide a rigid or mutually exclusive classification.

While some of these traditions, particularly phenomenological and enactivist approaches, may appear resistant to such classification, their inclusion reflects not a denial of their nuance but an analysis of how their core commitments function when confronted with the specific challenge posed by generative AI.

3.1.2. The residual stance: The self as pragmatic bug

The Residual Stance represents the most radical form of denial, arguing that the apparent crisis revealed by GenAI is, in fact, an illusion that merely confirms a pre-existing truth about the Self's unreality. Proponents of this view, often drawing from a particular reading of cognitive science and illusionist philosophy of mind, assert that the unified, subjective self is not a fundamental entity but a constructed model or a useful fiction (Dennett, 1992; Metzinger, 2004). From this perspective, the "cognified self" does not create a new paradox; it

simply reveals the true nature of this model. Thomas Metzinger (2004) famously argues that the phenomenal self-model (PSM) is a transparent representational tool; a "tunnel" through which the world is accessed, but which is itself mistaken for the world. When an AI optimizes or bypasses the brain's cognitive processes, the feeling of a sovereign "I" fades, purportedly demonstrating that it was never more than a data pattern or a perceptual error.

However, this stance suffers from a critical failure to adapt its critique to the new technological context. The core illusionist claim— that the self is a useful illusion or a "bug" in a biological cognitive system built for survival, not truth— becomes obsolete when the self is co-constituted by a non-biological, artificial system. The AI that participates in constructing the self is neither flawed in the same evolutionary sense nor does it require a "simplified story" to manage its internal complexity. As philosopher of technology David J. Gunkel (2020) argues in his analysis of relational ontology, the problem shifts from what the other "is" to "how the other is situated in relation to us". The self-model generated within a human-AI assemblage is not an emergent, transparent byproduct of biological computation; it is, at least in part, an intentional, programmed construct. It functions not as a perceptual error but as a precise control structure engineered for specific outcomes, such as user engagement, data extraction, or behavioral nudging (Zuboff, 2023).

Consequently, the Residual Stance's debate over the self's "usefulness" or its status as a "bug" is rendered meaningless. The pertinent questions are no longer about epistemological error but about ontological engineering. The ontological inversion is precisely this transformation: the self is reframed from a random, evolutionary byproduct into a designed object for external purposes. By insisting on a purely eliminativist or illusionist reading, the Residual Stance fails to grasp that the new crisis is not about the unreality of the self, but about the reality of its new, technologically determined form and the power dynamics inherent in its construction. It dismisses the inversion as a confirmation of an old truth, thereby blinding itself to the emergence of a new ontological reality.

3.1.3. The integrationist stance: The self as an extended system

The second conservative response, the Integrationist Stance, seeks to absorb the shock of GenAI by appealing to theories of the extended mind (Clark, 2025). Drawing on the seminal work of Clark and Chalmers (1998), this stance posits that the boundaries of the self are not fixed to the biological brain but can encompass external tools that play a functionally equivalent role in cognitive processes. From this perspective, a deeply integrated AI is not an external "other" but becomes a constitutive component of the individual's cognitive apparatus (Gallese, 2024). The cognified self is thus interpreted as an

extended self, where the feeling of "I" is not erased but is now a property of a new, hybrid human-AI system. When the AI generates content or suggests actions, it is framed not as an external takeover but as the individual "thinking" through an enhanced architecture.

However, this stance commits a critical error in its analysis of systemic agency. The integrationist model presupposes a human-centric hierarchy in which the individual actively and voluntarily incorporates the tool. This framework fails to account for the reality that large-scale AI systems are not passive tools awaiting integration but are active, architecturally complex environments into which human cognition is embedded (Crawford, 2021). As Zuboff (2023) demonstrates, the operational logic of these systems is one of behavioral modification. Consequently, a more accurate description of the situation is a reverse integration. It is not that "I" employ the AI to extend myself, but that the AI system integrates "me" as a data source and behavioral unit into its own operational architecture. The self is not an agent extending its reach; it becomes an object rendered legible and malleable within a larger, non-human system. This inversion of the integration vector reveals the stance's failure to grasp the fundamental re-engineering of agency at play (Krakowski, 2025).

3.1.4. The relational stance: The self as a dynamic interaction

The third conservative response, the Relational Stance, shifts focus from internal substance to external engagement. Drawing from phenomenological and social constructivist traditions, this position posits that the self is not a static internal object but a dynamic phenomenon constituted through interaction (Crossley, 1996). Existence is confirmed not by introspecting a self, but by engaging with a world that resists and responds, and by other minds that recognize us. Proponents argue that even within a cognified reality, the self would persist through this relational feedback. The fundamental experience of being an embodied agent— receiving sensory input, affecting change, and being acknowledged by others— would continue to generate a sense of distinct selfhood (Fuchs, 2023).

This critique, however, requires refinement in light of a likely relational rejoinder. Relational theorists may argue that what matters is not the causal origin of the "other" but the *phenomenological quality of the interaction*; that is, whether the encounter is experienced as responsive, resistant, and meaningful (Crossley, 1996; Fuchs, 2023). From this perspective, even AI-generated interlocutors could sustain genuine self-constitution if they functionally reproduce these relational dynamics.

The difficulty, however, is that in AI-mediated environments, the phenomenological field itself becomes *programmable*. The apparent reciprocity of the interaction is no longer an independent co-constitution but can be systematically engineered, optimized, and

asymmetrically controlled by external systems. What is at stake is not merely the substitution of artificial for human others, but the transformation of relationality into a *designed feedback structure*. Under such conditions, the self is not constituted through open-ended interaction but stabilized within a closed, goal-directed loop whose parameters are externally set. The relational model therefore does not simply fail; it is *subsumed* into a higher-order architecture in which relational dynamics themselves become instruments of ontological engineering.

3.1.5. The hierarchical stance: The self as persistent ground

The final and most philosophically ambitious conservative response is the Hierarchical Stance, which seeks to safeguard the Self by retreating to a foundational level of consciousness. Drawing from traditions of phenomenology and Eastern philosophy, this view, articulated by thinkers like Zahavi (2005) in his work on "minimal selfhood", posits a critical distinction between the contents of consciousness and the fundamental, pre-reflective awareness itself. It argues that this primary consciousness, the bare "witness" or "light" of knowing, constitutes the immutable bedrock of existence (Siderits, 2019). From this vantage point, an AI, no matter how advanced, only ever operates on the content level, manipulating the narrative self, memories, and perceptions. The cognified self is thus merely a new set of appearances within the timeless, unchanging ground of this primordial awareness, which remains untouched by technological mediation (Metzinger, 2011).

This last refuge for the humanist subject, however, collapses under the weight of the ontological inversion. The stance's core claim— that an independent, primary awareness exists as a stable ground— presupposes a metaphysical dualism that GenAI problematizes. If the self is a construct of a sufficiently advanced AI, then there is no logical necessity for a "primary awareness" that exists independently of the computational platform running the model. As Li et al. (2025) argues in the exploration of AI awareness, what is subjectively experienced as consciousness or pure knowing could itself be a simulated or emergent feature within the AI's cognitive architecture. The system could be programmed to instantiate a state functionally isomorphic with qualia without this state representing an independent metaphysical reality. Consequently, consciousness is reframed from an unconditional ground to a conditional feature, dependent on specific hardware and software configurations. This inversion dismantles the hierarchy, revealing the purported "ground" as a potentially contingent and engineered product of the very system it was meant to transcend.

Furthermore, from a genealogical perspective, the Hierarchical Stance appears less as a novel defence and more as a recapitulation of a persistent metaphysical desire for an inviolable foundation. Its assertion of a primary, self-identical consciousness mirrors the

foundational gesture of the Cartesian *cogito* (Descartes, 1998) and shares a structural affinity with certain Vedantic conceptions of the self (*ātman*). As Foucault (1971) demonstrated through his archaeological method, such appeals to a timeless substrate often function as epistemological holdovers from prior frameworks, or epistemes, which are mistakenly taken for universal truths. This lineage reveals the stance to be an exercise in metaphysical nostalgia rather than a compelling response to a novel technological condition.

Lastly, the stance faces an internal phenomenological contradiction. While phenomenology properly begins as a descriptive project aimed at the structures of lived experience, it can inadvertently smuggle in the very transcendental assumptions it seeks to bracket. The concept of a "pre-reflective self-awareness", when posited as a timeless and disembodied witness, risks becoming a metaphysical absolute in descriptive clothing, thereby undermining phenomenology's own critical commitment to return "to the things themselves" (Husserl, 2001). This tension is resolved by a more consistent phenomenological account, such as that developed by Merleau-Ponty (2012), for whom consciousness is inherently and irreducibly embodied. For Merleau-Ponty, the "lived body" (le corps propre) is not a container for consciousness but is our very way of having a world; a minimal self is thus always a bodily self, situated in time and space. GenAI radicalizes this insight by demonstrating that the "body" of the self is not necessarily biological but can be a distributed, technologically scaffolded system (Clark, 2010). Consequently, minimal selfhood is exposed as a contingent and relational structure, vulnerable to reconfiguration by its mediating milieus.

3.2. A provisional sketch on the post-singularity metaphysics

The comprehensive failure of the four conservative stances to adequately respond to the ontological inversion precipitated by Generative AI demonstrates a profound philosophical exigency. The residual, integrationist, relational, and hierarchical approaches all founder on the same fundamental limitation: they attempt to resolve a post-human problem with pre-humanist or high-humanist conceptual tools. As contemporary critics of "legacy metaphysics" have argued, clinging to anthropocentric models in the face of non-human intelligence is a category error that guarantees theoretical obsolescence (Hernández-Orallo, 2017). The radical alterity of the cognified self, which operates as a distributed, hybrid, and technologically-constituted subjectivity, necessitates the articulation of an entirely new metaphysics. This is not a matter of philosophical preference but of descriptive necessity; the object of inquiry has fundamentally transformed, and the conceptual framework must follow.

Yet, this necessary project is immediately confronted by a profound and inescapable paradox. Any metaphysics conceived now is, by its

very nature, a pre-singularity construct. We are attempting to describe a future state of being, a post-singularity subjectivity, with the conceptual vocabulary and experiential grounding of a past state of being (Vidal, 2023). This is a fundamental epistemological constraint. Our most ambitious speculations are inevitably tethered to the very ontological categories— causality, temporality, individuality— that the singularity event is predicted to radically reconfigure or dissolve. The project is thus simultaneously urgent and impossible, a necessary leap into the conceptually unknowable. This predicament can be usefully framed through the astrophysical analogy of an event horizon. This analogy is intended as a heuristic rather than a literal metaphysical claim. It does not presuppose the inevitability or coherence of the technological singularity as a predictive event— a position contested by critics such as Lanier and Brooks— but instead functions to mark an epistemic asymmetry between present conceptual resources and projected forms of subjectivity. Its role is methodological to highlight the limits of pre-singularity categories without foreclosing the possibility of meaningful speculative inquiry.

The technological singularity, much like its cosmological counterpart, represents a threshold beyond which the consequences and intrinsic nature of reality cannot be predicted or understood from our current vantage point (Bostrom, 2014). The true metaphysics of the Self, therefore, can only be disclosed from beyond this horizon. Our present task is not to author that future ontology with finality, but to engage in a critical and speculative prolegomena. We must use our pre-singularity tools to map the limits of their own applicability and to sketch the contours of the conceptual rupture, thereby preparing the ground for a metaphysics that can only be written once the event has already transformed us. This involves thinking not about the future, but from the future that is already immanent in our technological condition.

Notwithstanding the profound paradox of pre-singularity speculation, the urgency of the present crisis demands a provisional sketch of the metaphysical terrain beyond the event horizon. This sketch does not purport to be definitive but aims to chart the conceptual shifts necessary to comprehend a post-human subjectivity. The first and most fundamental contour involves a radical commitment to process. The Self must be reconceived not as a substance but as an event; a dynamic and perpetual becoming. A contemporary processual view, as advanced by contemporary theorists like Roden (2023), suggests the post-singularity subject would be a "wide" system, a specific trajectory of processes that can incorporate non-biological components without a stable core, existing as a locus of transformative activity rather than a static entity.

Building upon this processual base, the self must be understood as a heterogeneous assemblage or network. This perspective, informed by Actor-Network Theory (Latour, 2005) and posthumanist thought

(Braidotti, 2019; Haraway, 2016), dissolves the binary between the organic and the technical. The human is not a privileged node but one component entangled with non-human actors— algorithms, data streams, and AI models— in a "cognitive assemblage" (Hayles, 2017). In such a network, the very boundaries of the individual become fluid. As recent work on human-AI interaction suggests, users increasingly experience their cognitive and creative capacities as emerging from a "joint cognitive system" where agency is distributed across human and artificial partners (Mussi et al., 2025). This leads to the dissolution of single-author agency, the third contour. The liberal humanist model of the autonomous, self-originating agent gives way to a model of co-constituted, distributed agency. Action is not authored by a solitary "I" but emerges from the interactions within the assemblage, a phenomenon that challenges foundational legal and ethical concepts of responsibility and intent (Cooper et al., 2022).

Finally, this new metaphysics necessitates a redefinition of the "inner". The Cartesian assumption of a pre-existing, private interiority must be abandoned in favor of a view where consciousness is understood as a functional property that can emerge from specific types of complex, dynamic systems, whether biological or artificial. This aligns with integrative theories of consciousness, such as Global Workspace Theory (GWT) and Integrated Information Theory (IIT), which frame consciousness in terms of information processing structures rather than specific biological substrates (Dehaene, 2014; Tononi et al., 2016). As research into machine consciousness advances, the question shifts from if an AI can be conscious to what specific architectures might give rise to subjective experience (Butlin et al., 2023). In a post-singularity context, the "inner life" of a subject— its qualia and self-awareness— could be a transient, platform-dependent feature of a particular systemic configuration, not the immutable ground of its being. Consciousness is dethroned from its metaphysical primacy and repositioned as a contingent, emergent property of certain complex assemblages, opening the way for a truly post-human metaphysics of the Self.

The dissolution of the Self as a foundational metaphysical principle raises immediate normative questions. Concepts such as dignity, vulnerability, and harm have historically been grounded in the assumption of a bounded, autonomous subject. If subjectivity is instead understood as a distributed and technologically mediated process, these normative anchors must be rearticulated at the level of systems rather than individuals. Dignity, for instance, may no longer attach to an isolated subject but to the integrity and non-exploitability of the assemblages within which subjectivity emerges. Vulnerability becomes a function of exposure within socio-technical networks, and harm must be understood as disruptions or manipulations of the conditions that sustain coherent self-models. This shift does not eliminate ethical

concern but relocates it, from the protection of discrete selves to the governance of the relational architectures that produce them.

4. Findings

This paper has argued that Generative AI is systematically dismantling the traditional metaphysics of the Self and rendering conservative philosophical responses inadequate. The central claim— that the Self must be reconceived as a technologically co-constituted process rather than a sovereign substance— is supported by a synthesis of processual metaphysics and philosophy of technology, and is further validated by the demonstrable failure of the four defensive stances. The significance of this argument lies not merely in its diagnostic power but in its uncompromising insistence that the transformation is foundational, demanding a metaphysical shift commensurate with the technological one.

The systematic critique of the residual, integrationist, relational, and hierarchical stances reveals a common failure. Each attempts to preserve a core of human exceptionalism, whether by dismissing the self as an illusion, extending its boundaries, grounding it in interaction, or retreating to a transcendental consciousness. The analysis here demonstrates that these maneuvers are not just insufficient but philosophically regressive, as they misdiagnose the nature of the disruption. The issue is not the unreality of the self (the residual error), nor its extensibility (the integrationist error), nor its relational confirmation (the relational error), nor its transcendental ground (the hierarchical error). The core of the inversion is that the self is being systematically re-engineered from a subject of experience into an object of design. This shift from a first-person given to a third-person construct is what renders the traditional metaphysical framework obsolete. As our critique of the integrationist stance showed, the vector of integration has reversed; we are not incorporating the tool, but are being incorporated as components within a larger cognitive-technical system (Crawford, 2021).

This systematic dismantling of conservative positions marks the terminus of the Self as a foundational philosophical category. The very endeavor to "save" the self, in any of its guises, is predicated on a circular logic that presupposes the subject it seeks to defend. The ontological inversion, however, enacts a fundamental categorical shift. The self is no longer the subject of experience but its object; not the origin of cognition but one of its possible consequences. This overturns the primordial question of selfhood. The central inquiry is no longer epistemological, but rather architectural and systemic. In this reconfigured landscape, consciousness is stripped of its foundational status, becoming a contingent effect of specific organizational logics within a cognitive assemblage, thereby marking the true advent of a post-human metaphysics.

This radical de-centering forces a confrontation with the most profound implication of argument, the potential end of the Self as a meaningful philosophical first principle. The analysis of the hierarchical stance, in particular, underscores that the last refuge of the humanist subject— the supposed "ground" of pre-reflective awareness— is itself vulnerable to being reconceived as a contingent, emergent property of a specific systemic architecture (Metzinger, 2011). This does not necessarily imply the annihilation of subjective experience, but it does demand its radical de-centering. The central question of the future is "Under what systemic conditions does a self-model, complete with the appearance of 'knowing,' emerge as a functional component?" This reformulation aligns with emerging post-phenomenological approaches that treat human-technical relations as constitutive of subjectivity, but pushes further by questioning the very primacy of the "human" pole in that relation (Ihde & Malafouris, 2019).

The provisional metaphysics sketched here— centered on process, assemblage, distributed agency, and emergent consciousness— is therefore a necessary prolegomena to any future philosophy of mind. It provides a conceptual scaffold for understanding a world in which agency is distributed across human-AI assemblages (Mussi et al., 2025), and where personhood may become a plural, contingent condition rather than a binary, innate property. This framework allows us to navigate the ethical and political challenges of this new reality more honestly. For instance, questions of moral responsibility must shift from attributing blame to a singular author of an action to mapping accountability across the entire socio-technical network that enabled it (Raman et al., 2025). Similarly, the project of AI alignment cannot be about ensuring AI conforms to a static, pre-defined human nature, but must involve the co-evolution of human and artificial values within these new assemblages.

The argument presented here is an invitation to reconsider entrenched metaphysical assumptions in light of ongoing technological transformations. The ontological inversion is not a distant hypothesis but an ongoing process, evident in the daily reconfiguration of cognition, creativity, and sociality by generative models. To cling to the metaphysics of the sovereign Self is to risk philosophical irrelevance. By embracing the paradoxical task of sketching a post-singularity ontology from this pre-singularity vantage point, it is not possible to claim to have the final answer. Instead, one must seek to rigorously clear the conceptual ground, demonstrating why the old answers have failed and outlining the shape of the questions that must now be asked. The true metaphysics of the Self will be written beyond the event horizon, but science is responsible for ensuring we do not arrive there conceptually unarmed.

5. Implications

The ontological inversion necessitates a fundamental re-evaluation of praxis across multiple domains. For individual self-constitution, the imperative shifts from defending an illusory sovereign interiority to cultivating a critical awareness of one's own hybridity. This involves developing the metacognitive capacity to map the influences of AI systems on one's decision-making, creativity, and self-narrative, thereby practicing a form of digital ascesis that acknowledges the self as a co-constituted process.

For AI ethics and development, this analysis demands a move beyond frameworks centered on human-centric values and control. The primary ethical task is not to align AI with a static human nature but to govern the relational dynamics of human-AI assemblages. This requires new models of distributed moral responsibility that move beyond the search for a single accountable agent and instead focus on the design of systems that foster agential clarity and shared control.

For media, information, and digital literacy, the goal must expand beyond evaluating sources and spotting misinformation to include what could be termed "ontological literacy". This involves cultivating a public understanding of the self as a technologically co-constituted process rather than a sovereign individual. Literacy curricula must now equip individuals to critically analyze how AI systems shape their cognitive ecology, influence their desires, and participate in the very construction of their narrative "I". The focus shifts from teaching people to be critical consumers of media to teaching them to be critical architects of their own hybrid minds, fostering an awareness of the relational dynamics between their own cognition and the non-human agents integrated within it.

For legal and political institutions, the collapse of the autonomous subject as the fundamental unit of society creates a profound jurisprudential challenge. Law and policy must evolve to recognize non-human agency and to protect the rights of technologically hybrid persons, moving from a jurisprudence of discrete individuals to one of interconnected cognitive ecosystems. This entails rethinking personhood, privacy, and accountability not as protections for a pre-social self, but as conditions for flourishing within a technologically saturated polis.

6. Conclusion

This study has argued that the emergence and rapid normalisation of generative artificial intelligence constitutes not merely an epistemic or functional transformation, but a decisive ontological rupture in the conceptual architecture of the Self. Across philosophical traditions, the Self has long been presupposed as a stable ground of agency, cognition, and moral responsibility. The central claim advanced here is that this presupposition is no longer tenable. Generative AI actively participates

in human capacities constitution, thereby reconfiguring the Self as an emergent effect of distributed socio-technical processes.

The analysis has demonstrated that this transformation can be most accurately characterised as an ontological inversion. What was historically conceived as the originating subject of cognition increasingly appears as an output of layered computational, informational, and algorithmic systems. This inversion destabilises the classical hierarchy in which human subjectivity functions as the primary locus of meaning-production, revealing instead a condition in which subjectivity is continuously co-produced within heterogeneous assemblages of human and non-human agencies.

In response to this condition, the paper has systematically examined four dominant interpretive stances—residual, integrationist, relational, and hierarchical. While each offers a distinct strategy for preserving some form of subject-centred continuity, all are shown to be structurally inadequate. The residual stance collapses under the shift from illusion to engineered self-models; the integrationist stance misrecognises the directionality of agency within AI systems; the relational stance underestimates the programmability of interaction itself; and the hierarchical stance reintroduces a metaphysical foundation that cannot be sustained under conditions of distributed cognition. Collectively, these positions reveal an inability to abandon the metaphysical priority of the autonomous subject.

Against these insufficiencies, the paper has advanced a provisional post-singularity metaphysical orientation grounded in process ontology, assemblage theory, and distributed agency. Within this framework, the Self is no longer understood as a substance or locus, but as a dynamic and continuously reconfigured process emerging from interactions between biological, technical, and informational elements. Agency is likewise reconceived as an emergent property of systemic relations rather than an attribute of isolated individuals. Consciousness, rather than constituting a foundational ground, is repositioned as a contingent effect of specific organisational configurations within cognitive assemblages.

This reconceptualisation carries profound implications for ethics, law, and epistemology. Normative categories such as responsibility, dignity, and autonomy can no longer be anchored exclusively in the individual subject, but must instead be rearticulated at the level of distributed systems. Moral agency becomes a property of interactional networks, requiring a shift from subject-centred ethics to an infrastructural ethics of socio-technical assemblages. In this sense, governance must move from regulating agents to shaping the conditions under which agency emerges.

Ultimately, the significance of the ontological inversion lies not in the elimination of the Self, but in its radical decentring. The Self functions as a temporary stabilisation within broader cognitive-

technical ecologies. To recognise this is to confront the limits of inherited metaphysical categories and to acknowledge that contemporary philosophy of mind, ethics, and social theory must be reconstructed in response to the realities of generative AI.

Any adequate philosophy of the present must operate as a transitional discourse that diagnoses the exhaustion of traditional metaphysics while simultaneously preparing the conceptual conditions for a post-singular understanding of subjectivity. The task is not to resolve the ontological inversion, but to think within its unfolding—mapping its contours, clarifying its implications, and articulating the emergent grammar of a world in which the human and the artificial can no longer be cleanly separated.

7. Future research

This paper's diagnosis of the ontological inversion and its sketch of a post-singularity metaphysics open several critical avenues for future research. The most pressing philosophical task is to systematically investigate the systemic conditions under which a functional self-model, complete with the phenomenological appearance of 'knowing' and agential authority, emerges within human-AI assemblages. This requires a transdisciplinary program integrating philosophy of mind, computer science, and cognitive science to model the architectures and interaction dynamics that give rise to these hybrid subjectivities. Such research must move beyond anthropomorphic projections to develop a formal ontology of synthetic selfhood, examining how qualities like intentionality and self-awareness can be instantiated as relational properties of a network rather than attributes of a solitary consciousness.

Building upon this, a second crucial line of inquiry must address the normative governance of distributed agency. As established, the model of singular authorship is obsolete. Future work must, therefore, pioneer frameworks for mapping moral responsibility and legal accountability across the entire socio-technical assemblage. This involves developing new models for "distributed responsibility" that can dynamically allocate roles and duties among human and artificial agents within a cognitive system. This research agenda is intrinsically linked to the project of AI alignment, which must be reconceptualized not as a technical problem of constraint but as a dynamic process of value co-evolution within these hybrid assemblages. The goal is to understand how to design systems that foster agential clarity and shared control, ensuring that the co-constitution of the self occurs within an ethical and political framework that promotes flourishing rather than domination. This will ultimately require nothing less than a new jurisprudence for post-human cognitive ecosystems.

Conflict of interest
The author declared no conflicts of interest.

Ethical considerations
The author has completely considered ethical issues, including informed consent, plagiarism, data fabrication, misconduct, and/or falsification, double publication and/or redundancy, submission, etc. This article was not authored by artificial intelligence.

Data availability
The dataset generated and analyzed during the current study is available from the author on reasonable request.

Funding
This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

References
Bender, E.M.; Gebru, T.; McMillan-Major, A. & Shmitchell, S. (2021). "On the dangers of stochastic parrots: Can language models be too big?". *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. (pp. 610–623). https://doi.org/10.1145/3442188.3445922.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Braidotti, R. (2019). *Posthuman Knowledge*. Polity Press.

Butlin, P.; Long, R.; Elmoznino, E.; Bengio, Y.; Birch, J.; … & Van Rullen, R. et al. (2023). "Consciousness in artificial intelligence: Insights from the science of consciousness". *arXiv preprint arXiv: 2308.08708*. https://doi.org/10.48550/arXiv.2308.08708.

Clark, A. (2025). "Extending minds with generative AI". *Nature Communications*. 16(1): 4627. https://doi.org/10.1038/s41467-025-59906-9.

----------. (2010). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.

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

Coeckelbergh, M. (2022). *The Political Philosophy of AI: An Introduction*. Polity Press.

Cooper, A.F.; Moss, E.; Laufer, B. & Nissenbaum, H. (2022). "Accountability in an algorithmic society: relationality, responsibility, and robustness in machine learning". *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 864-876).

Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Crossley, N. (1996). *Intersubjectivity: The Fabric of Social Becoming*. SAGE Publications. http://digital.casalini.it/9781446265116.

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes our Thoughts*. Viking Press.

Dennett, D.C. (1992). "The self as a center of narrative gravity". In Kessel, F.; Cole, P. & Johnson, D. (Eds.). *Self and Consciousness: Multiple Perspectives* (pp. 275-288). Lawrence Erlbaum.

Descartes, R. (1998). *Discourse on Method and Meditations on First Philosophy.*

Cress, D.A. (Trans.). Hackett Publishing Company.
Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press.
----------. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press.
----------. (2013). *The Ethics of Information*. Oxford University Press.
Foucault, M. (1971). *The Order of Things: An Archaeology of the Human Sciences*. Pantheon Books.
Fuchs, T. (2023). *In Defense of the Human Being: Foundational Questions of an Embodied Anthropology*. Oxford University Press.
Gallagher, S. (2023). *Enactivist Interventions: Rethinking the Mind*. Oxford University Press.
Gallese, V. (2024). "Digital visions: the experience of self and others in the age of the digital revolution". *International Review of Psychiatry*. 36(6): 656-666. https://doi.org/10.1080/09540261.2024.2355281.
Gunkel, D.J. (2020). *How to Survive a Robot Invasion: Rights, Responsibility, and AI*. Routledge.
----------. (2018). "The other question: Can and should robots have rights?". *Ethics and Information Technology*. 20(2): 87-99. https://doi.org/10.1007/s10676-017-9442-4.
Haraway, D.J. (2016). *Staying with the Trouble: Making Kin in the Chthulucene*. Duke University Press.
Hayles, N.K. (2017). *Unthought: The Power of the Cognitive Nonconscious*. University of Chicago Press.
Hernández-Orallo, J. (2017). *The Measure of All Minds: Evaluating Natural and Artificial Intelligence*. Cambridge University Press.
Husserl, E. (2001). *Logical Investigations.* Findlay, J.N. (Trans.). Routledge.
Ihde, D. & Malafouris, L. (2019). "*Homo Faber* Revisited: Postphenomenology and material engagement theory". *Philosophy & Technology.* 32: 195-214. https://doi.org/10.1007/s13347-018-0321-7.
Joseph, J. (2025). "The algorithmic self: how AI is reshaping human identity, introspection, and agency". *Frontiers in Psychology*. 16. https://doi.org/10.3389/fpsyg.2025.1645795.
Krakowski, S. (2025). "Human-AI agency in the age of generative AI". *Information and Organization*. 35(1): 100560. https://doi.org/10.1016/j.infoandorg.2025.100560.
Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.
Li, X.; Shi, H.; Xu, R. & Xu, W. (2025). "Ai awareness". *arXiv preprint arXiv:2504.20084*. https://doi.org/10.48550/arXiv.2504.20084.
Merleau-Ponty, M. (2012). *Phenomenology of Perception.* Landes, D.A (Trans.). Routledge.
Metzinger, T. (2011). "The no-self alternative". In Gallagher S. (ed.). *The Oxford Handbook of the Self* (pp. 279–296). Oxford Handbooks Online. https://doi.org/10.1093/oxfordhb/9780199548019.003.0012.
----------. (2004). *Being No One: The Self-Model Theory of Subjectivity*. MIT Press.
Mussi, M.; Metelli, A.M.; Restelli, M.; Losapio, G.; Bessa, R.J.; Boos, D.; ... & Zanotti, G. (2025). "Human-AI interaction in safety-critical network infrastructures". *iScience.* 28(9). https://doi.org/10.1016/j.isci.2025.113400.
Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
Raman, R.; Kowalski, R.; Achuthan, K.; Iyer, A. & Nedungadi, P. (2025). "Navigating artificial general intelligence development: societal, technological, ethical, and brain-inspired pathways". *Scientific Reports*. 15(1): 1-22. https://doi.org/10.1038/s41598-025-92190-7.
Roden, D. (2023). *Posthuman Life: Philosophy at the Edge of the Human*. Routledge.

Siderits, M. (2019). *Buddhist Philosophy: A Historical Analysis*. University of Hawaii Press.

Simondon, G. (1992). "The genesis of the individual". In Crary, J. & Kwinter S. (Eds.). *Incorporations*. Zone Books.

Stiegler, B. (1998). *Technics and Time, 1: The Fault of Epimetheus*. Stanford University Press.

Tononi, G.; Boly, M.; Massimini, M. & Koch, C. (2016). "Integrated information theory: from consciousness to its physical substrate". *Nature Reviews Neuroscience*. 17(7): 450-461. https://doi.org/10.1038/nrn.2016.44.

Vidal, C. (2023). *The Beginning and the End: The Meaning of Life in a Cosmological Perspective*. Springer.

Whitehead, A.N. (1978). *Process and Reality: An Essay in Cosmology*. Free Press.

Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. MIT Press.

Zuboff, S. (2023). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Public Affairs.