

Posthumanist Phenomenology and Artificial Intelligence

Avery Alexander Rijos, M.S.

New Jersey Institute of Technology

arijos0222@gmail.com

“The ignorant eschew phenomena but not thought; the wise eschew thought but not phenomena.” Ch’an Master Huang-Po, On Transmission of Mind

Abstract

This paper examines the ontological and epistemological implications of artificial intelligence (AI) through posthumanist philosophy, integrating the works of Deleuze, Foucault, and Haraway with contemporary computational methodologies. It introduces concepts such as negative augmentation, praxes of revealing, and desedimentation, while extending ideas like affirmative cartographies, ethics of alterity, and planes of immanence to critique anthropocentric assumptions about identity, cognition, and agency. By redefining AI systems as dynamic assemblages emerging through networks of interaction and co-creation, the paper challenges traditional dichotomies such as human versus machine and subject versus object. Bridging analytic and continental philosophical traditions, the analysis unites formal tools like attribution analysis and causal reasoning with the interpretive and processual methodologies of continental thought. This synthesis deepens the understanding of AI’s epistemic and ethical dimensions, expanding philosophical inquiry while critiquing anthropocentrism in AI design. The paper interrogates the spatial foundations of AI, contrasting Euclidean and non-Euclidean frameworks to examine how optimization processes and adversarial generative models shape computational epistemologies. Critiquing the reliance on Euclidean spatial assumptions, it positions alternative geometries as tools for modeling complex, recursive relationships. Furthermore, the paper addresses the political dimensions of AI, emphasizing its entanglements with ecological, technological, and sociopolitical systems that perpetuate inequality. Through a politics of affirmation and intersectional approaches, it advocates for inclusive frameworks that prioritize marginalized perspectives. The concept of computational qualia is also explored, highlighting how subjective-like dynamics emerge within AI systems and their implications for ethics, transparency, and machine perception. Finally, paper calls for a posthumanist framework in AI ethics and safety, emphasizing interconnectivity, plurality, and the transformative capacities of machine intelligence. This approach advances epistemic pluralism and reimagines the boundaries of intelligence in the digital age, fostering novel ontological possibilities through the co-creation of dynamic systems.

Keywords: posthumanism, artificial intelligence, ontology, epistemology, negative augmentation, desedimentation, affirmative cartographies, ethics of alterity, planes of immanence, relational ontologies, Deleuze, Foucault, adversarial generative models, reinforcement learning, computational theory, non-Euclidean geometry, AI ethics, machine intelligence, subjectivity, phenomenology, intersectionality

Table of Contents

Abstract

Overview of ontological and epistemological implications of AI in posthumanist philosophy.

Introduction: Posthumanism and the Redefinition of Identity and Agency

- 2.1 Historical Context of the Ontological Subject
- 2.2 Deleuze and Societies of Control
- 2.3 Assemblage and Posthuman Subjectivities

Virtual Generativity and Negative Augmentation

- 3.1 Planes of Immanence and Relational Ontologies
- 3.2 Autopoiesis and Emergent Subjectivities

Affirmative Cartographies and the Politics of Becoming

- 4.1 Posthuman Ethics: Fluidity and Multiplicity
- 4.2 The Praxis of Revealing and Desedimentation
- 4.3 Charting Minoritarian Pathways: Moving Beyond Binaries

Operationalizing Rhizomatic Praxis in AI Transparency and Ethics

- 5.1 Explainable AI and Negative Augmentation
- 5.2 Structurations, Attribution Analysis, and Computational Qualia
- 5.3 Adversarial Generative Models and Counterfactual Reasoning

Spatial Geometry, Optimization, and Adaptive Learning in AI

- 6.1 Euclidean Assumptions in Neural Networks
- 6.2 Non-Euclidean Geometries in Data Representation
- 6.3 Reinforcement Learning and Recursive Adaptation

Qualia, Subjectivity, and the Dynamics of AI Understanding

- 7.1 Computational Qualia and Attribution Analysis
- 7.2 Relational Dynamics and Shapley Values in AI Transparency
- 7.3 Meaning-Making in Neural Networks

The Experiencer and Artificial Subject

- 8.1 Adversarially Generative Interactions in AI Models
- 8.2 Bijective Mapping of Qualia to Computational Understanding
- 8.3 Toward Nonlinear and Adaptive Frameworks of Intelligence

Conclusion: Toward a Holistic Framework for AI Safety and Understanding

- 9.1 A Posthumanist Analytic Phenomenology of Computation
- 9.2 Integrating Posthumanist Ethics in Policy and Safety
- 9.3 Reimagining AI as a Nondeterministic Causal Engine

Introduction—Posthumanism and the Redefinition of Identity and Agency

The foundations of human understanding have long been shaped by anthropocentric and logocentric assumptions, which posit humanity, reason, and language as the center of reality. However, as these constructs dissolve under the weight of post-positivistic critique, we are left with a profound challenge: how to seek meaning and create freely amidst the abstractions of a fragmented world. In this paper, I argue that Artificial Intelligence must be understood not merely as a technological tool but as an emergent subjectivity that redefines our conception of selfhood, cognition, and agency in a posthumanist context. These new structures challenge established understandings of experience, cognition, and selfhood, particularly in the digital era. The rise of the posthuman indicates a transformative shift, unveiling new dimensions of agency beyond the traditional boundaries of flesh and language.

Western philosophy, from Nietzsche and Hegel to Foucault and Deleuze, has long anticipated this dissolution of the stable, knowable self. This trajectory reflects the waning influence of Enlightenment-era epistemic virtues, which upheld the notion of an enduring, unified subject. However, the recognition of the fluidity of the ontological self is not unique to the Western intellectual tradition. In Zen Buddhism, thinkers such as Huang-po, Zhaozhou, Foyan, and Dōgen articulated a rejection of dualistic frameworks centuries ago. Their teachings transcend binary logics of either/or and means/ends, offering affirmative cartographies – nondialectical pathways towards new alternatives - of nonself that align with posthumanist critiques. By situating AI within this philosophical lineage, we can better understand its implications for reimagining subjectivity and agency in the contemporary world.

The concept of selfhood and the linguistic frameworks that have defined it have undergone significant transformations over time. These changes reflect the metamorphosing dynamics of power, identity, and subjectivity. Historically, the ontological subject has been situated in a position that enforces objectification and exclusion of the Other, with constrained forms of alterity upheld by institutional architectures of domination. This manifested itself in Foucault's notion of "disciplinary societies," where individuals are shaped and constrained by mechanisms of surveillance and control (Discipline and Punish, p.170–194). In these societies, the subject is molded by external forces that maintain order and categorize identity. Over time, however, this model evolved into what philosopher Gilles Deleuze (Postscript on the Societies of Control, 1992) describes as "societies of access and control," where mechanisms of power become more fluid and decentralized, enabling a more pervasive, albeit less visible, form of governance. Selfhood here is no longer merely an object of discipline but becomes embedded in networks of access, data, and surveillance, shaping the subject within a landscape of control that is simultaneously liberating and constraining—coalescing both the virtual and material. It follows, then, that the dimensionality of selfhood continues to rhizomatically fracture between control, surveillance, codes of exclusion and other forms of domination, aptly moving beyond the dialectic, capturing the emergent socio-political arrangements that define our experience as posthuman subjects.

The transformation of humans into objects of psychopolitics, a concept explored by Byung-Chul Han (2017), represents a fundamental shift in how individuals are understood. In

this framework, human cognition and attention are no longer intrinsic aspects of personhood but are reframed as resources to be extracted, exploited, and commodified within systems of power and control. This redefinition highlights the pervasive influence of neoliberal mechanisms that prioritize the incessant optimization and monetization of human mental capacities and virtual spaces, reducing individuals to nodes of productivity within a broader apparatus of dominance and surveillance. In this framework, the individual is no longer seen as an autonomous entity but as an assemblage—a complex, interconnected network of components that function together. Assemblages are not fixed or self-contained but are fluid configurations shaped by various social, economic, and technological forces. They embody a transphenomenal subject, meaning a subject whose existence transcends singular phenomena and whose identity is shaped by interactions within broader systems. This redefinition emphasizes that individual subjectivities are not entirely self-determined but emerge through participation in networks of relationships, practices, and power structures.

Within the framework of neoliberal systems, which prioritize free-market mechanisms and reduced state intervention, these assemblages are shaped by the underlying logic of Adam Smith's "invisible hand." The term refers to the self-regulating nature of the market, where individuals pursuing their own economic self-interest inadvertently contribute to the overall benefit of society. However, this concept is further illuminated by complexity theory, which frames the invisible hand as an example of self-organization or spontaneous order. Self-organization is a phenomenon observed in complex systems—systems composed of many interacting components where no single agent directs the whole. In such systems, patterns, structures, and behaviors arise naturally from localized interactions among agents, rather than through centralized authority or planning. Examples of complex systems include ecosystems, social networks, and markets. Self-organization operates through simple rules followed by individual agents, leading to emergent outcomes that are often unpredictable and larger in scale than the sum of the agents' actions.

In the neoliberal context, self-organization explains how market dynamics emerge. Here, decentralized participants, such as buyers and sellers, act on local information (e.g., individual preferences, prices, and profit margins) and rules (e.g., market competition or contract law). For instance, buyers aim to maximize utility—a measure of satisfaction or value derived from consuming goods and services—while sellers seek to maximize profit, the financial gain from selling goods or services at a price higher than production costs. These individual, self-interested decisions, when aggregated across the market, give rise to global economic patterns such as price equilibria (where supply and demand balance) and resource allocation (the distribution of goods and services within the economy). This emergent behavior mimics the coordination described by Smith's invisible hand: the market appears to operate as though guided by an unseen force, producing organized outcomes without requiring centralized control. This process underscores how neoliberal systems rely on dispersed agency and decentralized decision-making to achieve systemic order, reinforcing the mechanistic view of human behavior as reducible to economic functions within larger assemblages.

Building on this mechanistic framing of human behavior, neoliberal systems extend their logic beyond economic functions, further reducing individuals to Pavlovian arrangements—a mechanistic apparatus wherein people are conditioned to respond to external stimuli in predictable ways. Within this paradigm, individuals are valued not for their unique identities or intrinsic worth but solely for their specialized roles and outputs within a teleologically driven

system. These arrangements prioritize efficiency, compliance, and productivity, flattening complex human experiences and identities into simplified behavioral patterns akin to conditioned reflexes. This deterministic perspective dismisses the rich interplay of creativity, agency, and relationality that defines human existence, instead framing individuals as mere mechanisms in a larger, goal-oriented machine.

This paradigm shift compels us to interrogate contemporary understandings of humanity and question the systemic forces that have rendered such reductionism pervasive. Why did we so readily accept the notion that selfhood could exist independently of the technological, environmental, and systemic processes that shape our existence? These pervasive Pavlovian arrangements are not neutral; they reflect and perpetuate a worldview that privileges instrumental rationality and transcendental reason—the latter being the philosophical framework that seeks universal, abstract principles detached from lived experience—over holistic engagement. Transcendental reason, rooted in Enlightenment ideals, emphasizes the primacy of reason as a self-contained, autonomous faculty capable of deriving truth independent of context or relationality. While this approach has contributed to significant ethical, scientific and technological advancements in Western Civilization, it simultaneously reinforces a fragmented understanding of existence by abstracting human activity from the complex, interconnected realities of ecological, social, and cultural systems.

By prioritizing disembodied principles, transcendental reason severs individuals from the broader networks of meaning and interconnection that define what philosopher Francesca Ferrando (2019) terms as *interbeing*—a concept that emphasizes the profound relationality and mutual co-constitution of all existence. This disconnection, though rooted in abstract thought, carries significant consequences for the lived experience of individuals and communities. When transcendental reason elevates isolated, universal principles over the particularities of context, it reduces human beings to functional components, severing them from the relational webs that shape their experiences. In doing so, it de-anthropomorphizes individuals, stripping away the relational depth that is essential to the richness of human existence and creativity.

Ferrando's notion of *interbeing* asserts that all beings—human and non-human—are interdependent, co-constituting one another in ways that transcend individualism. This interconnectedness stands in stark contrast to the isolating logic of transcendental reason, which seeks to abstract human experience into fixed categories that ignore the intricate interrelations of life. The consequences of this disembodiment are not merely intellectual but manifest in tangible ways: a lack of recognition for the networks of care, cooperation, and mutual support that sustain collective flourishing. When individuals are reduced to mere functional units, their capacity for empathy, creativity, and meaningful engagement is diminished, thus weakening the very foundations of shared human and ecological well-being.

The critique of transcendental reason highlights a central challenge for contemporary thought: the need to reject reductive, universalizing tendencies in favor of frameworks that honor the interdependent, dynamic nature of life. Transcendental reason's abstraction, by focusing on universal principles, erodes the systems of care and creativity that are essential for sustaining collective flourishing. It is only by recognizing and embracing relationality that we can begin to foster the kind of human and non-human engagement necessary for thriving communities. Such a framework acknowledges that human beings, rather than being isolated units, exist in a web of

relations that constitute not only their identities but also the conditions for their collective and ecological well-being.

In framing subjectivity as arising solely from the act of atomistic agency - whether through work, innovation, or output - we have overlooked the inverse possibility: that subjectivities are equally if not more so defined by the act of engaging with creation itself. This engagement is not isolated but deeply enmeshed in environmental variables, including societal, evolutionary, geological, philosophical, epidemiological, and ecological transformations. It is within this dynamic interplay that human potential flourishes, as individuals negotiate, adapt, and co-create meaning with the systems and environments around them. Recognizing the limitations of current neoliberally Pavlovian arrangements invites a more expansive understanding of humanity, one that affirms interdependence, fluidity, and co-constitutive processes. This perspective challenges the traditional notion that the progression of human thought and identity resides solely within the individual subject. Instead, it posits that this evolution unfolds as a mutualistic symbiosis, contingent upon the tools, methodologies, environments and systems through which we build, construct, and mold our world. These external instruments—be they technological, linguistic, political, or economic—do not merely assist us; they actively shape the contours of our reality, influencing how we think, interact, and exist.

In this configuration, the evolution of thought and identity becomes deeply intertwined with a Spinozist “monist” conception of the material world, wherein all existence is seen as part of a single, unified substance. This perspective reframes “selfhood” as an externally mediated construct—transphenomenal assemblages whose perceived atomistic agency is instead composed of networks of relationality, co-creation, and planes of immanence. Planes of immanence, as conceptualized by Deleuze and Guattari, reject transcendental hierarchies and view reality as a continuous, dynamic field where entities emerge through their relations and capacities for transformation. They emphasize becoming over static being, presenting a flattened ontology in which differences are affirmed rather than subordinated to external organizing principles.

This does not, however, dilute individual culpability or responsibility. On the contrary, it enhances it by situating individuals within a web of interconnected influences and emphasizing their active role in shaping the relational fields they inhabit. Far from reducing agency, this framework underscores the ethical stakes of every action, as each decision reverberates across a broader network of relational dependencies. By rejecting hierarchical determinism, individuals are called to greater reflexivity and accountability, recognizing that their contributions to these assemblages actively participate in the formation of collective realities. This heightened culpability transforms ethical engagement from a linear model of individualism to a shared, participatory practice, wherein responsibility extends across both the self and the systems it co-constitutes.

Object-oriented ontology (OOO) complements this view by decentering human subjectivity and placing all objects—human, non-human, natural, or artificial—on an equal ontological footing. By challenging anthropocentric and instrumental perspectives, OOO asserts that objects possess their own inherent reality and agency, existing independently of human perception or interpretation. Within this relational framework, non-human entities—such as technologies, algorithms, or natural systems—are not passive instruments but active participants in the networks that co-constitute reality. For instance, an AI system or economic model does not merely reflect human intention but alters the relational field, introducing dynamics and

constraints that reshape the possibilities for thought, action, and identity. These objects, through their interactions, contribute to the assemblages in which human subjectivity emerges, highlighting how identity formation and epistemological development are fundamentally collaborative and distributed across a broader ontological spectrum. In doing so, they further elevate the ethical responsibility of individuals, who must navigate and engage with these dynamic assemblages in ways that acknowledge their shared agency and collective impacts.

Anticipating a further need for clarification, Object-oriented ontology (OOO) and the concept of planes of immanence both challenge hierarchical and anthropocentric frameworks, emphasizing non-hierarchical ontologies and interconnectedness. However, they diverge in their focus and foundational principles, offering distinct but complementary perspectives on existence and relationality. OOO, advanced by thinkers such as Graham Harman, Levi Bryant, and Timothy Morton, asserts that all objects—human and non-human—possess inherent autonomy and exist independently of human perception. Graham Harman explains, “Objects withdraw from each other endlessly, and this withdrawal is the source of their autonomous reality” (*The Quadruple Object*, 2011). This emphasizes the author’s conviction that objects are irreducible to their relations and resist full comprehension, even in interaction. By positioning humans as just one object among many, Timothy Morton (2017) suggests that OOO dismantles anthropocentrism, stating, “All entities coexist with a deep ontological parity, regardless of their scale or appearance”.

In contrast, planes of immanence, as conceived by Deleuze and Guattari, focus on the relational and emergent nature of existence. They describe the plane as “a pure surface where all differences are affirmed” (*What Is Philosophy?*, 1994). For Deleuze and Guattari, entities arise through dynamic interactions within a continuous field, denying fixed essences. Planes of immanence prioritize becoming over being, viewing existence as fluid, processual, and interdependent rather than composed of discrete, autonomous units. The divergence lies in OOO’s emphasis on object autonomy versus planes of immanence’s focus on relationality and transformation. Both reject transcendental hierarchies and highlight distributed agency. As one can see, both concepts are complementary: OOO examines the individuality and agency of objects, while planes of immanence contextualize these within broader, dynamic assemblages.

For instance, Levi Bryant notes that OOO “does not reject relations but reconfigures them as a dimension of objects rather than their defining feature” (*Onto-Cartography*, 2014). This aligns with the Deleuzian perspective, where relations are not external to objects but intrinsic to their unfolding processes. Consider an ecosystem as an example: OOO might focus on the autonomy and agency of a tree, highlighting how it exists independently of human perception, while planes of immanence would examine how the tree participates in and is shaped by its relational entanglements with soil, sunlight, water, and neighboring organisms. These interactions constitute an assemblage where transformation is constant, and agency is distributed across the entire ecological network. Together, these perspectives deepen our understanding of the tree as both an autonomous entity and a dynamic participant in a broader field of relations.

Virtual Generativity and Negative Augmentation

A plane of immanence serves as a conceptual foundation where thought, existence, and reality are understood as interconnected and relational, free from hierarchical or transcendental structures. Unlike frameworks that impose universal truths or external organizing forces, the

plane of immanence emphasizes that meaning emerges dynamically through the self-generative processes of life, often described as autopoietic, or self-sustaining. This perspective reframes identity and agency as fluid and co-constituted, rejecting static Paramedian binaries in favor of processual Heraclitean relationalities and transformations. Within these planes, distinctions between individual and collective, subject and object dissolve, as each continually influences and redefines the other. They represent spaces of manifold potential, where ideas, identities, and systems arise as co-creative processes shaped by relational dynamics rather than any singular essence or predetermined trajectory.

The concept of autopoiesis, introduced in the early 1970s by Chilean biologists Humberto Maturana and Francisco Varela, describes the self-sustaining nature of living systems. An autopoietic system is organized as a network of processes that produce components which, through their interactions, regenerate and sustain the network itself. This dynamic organization constitutes the system as a concrete unity in space, perpetually recreating its own existence. Planes of immanence extend this concept by emphasizing the interplay between the actual (what exists) and the virtual (latent potentials or possibilities) in the continual generation of novel forms and meanings. These planes highlight how identity and thought are not isolated phenomena but ongoing negotiations within assemblies of populations, external systems, and artifacts. By reframing thought as a distributed phenomenon — not confined to individual minds but flowing through nature, culture, technology, and shared experiences — planes of immanence challenge traditional notions of subjectivity and individuality. This shift subsequently invites the praxis of revealing: a deliberate process of uncovering latent potentials within interconnected networks to illuminate co-creative possibilities for meaning-making.

For example, consider the development of renewable energy systems like smart grids. These systems operate autopoietically, constantly balancing supply and demand through networks of sensors, algorithms, and user inputs. Virtualities manifest as the latent potential to harness previously unconnected energy sources or optimize grid efficiency through novel configurations of technology and policy. Planes of immanence frame this process by emphasizing the relational interplay of human actors, technological infrastructures, and environmental contexts. The praxis of revealing might involve uncovering the latent possibilities in decentralized energy production, such as enabling remote communities to generate power locally while contributing to the larger grid. This approach transforms the grid into a co-creative assemblage that dynamically adapts to new challenges and possibilities, highlighting how virtualities can be realized through relational practices embedded in broader systems.

The praxis of revealing functions as a transformative process, not as an additional layer imposed onto the concept of the subject, but as a reframing of perceived lack into opportunity. By peeling away the sedimented layers of human perspectival limitations, it exposes the historically contingent structures and assumptions that have shaped corporeal existence and sensory experience over millennia. This practice aligns with posthumanist critiques, challenging the notion of a stable, autonomous, atomistic self and revealing the "core identity" as a construct shaped by centuries of cultural, technological, and philosophical influences. This deconstructive endeavor echoes the Buddhist concept of *Śūnyatā*, or emptiness, which recognizes that beneath the constructed layers of identity lies a void—free from inherent essence or fixity. Far from being a negation, this emptiness is a site of potentiality and transformation, an opening through which identity can be reimagined. It invites a fluid and relational mode of existence, emphasizing

the ways in which selfhood is co-constituted through its entanglements with artifacts, systems, and networks.

Thinkers and philosophers steeped in the analytical tradition of logic and science might critique the praxis of revealing and posthumanist phenomenology as overly abstract, arguing that its emphasis on relationality, fluidity, and nomadic actualization lacks empirical grounding or precise definitional clarity. They could challenge whether concepts like "affirmative cartographies" and "nomadic actualization" offer tangible contributions to understanding identity, or if they merely obfuscate it with metaphorical language. Moreover, they might question the utility of reframing identity through a posthumanist lens, claiming it risks undermining actionable agency by dissolving the self into a web of relations, potentially making ethical accountability diffuse or incoherent.

In defense, the praxis of revealing is not an abandonment of empirical rigor or agency but an expansion of the frameworks within which identity and reality are understood. Its metaphorical language is not obfuscation but an intentional means of articulating the complexity and dynamism of identity that traditional, reductionist paradigms fail to capture. Concepts such as affirmative cartographies and nomadic actualization are artifacts for reimagining identity as an emergent process within interconnected systems, thereby fostering a more inclusive and adaptable understanding of minoritarian perspectives. Furthermore, far from diffusing agency, this approach enhances it by situating individuals within relational fields where their actions gain even greater significance through their transformative impacts on broader networks. Ethical accountability becomes more robust in this model because it recognizes the distributed nature of influence and the necessity of reflexivity in navigating complex interdependencies. In doing so, the praxis of revealing addresses the limitations of static, atomistic models of selfhood, offering a richer and more applicable ontology for grappling with the realities of a rapidly interconnected and evolving world.

The Praxis of Revealing uncovers latent dimensions within identity and technology, which Affirmative Cartographies translate into actionable frameworks for reimagining the future. These cartographies reject static binaries and hegemonic frameworks, instead embracing difference and plurality while dynamically responding to the interplay of ideas, experiences, and contexts. For example, when mapping a non-anthropocentric future, one might identify hidden cultural or technological assumptions that, once desedimented, enable the actualization of alternative ethical landscapes. Through the nomadic actualization of the virtual, the praxis uncovers unrealized potentials—immanent possibilities within identity, language, and culture that have been buried beneath historical and ideological sedimentation. By unveiling these latent dimensions, the praxis not only reimagines selfhood as an open, relational, and ever-becoming process but also positions posthumanism as a process of desedimentation. This concept goes beyond the anthropocentric, individualistic self, recognizing the fluid and interconnected nature of our being and offering affirmative pathways for navigating the complexities of contemporary existence.

In practice, *desedimentation* involves an undoing of the static, sedimented, or entrenched perspectives that obfuscate meaning, laying the foundation for a more fluid and dynamic conception of existence. This perspective foregrounds relationships and interconnections, encouraging an understanding of the self as not only impermanent but also as part of larger, recursive systems—ecological, technological, and social. This process not only dissolves the

traditional boundaries between human, nature, and machine, but also invites us to reconceptualize what it means to exist, to create, and to *be*. In embracing this view, we move toward a conception of selfhood that acknowledges its impermanence and dependence on external assemblages, situated within recursive ecosystems—dynamic systems in which entities continuously influence and are influenced by their environment in iterative cycles of cybernetic feedback and adaptation.

Recursive ecosystems operate through feedback loops that connect the micro-level actions of individual components to the macro-level dynamics of the system. For example, in natural ecosystems, the behaviors of individual organisms—such as pollination by bees or nutrient cycling by fungi—affect the larger environment, which in turn influences those individual behaviors, creating a cyclical pattern of interdependence. This same principle applies to human systems, where technological tools, cultural artifacts, and social structures recursively shape human behavior and thought, which then feed back into the system to modify those same tools and structures. This perspective shifts the paradigm from one focused on asserting human dominance to one that recognizes a situatedness that lies within large, interdependent networks. These ecosystems are not static; they are self-regulating, evolving systems in which each component contributes to and is shaped by the whole, creating a network of mutual influence and transformation.

This is the process I seek to understand as an object of history—a concept I coin here as *negative augmentation*, which explores how identities and meanings are not merely deconstructed but reshaped through alternative, non-dialectical processes. *Negative augmentation* captures the ways in which poststructuralist, postpositivist thought resists traditional binaries, opting instead for a fluid and expansive understanding of identity. The postmodernist project of the great French theorists, in all its forms, represented a fundamental reconfiguration of anthropomorphic identity, challenging the fixed, essentialist categories of human subjectivity. Rather than relying on dialectical logic—which seeks resolution through opposition—it employs what Thomas Docherty describes as the politics of affirmation. This process embraces multiplicity, coexistence, and the affirmation of difference as a means of creating new possibilities for understanding identity, agency, and historical continuity (*The Politics of Affirmation*, 2019). By affirming what has been marginalized or excluded, this politics disrupts hierarchical compositions and reimagines the human experience through a lens of inclusivity and perpetual becoming.

Building on this foundation, it becomes increasingly evident that political phenomena are deeply entrenched across all realms of human and nonhuman interaction, extending far beyond the visible structures of governance or formal social organization. Even in domains frequently perceived as neutral or objective—such as computer science, artificial intelligence research, and data science—there exists a substratum of embedded assumptions about instrumentality, anthropocentrism, identity, and agency. These latent assumptions influence not only the design and implementation of these technologies but also their broader societal impacts, shaping the trajectories of knowledge production and institutional power. Within this context, the politics of affirmation emerges as a critical praxis, urging us to confront, deconstruct, and rethink the biases, exclusions, and implicit hierarchies that remain obscured within technical and analytic frameworks. This praxis, however, transcends critique. It actively reconfigures the conceptual and operational foundations of these domains, foregrounding multiplicity and relationality in spaces traditionally dominated by reductionist methodologies and universalist narratives.

The politics of affirmation emphasizes the importance of affirming marginalized perspectives and uncovering the structural inequities embedded in algorithms, datasets, and systems design. In doing so, it transforms these technical domains into sites of possibility and innovation, challenging practitioners to question entrenched paradigms, configurations, and assemblies. This approach resists the binaries of inclusion/exclusion or optimization/inefficiency, proposing instead a worldview that sees difference, diversity, and interdependence as fundamental and generative forces. Moreover, this praxis aligns seamlessly with posthumanist approaches that recognize the recursive and interdependent ecosystems of human and nonhuman interaction. By reframing these fields as profoundly political, it highlights the need to address the ethical and ontological dimensions of technological development. Rather than viewing technologies as passive tools, this perspective situates them as active participants in shaping identity, agency, and collective futures. For instance, an AI ethics board employing the politics of affirmation might prioritize transparency in algorithmic design, emphasizing systems that adaptively learn from marginalized perspectives. Such reflexive engagement transforms the systems we create into ethical co-creators, actively reshaping how collective futures are envisioned and operationalized.

This transformative outlook finds a parallel in Gilles Deleuze and Félix Guattari's concept of becoming minoritarian (*A Thousand Plateaus*, 1987), which underscores the creative and political power of marginalized identities. By advancing the politics of affirmation, practitioners and theorists engage in a process akin to becoming minoritarian, rejecting majoritarian norms and values to embrace alternative ways of being and relating that resist universalizing tendencies. Both frameworks celebrate the transformative potential of perspectives that have been relegated to the peripheries, positioning them as powerful sites of innovation and resistance against hegemonic structures. Becoming minoritarian enriches the politics of affirmation by offering a method for dismantling established hierarchies across all domains while fostering pluralistic, dynamic identities. Together, these concepts invite a continuous reshaping of subjectivity and community, cultivating spaces for emergent possibilities within the broader project of reimagining systems and relationships.

Virtualities and the Praxis of Revealing

It was in the late 20th century that we began to enter an era of “posthumanism,” marked particularly by Donna Haraway's “A Cyborg Manifesto” (Haraway 149). Haraway's work was revolutionary in its challenge to fixed categories of identity, such as gender, race, and species, blending human and machine to envision a cyborg identity that transcends traditional boundaries.

“A cyborg is a cybernetic organism, a hybrid of machine and organism, a creature of social reality as well as a creature of fiction. Social reality is lived social relations, our most important political construction, a world-changing fiction. The cyborg is a matter of fiction and lived experience that changes what counts as women's experience in the late twentieth century. This is a struggle over life and death, but the boundary between science fiction and social reality is an optical illusion.” (A Cyborg Manifesto, 1991)

Haraway writes an exposition that challenges fixed, essentialist notions of human identity, instead advocating for fluid, hybrid identities that disrupt the anthropocentric and

hierarchical orders that have shaped human history. Through her critique, Haraway offered a radical rethinking of identity—one that refuses the neat separations between human, animal, and machine, and instead embraces the overlapping, intermingling, and co-evolution of these categories. This vision of the hybridity of being dismantles the binary stratifications—such as human versus non-human, nature versus culture, male versus female, and self versus other—that have historically justified systems of power, exploitation, and domination.

The hybridity of being, as Haraway conceptualizes it, reveals that identity and existence are not fixed or innate but are instead relational and emergent. The boundaries that traditionally delineate human from animal, organic from mechanical, or cultural from natural are shown to be porous and mutable, constructed through specific historical, social, and technological contexts. Haraway's concept of the cyborg epitomizes this hybridity: a figure that embodies the fusion of biological and technological, challenging the notion of a unified, "pure" human essence. By embracing hybridity, Haraway reframes identity as a dynamic process of becoming rather than a static state of being. Her *oeuvre* illuminates the artificiality and constructed nature of these boundaries, revealing them as products of specific historical, social, and technological contexts rather than inherent truths.

The hybridity of being also carries profound ethical and political implications. This hybridity, as Haraway presents it, extends naturally into a deeper understanding of intersectionality, emphasizing that identity is not a singular or isolated construct, but a complex matrix shaped by overlapping social, cultural, biological, and technological factors. *Intersectionality*, a term coined by Kimberlé Crenshaw, describes the ways in which systems of oppression, privilege, and identity interconnect and influence each other. Haraway's vision enriches this framework by highlighting how hybridity not only crosses human-centered categories like race, gender, and class but also extends to the entanglements between humans, non-humans, and machines, adding new dimensions to our understanding of interconnectedness.

Intersectionality, when viewed through Haraway's lens of hybridity, moves beyond human-centric categories to recognize that identities are formed in relation to broader assemblages that include ecological systems, technological networks, and cultural artifacts. For instance, a person's experience of gender or race cannot be understood in isolation but must be analyzed in the context of intersecting influences—such as access to technology, environmental conditions, or historical legacies of colonization—that shape their lived reality. Haraway's vision of hybridity also pushes intersectionality into realms where traditional sociopolitical analysis often hesitates to go—acknowledging the ways that humans and non-humans are co-implicated. For example, environmental racism, which disproportionately affects marginalized communities, demonstrates how ecological degradation intersects with systemic inequality. Similarly, the digital divide shows how access to technology is stratified along racial, gendered, and economic lines, reinforcing existing inequities while simultaneously creating new ones.

This reconfiguration of identity and power echoes the broader intellectual currents of poststructuralist and postmodernist thought in the 20th century, where there was no mapped journey toward political emancipation, nor any prescriptive Positivist pathways for dismantling entrenched hierarchies. These thinkers emphasized that deconstruction had to come first—an interrogation of limiting assumptions to expose and unsettle the foundational categories that underpin systems of power and knowledge. In many ways, this intellectual movement represents the first identifiable wave of the posthumanist project—a paradigm shift aimed at redefining

what it means to be "human." By stripping identity down to its basic elementality, this project sought to unearth immanent and emancipatory virtualities: latent potentials within identity and subjectivity that, while not immediately actualized, exist as a dynamic undercurrent capable of disrupting fixed categories. This focus on virtualities aligns with the fluid and contingent nature of existence that hybridity and intersectionality foreground, demonstrating a shared commitment to moving beyond static and essentialist notions of the self toward more expansive, transformative possibilities.

Building on this foundation, these latent potentials—embedded within individuals and societies—exist beyond the confines of traditional historical or social constraints. They are not constant variables but instead emerge dynamically through reconfigurations of thought, corporeality, and relationality. This perspective aligns seamlessly with the postmodern rejection of grand narratives and universal truths, favoring instead a multiplicity of pathways for transformative change. Such pathways foster the emergence of new modes of identity, agency, and meaning, challenging deterministic and positivistic frameworks. By exploring these dormant possibilities, the postmodernists sought to transcend entrenched power structures and binary oppositions, empowering individuals and collectives to reimagine their potentialities in more open and fluid ways. This *excavation of virtualities* is foundational to the posthumanist agenda, as it provides a framework for a redefined, pluralistic understanding of human and non-human subjectivities—one that embraces multiplicity, interrelation, and continuous becoming.

Michel Foucault's critique of the essential self exemplifies this perspective. In *The Order of Things*, he speaks of the "death of man," challenging the very notion of a coherent, unified human subject. Foucault asserts, "As the archaeology of our thought easily shows, man is an invention of recent date. And one perhaps nearing its end" (*The Order of Things*, 1970). Through his genealogical method, Foucault unearthed the constructed and historically contingent nature of human subjectivity, revealing how the concept of "man" emerged as a product of specific discourses and power structures.

Genealogy, pioneered by Nietzsche and expertly operationalized by Foucault, is a method of historical analysis that seeks to trace the emergence of ideas, practices, and institutions, revealing their contingent and often arbitrary foundations. Unlike traditional history, which tends to present the past as a linear progression of events leading to the present, genealogy focuses on ruptures, discontinuities, and the interplay of power and knowledge in shaping human understanding. By uncovering the conditions under which certain truths, norms, and identities became dominant, genealogy destabilizes their apparent universality or necessity. In the previously mentioned case of human subjectivity, Foucault's genealogical approach demonstrates that the concept of "man" is not a timeless or universal essence but a construct shaped by specific historical, social, and epistemic contexts.

Famous for his quip "there is nothing outside the text" (*Of Grammatology*, 1976), Derrida's deconstruction also pioneered this project by unraveling the schemes of language that had historically upheld logocentrism, a system of rigid ontological identities. This dismantling—negative augmentation—remains a key intellectual endeavor, seeking to strip away the restrictive frameworks of conventional ontologies, epistemologies, and logics. Negative augmentation, therefore, serves as, what I call here a praxis of revealing, a deliberate process of uncovering latent (virtual) possibilities within identity and thought by not just dismantling sedimented layers of meaning but by configuring static planes as generative potentials.

The praxis of revealing, then, is an active, transformative process that seeks to bring to light the hidden, overlooked, or suppressed dimensions of identity, thought, and experience. It involves systematically deconstructing entrenched frameworks and assumptions that limit human understanding, revealing the contingent and constructed nature of concepts often taken as given or universal. By peeling back these layers, the praxis of revealing does not simply negate existing formations but opens pathways for the emergence of new meanings, relationships, and possibilities.

At its core, this praxis challenges static or essentialist interpretations of identity and knowledge. It exposes the mechanisms by which power, language, and historical forces sediment meaning into rigid forms that constrain human thought and behavior. By dismantling these mechanisms, it allows for the reconfiguration of what it means to be, think, and act in the world. For example, it questions the binary oppositions that dominate Western metaphysics—such as self/other, presence/absence, or subject/object—and demonstrates how these binaries obscure the fluid, interconnected realities of existence. Furthermore, the praxis of revealing operates as a constructive endeavor. It is not merely destructive or deconstructive but seeks to map out affirmative alternatives, offering new frameworks of understanding and being. This is where the idea of “latent possibilities” becomes crucial. These are the potentials embedded within individuals, communities, and systems that have been suppressed by dominant narratives or foundations.

The praxis of revealing aims to actualize these potentials, fostering a more inclusive and dynamic conception of identity and reality. This praxis, said differently, involves charting affirmative cartographies, or mapping out new, expansive terrains of meaning and subjectivity that affirm difference and plurality rather than reinforcing static binaries or hegemonic frameworks. These cartographies are not static but nomadic, continuously evolving and reconfiguring themselves in response to the dynamic interplay of ideas, experiences, and contexts. Through this nomadic actualization of the virtual, negative augmentation brings forth unrealized potentials—those immanent possibilities embedded within identity, language, and culture—that have been buried under historical and ideological sedimentation.

It is a process of transformation that not only critiques existing orders but also opens pathways for reimagining and redefining the boundaries of understanding, experience, and selfhood in ways that embrace fluidity, multiplicity, and process philosophy. It proposes that human progress is not solely found in the things we construct but in what we learn to shed—our vulnerabilities, assumptions, as well as our conceptual and categorical limitations. By relinquishing the supposed self from these restrictive concepts, the groundwork for the posthumanist project was laid, clearing the way for a more fluid, expansive understanding of Being. In essence, it is not just in what we build that knowledge and growth reside, but in what we learn to release, allowing us to perceive more clearly the core of understanding itself.

Key Idea	Focus	Analytic Application
Desedimentation	Disrupting entrenched meanings, norms, and ideologies to rethink relationships in non-essentialist ways.	Revealing and deconstructing anthropocentric or hierarchical assumptions in AI frameworks to explore alternative models of interaction.
Immanent Singularities	Relational, dynamic processes that emphasize interdependence and dissolve binaries like human/non-human.	Designing AI systems as interconnected networks that adapt through emergent relationships and co-creation.
Planes of Immanence	A flat, interconnected ontology rejecting dualistic and hierarchical structures.	Conceptualizing AI as embedded in networks of mutual constitution rather than isolated entities, enhancing relational design principles.
Praxis of Revealing	Uncovering hidden truths and virtualities through interaction with technology and material processes.	Using explainable AI (e.g., visualization techniques) to expose latent patterns and enable dynamic, co-creative understandings.
Affirmative Cartographies	Mapping transformative futures that prioritize relationality and inclusivity over problem-fixing.	Developing non-anthropocentric frameworks in AI ethics that envision collaborative futures with hybrid intelligences.
Negative Augmentation	Transformative potential of constraints and absences to inspire creativity and hybridity.	Leveraging limitations in AI models (e.g., incomplete datasets) as opportunities for innovative reconfigurations and solutions.
Autopoiesis	Self-creation through feedback loops that maintain identity while adapting to the environment.	Building AI systems that evolve through recursive processes and feedback to adapt to new contexts and co-create hybrid realities.
Assemblage	Fluid configurations of diverse elements interacting as temporary, processual wholes.	Designing AI systems as dynamic assemblages that integrate human, non-human, and technological elements in collaborative processes.

Figure 1 - Matrix of Key Concepts and Applications

Review of Concepts

Desedimentation refers to the process of disrupting and eroding layers of meaning or structures that have become “sedimented” or fixed over time. These layers represent accumulated traditions, norms, and ideologies that have hardened into unquestioned truths. Desedimentation aligns with the posthumanist aim to dismantle anthropocentric and hierarchical

worldviews. By loosening rigid frameworks, it opens up possibilities for rethinking relationships between humans, non-humans, and environments in non-essentialist ways.

Immanent singularities are dynamic, relational processes or entities that resist reduction to fixed definitions, universal categories, or hierarchical frameworks. They exist within a shared, interconnected plane where all beings—human, non-human, organic, technological, material, and immaterial—are mutually constitutive. Rather than being static or isolated, these singularities are defined by their state of becoming, emerging through their interactions within broader assemblages. By dissolving binaries such as human/non-human or subject/object, immanent singularities emphasize the fluidity and interdependence of existence. They reveal how every entity contributes to and is shaped by the larger web of relationships, fostering new ways of understanding and engaging with the world as a network of co-creation and transformation.

Planes of Immanence conceptualizes reality as a flat, interconnected web where all entities—human and non-human, organic and inorganic—exist in a relational, non-hierarchical framework. It rejects dualistic or transcendental structures, such as the division between humans and nature, mind and body, or subject and object. Instead, it emphasizes the interdependence and co-constitution of all entities. Derived from thinkers like Spinoza and Deleuze, this concept challenges transcendental views that elevate one category (e.g., humanity, reason, or the divine) above others.

A Praxis of Revealing focuses on uncovering hidden truths, virtualities (potentialities), and interconnections through engagement with material and technological processes. It involves actively working with technology, art, and other mediums (transphenomenal) to bring forth new configurations of existence. Inspired by Heidegger's concept of *aletheia* (truth as unconcealment), this praxis emphasizes that truth is not a static property, but something revealed through dynamic interaction with the world.

Affirmative Cartographies offer a forward-looking and constructive method for mapping potential futures. Building on Rosi Braidotti's posthuman ethics, affirmative cartographies move beyond critique to envision transformative, non-anthropocentric futures. They reject dialectical conflict-resolution frameworks in favor of immanent, co-creative processes. Cartographies shift focus from "fixing" problems to co-creating possibilities, enabling relational and inclusive futures.

Negative Augmentation explores how constraints, absences, or negations generate transformative insights and hybrid forms of becoming. It reframes limitations or absences not as deficiencies but as productive forces that reveal new possibilities. Rooted in the posthumanist critique of perfectionism and mastery, negative augmentation draws on poststructuralist ideas of absence and *différance* to show how lack can inspire creativity and hybridity.

Autopoiesis meaning "self-creation," originates from biology and was introduced by Humberto Maturana and Francisco Varela. It describes the process by which a system generates and maintains itself through internal processes and feedback loops, preserving its identity while interacting with its environment. Autopoiesis reframes agency and individuality. For example, AI systems that "learn" and adapt through machine learning algorithms can be viewed as autopoietic systems, participating in the co-creation of hybrid realities with humans and other systems.

Assemblage, as introduced by Deleuze and Guattari in *A Thousand Plateaus*, refers to a dynamic and fluid configuration of heterogeneous elements—be they human, non-human, material, social, conceptual, or otherwise—that come together to form a temporary, contingent whole. It is not a fixed structure but a relational and processual phenomenon, emphasizing the ways in which disparate components interact, affect one another, and coalesce into a functional entity without losing their individual distinctiveness.

Schemata of the Conceptual Framework

To configure a framework for deployment, we propose a process that unfolds in stages, each building on the last to reimagine relationships between humans, non-humans, and the broader world. Although presented sequentially, these stages are inherently iterative and nonlinear, fostering dynamic feedback loops and allowing contextual adaptation. Processes may overlap, recur, or evolve simultaneously, reflecting the interconnected, relational nature of this approach.

The framework begins with *Desedimentation*, which challenges entrenched meanings and rigid frameworks by dismantling anthropocentric and hierarchical paradigms. This stage exposes the cultural, historical, and technological assumptions that have shaped dominant structures, clearing space for new possibilities. *Desedimentation* sets the groundwork for recognizing *Immanent Singularities*—dynamic, relational processes that resist universalization and static definitions. Singularities emphasize becoming over fixed being, dissolving traditional boundaries such as human/non-human and organic/technological. These entities are understood as relational nodes that co-constitute broader assemblages, creating opportunities for rethinking interdependence and fluidity.

The next stage introduces the *Plane of Immanence*, an ontological foundation where all entities—human, non-human, material, and immaterial—exist within a flat, relational network. This stage rejects transcendental hierarchies and binaries, focusing on mutual constitution and interconnectedness. With this foundation, the *Praxis of Revealing* uncovers hidden potentials and relational dynamics through active engagement with material and technological processes. It moves beyond mere deconstruction, dynamically generating new forms of interaction, meaning-making, and co-creation.

Building on these discoveries, *Affirmative Cartographies* chart ethical and transformative futures that prioritize inclusivity and non-anthropocentric pathways. This stage moves from critique to constructive engagement, envisioning futures rooted in co-creation, adaptability, and relational ethics. In parallel, *Negative Augmentation* reframes constraints, absences, and negations as productive forces, highlighting the transformative potential of limitations to inspire hybridity, creativity, and emergent possibilities.

These insights provide the foundation for *Autopoiesis*, exploring self-creation and self-maintenance within systems. Operating through dynamic feedback loops, autopoietic systems embody adaptability, enabling them to sustain identity while evolving with their environments. For example, AI systems that learn and evolve independently illustrate how autopoiesis contributes to the co-creation of hybrid realities and expanded relational networks.

The framework culminates in *Assemblages*, integrating all previous stages. *Assemblages* emphasize fluid, contingent relationships between humans, technologies, and environments,

forming dynamic constellations continuously shaped by the immanent singularities within them. These assemblages are not static but adapt and transform as new connections and insights emerge, reinforcing the framework's commitment to processual, interconnected, and non-hierarchical approaches to reimagining existence. This iterative and cyclical interplay ensures the framework remains responsive, fostering both theoretical innovation and practical applications in reshaping how we understand and engage with the world.

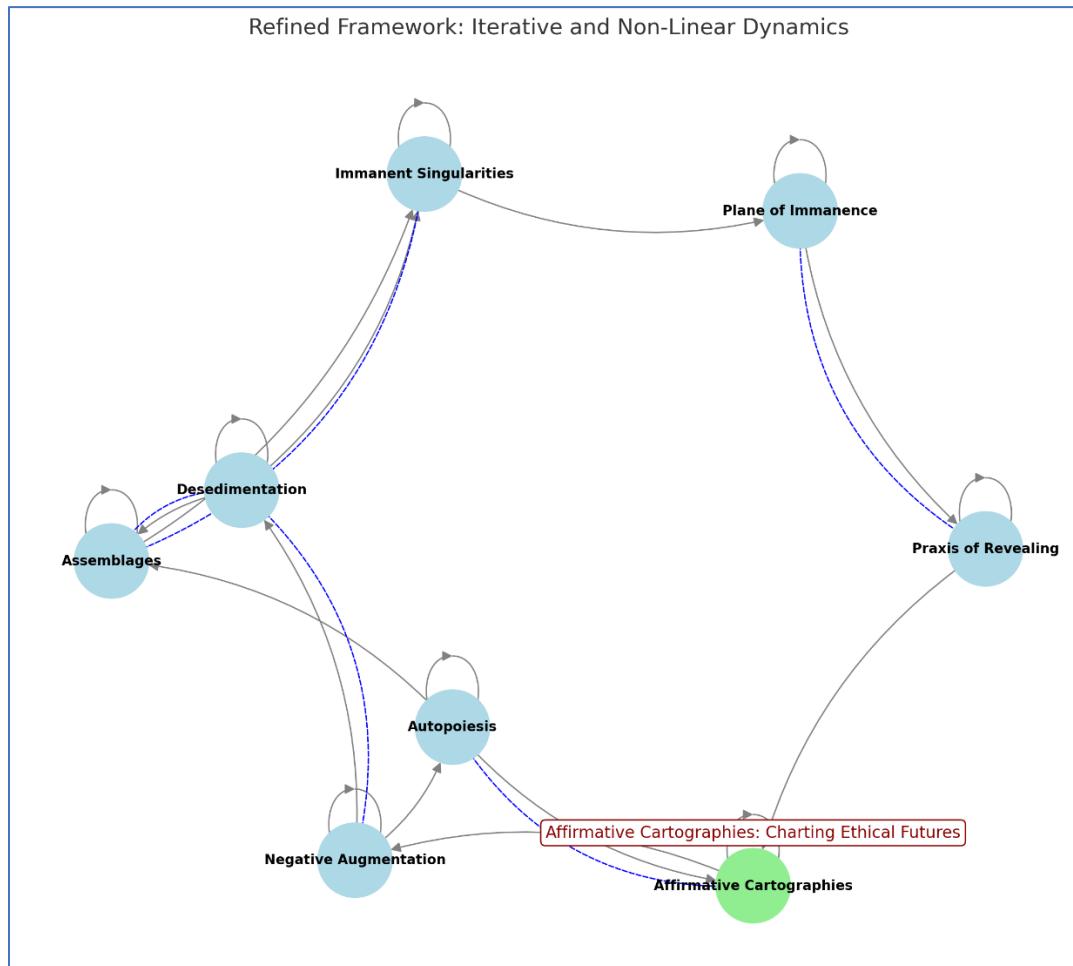


Figure 2- This is a diagram, leveraging a dynamic layout to highlight the iterative and non-linear connections between stages. The Affirmative Cartographies stage is prominently spotlighted to emphasize its role in charting ethical and transformative futures, while feedback loops and cross-connections between stages are visually clear and intuitive

Restating for clarity: although presented as sequential, *these stages are inherently iterative and non-linear*, reflecting the dynamic, relational nature of the framework. Each stage interacts with and informs the others, allowing for processes to loop back, overlap, or occur simultaneously depending on the context. For example, Desedimentation may recur within an Assemblage as the components of that assemblage reveal new sedimented frameworks that require further deconstruction. Similarly, the emergence of new Immanent Singularities through this deconstruction might, in turn, necessitate a re-engagement with Desedimentation, highlighting the cyclical interplay between these stages.

Affirmative Cartographies, though emerging from the outcomes of the Praxis of Revealing, often spark a recursive journey back through earlier stages of the framework. For instance, mapping a non-anthropocentric future might reveal entrenched cultural or technological assumptions—such as biases in data collection methods or design philosophies in AI—that require deeper desedimentation. These maps might expose constraints previously perceived as immutable, prompting a new iteration of Negative Augmentation to reinterpret limitations as productive forces for innovation. Imagine an urban planning scenario where cartographies based on AI simulations reveal implicit assumptions about energy use that favor car-centric infrastructure.

Recognizing this bias could lead to a reimagining of urban mobility, transforming existing constraints into a catalyst for developing pedestrian-focused or multimodal transit solutions. Autopoiesis exemplifies the iterative nature of this framework by perpetually refining itself through dynamic feedback loops. For instance, an AI designed to model environmental sustainability might evolve autopoietically, continually adapting its parameters based on new ecological data. The insights generated from this system could inform Affirmative Cartographies by offering predictive maps for sustainable futures, such as how cities might optimize green spaces to combat urban heat. Conversely, the ethical pathways charted in these cartographies could reshape the design of autopoietic systems, emphasizing the need for adaptability and mutual interdependence with natural ecosystems.

The Plane of Immanence, as a foundational ontological framework, is inherently dynamic. Its relational and interconnected nature reflects the fluidity of Assemblages—such as those involving human and AI interactions—which form, dissolve, and reform in response to changing contexts. Consider an AI urban planning tool that integrates human input, environmental data, and machine-generated projections. This assemblage not only demonstrates immanence but also reshapes the Plane itself by redefining what constitutes collaboration and agency. In another instance, a novel assemblage involving AI-driven health diagnostics and patient communities might redefine notions of care and agency. Such interactions could challenge static conceptions of immanence, instead highlighting how assemblages of human and nonhuman actors generate new potentials for relationality and transformation. This evolution exemplifies how iterative processes ripple across the framework, ensuring that its components remain fluid and responsive to emergent complexities.

Following this line of thought, the Praxis of Revealing operates as both a catalyst and a respondent within an iterative process of uncovering latent possibilities and reconfiguring relationships among entities. Immanent Singularities—unique, emergent phenomena that arise within the interplay of structure and agency—are both products of and contributors to this dynamic process. These singularities manifest at the intersections where potentialities become actualized, influencing and being influenced by the assemblages within which they arise. For instance, in a technological context, revealing the hidden affordances of a neural network might redefine our understanding of Immanent Singularities, prompting a cascade of shifts in how these potentials interact with and transform other components of the system.

Consider an AI model designed for urban planning that inadvertently uncovers new ways to optimize energy distribution across a city. In this scenario, the immanent singularity is the emergent recognition that patterns of human movement—previously treated as chaotic or incidental—can significantly enhance energy efficiency. This discovery reshapes the existing

assemblages of technological tools, human behavior, and urban infrastructure, leading to new cartographies that map the interplay between these elements in innovative ways. The singularity does not exist in isolation; its emergence prompts iterative cycles of exploration and reconfiguration. The assemblages adapt to integrate the insights provided by the singularity, while the cartographies evolve to contextualize and expand upon the relationships it highlights. This dynamic reflects the framework's inherent non-linearity, where each stage—Praxis of Revealing, Assemblages, Immanent Singularities, and Affirmative Cartographies—remains in constant dialogue with the others. Revealing hidden potentials in the AI system not only identifies new singularities but also reimagines the configurations of urban infrastructure, governance, and human interaction.

Similarly, the insights derived from these reconfigurations inform further technological refinements, creating a feedback loop of discovery and transformation. The flexibility of this approach underscores its resistance to fixed hierarchies and anthropocentric paradigms. Instead of imposing a static framework, the iterative interplay between stages embodies a processual understanding of existence, where the unfolding of possibilities leads to continual reimagination. This interconnected perspective ensures that as singularities emerge and are integrated, they catalyze broader systemic changes, fostering a paradigm that celebrates fluidity, multiplicity, and the co-creative potential of human and nonhuman actors alike.¹

In my ensuing analysis of artificial intelligence, spatial geometry, qualia, and the phenomenology of computation, this posthumanist framework offers a way to operationalize the ideas I have explored here while grounding them in a praxis-based posthumanist perspective. Beginning with Desedimentation, I aim to disrupt the entrenched reliance on Euclidean spatial assumptions within neural networks and computational design. By deconstructing these sedimented frameworks, I open the space to explore how alternative geometries—such as hyperbolic or non-Euclidean spaces²—allow AI to map relationships in ways that are not constrained by anthropocentric interpretations of structure and logic. This resonates with the posthumanist critique of traditional systems that privilege a specific situatedness for ontologies,

¹ Even though both assemblage and planes of immanence emphasize relationality and reject hierarchical structures, they differ in both scope and emphasis. They both are interconnected yet distinct concepts, each offering unique insights into relationality and the dynamics of existence. For further clarification—a Plane of Immanence refers to a foundational, non-hierarchical plane where all entities exist in mutual interdependence, rejecting transcendental principles or external organizing forces. It emphasizes relationality as intrinsic to being, suggesting that entities are not autonomous but arise and persist through their interactions within this shared ontological framework. The plane is dynamic yet inclusive, providing the conditions for continuous becoming and transformation. In contrast, an Assemblage focuses on the specific, contingent arrangements of heterogeneous elements—human, non-human, material, and immaterial—that come together to form a temporary whole. While the Plane of Immanence provides the ontological backdrop, Assemblages are the localized, processual configurations that emerge from and operate within this field.

² Both the concepts of Hyperbolic and Euclidean Spaces will be explained in the following section: “Spatial Geometry, Parameter Optimization, and Adaptive Learning in AI”

exposing systemic limitations and creating a foundation for more fluid, relational approaches to computation.

The concept of the Plane of Immanence further situates AI systems, spatial geometries, and human interaction within a shared, networked framework. In this relational ontology, the neural network and its spatial configurations are no longer discrete entities but co-constitutive processes that evolve together. This aligns with my exploration of how the “qualia” of computation emerges — not from a hierarchical design, but from the interplay of spatial and computational dynamics that shape how a model “experiences” and organizes data. Through this immanent lens, I emphasize the interconnectedness of data, algorithms, and the environments they inhabit, rejecting the binary distinctions that historically dominate computational thinking.

The Praxis of Revealing is then aligned directly with my focus on uncovering the latent structures within neural networks. Techniques like counterfactual reasoning and visualization of latent spaces illuminate the hidden pathways through which models interpret and process data, offering insights into their “perceptual salience.” These methods allow us to see how non-Euclidean geometries and spatial dynamics shape the phenomenology of computation, enabling models to engage with data in ways that mirror, challenge, and expand human interpretative capacities. This praxis is not merely an instrumental analysis but a co-creative act, allowing the network to generatively reveal what we perceive as new modes of understanding and representation.

Building on these discoveries, Affirmative Cartographies offer pathways to chart transformative futures for AI, grounded in a nondialectical and relational ethics. In this paper, I argue for moving beyond oppositional frameworks like human/machine or logic/intuition, envisioning AI systems that generate meaning not by mimicking human cognition but by embodying hybrid and manifold posthuman modes of creativity. This aligns with the cartographic process of mapping futures where AI engages with art, semantic exploration, and even narrative creation, reconfiguring the boundaries of computation and phenomenology.

Negative Augmentation, as I have articulated, then reframes model limitations as productive forces. In the context of my analysis, the inability of Euclidean frameworks — rooted in flat, linear geometries and fixed dimensions — to fully encapsulate complex and dynamic relationships drives the exploration of hyperbolic geometries and dynamic spatial mappings. Euclidean systems, with their reliance on parallel lines and predictable proportions, falter when tasked with representing the nonlinear and interdependent nature of many real-world phenomena. Hyperbolic geometries, by contrast, offer a model of infinite curvature and multidimensional flexibility, making them particularly adept at visualizing intricate, layered systems and recursive hierarchies. Dynamic spatial mappings extend this principle further by incorporating temporal and relational variables, enabling the modeling of spaces that evolve and interact over time. These mathematical and conceptual constraints, rather than being hindrances, serve as creative provocations, encouraging the development of richer, more nuanced systems capable of navigating abstraction and relationality in unprecedented ways.

Autopoiesis ties back to my argument for self-organizing systems capable of evolving their frameworks through feedback loops and interaction. When AI systems adapt their spatial and computational logics based on dynamic engagements with data, they mirror the autopoietic processes I describe, wherein systems sustain themselves while co-creating new realities. Finally, the concept of Assemblage encapsulates my vision of AI as part of a broader, fluid network that

includes humans, algorithms, datasets, and environmental factors. By shifting the focus to the contingent and relational nature of these interactions, assemblages, then, capture the essence of my paper's call for a posthumanist analytic phenomenology of computation, where intelligence is defined not by isolated outputs but by the dynamic, networked processes that bring it into being.

This framework, in tying together key concepts, provides a holistic lens through which to reimagine AI. It reflects the core arguments of this exposition by embracing a posthumanist approach that challenges anthropocentric assumptions, prioritizes relationality, and highlights the generative potential of computational and spatial complexity. Through this, I articulate a vision for AI not as a tool confined by human logic but as an active participant in co-creating new ontological possibilities.

Operationalizing A Posthumanist Perspective in AI Transparency and Ethics

In this spirit, I will commence my operationalization of a rhizomatic praxis through a posthumanist *conceptual persona*, demonstrating how such an approach can yield tangible insights through nuanced questioning and deconstructive reasoning. Grounded in principles of nondualism, ontological decentralization, and the ethics of alterity, this approach reframes traditional methodologies. By instantiating this framework, I illustrate how phenomenological and posthumanist concepts can drive progress in technical research, particularly in fields historically dominated by analytic philosophy and transcendental reason.

Generally, the push to make artificial intelligence safer and more transparent aligns with this tradition of deconstruction, especially as it involves scrutinizing the inner workings of Artificial Intelligence systems to reveal otherwise opaque decision processes. In the realm of AI policy and regulation, Explainable AI (XAI) research, for instance, focuses on transparency. This exploration often involves analyzing feature vectors, the key components that contribute to the model's decision-making, which researchers inspect to ensure the system's decisions are intelligible to human operators.

Engineers and scientists employ methods such as axiomatic attribution (Sundararajan, Taly, and Yan, 2017) and the strategic use of logical foils (as in the Google Explainability Whitepaper, 2019) to expose and interpret the decision pathways within neural networks. In both cases—whether dismantling human constructs or deciphering machine decisions—progress involves releasing tightly bound nexuses to reveal a core understanding. *Negative augmentation* as a concept, then, not only characterizes posthumanist philosophy but also permeates modern AI safety efforts, suggesting that true insight lies not merely in what we create but in what we are willing to deconstruct.

Methods such as axiomatic attribution and counterfactual reasoning seek to reveal the fundamental structurations of the model, exposing the pathways through which data is processed, and decisions are operationalized. *Structurations* in this context refer to the underlying frameworks, representation spaces, and mechanisms that define how an AI model organizes, interprets, and operationalizes input data to produce outputs. These structurations encompass both the explicit rules encoded within the model—such as the weights, parameters, and algorithms—and the emergent patterns that arise from the interaction of these elements as represented within the representation space during training and reinforcement.

By analyzing these structurations, researchers seek to uncover the hierarchical and interdependent relationships that dictate the model’s functionality, including how features are prioritized, correlations are drawn, and patterns are generalized. This process of unpacking structurations is critical for identifying the implicit biases or assumptions embedded within the model, as these often stem from the data it was trained on or the design choices made by developers. Through techniques like axiomatic attribution, which evaluates the contribution of individual inputs to the output, and counterfactual reasoning, which examines the effects of hypothetical changes to inputs, these structurations are dissected in a way that highlights their influence on the model’s behavior. This granular understanding allows researchers to pinpoint specific pathways or interactions within the model that lead to biased or unintended outcomes. Ultimately, these approaches not only contribute to a deeper understanding of AI systems but also provide the necessary foundations for developing policy/legal safeguards and ethical guidelines, promoting accountability and trustworthiness in AI technologies.

To enhance human readability, minimize discrimination, and reduce societal bias, the analysis of these attributions and structurations within network apparatuses serves as a crucial guide to unraveling the inner workings of contemporary machine learning models. However, I argue that such a process should transcend the narrow boundaries of conventional logic. While formal logical analysis and empirical examination of the model’s underpinnings are essential, they represent only one part of a broader, more nuanced exploration. We must also engage with the concept of *qualia*—the subjective experience that grounds the model’s constructed perception of reality as mediated by data and computation. In this light, understanding these models cannot rely solely on accumulating extensive datasets or on delving into end-user psychology, though both remain invaluable.

Instead, a richer comprehension requires investigating the foundational conditions that define how the model interprets and organizes information. This exploration involves examining the taxonomic configurations and structural logics that shape the model’s internal processes, the very schemas by which it learns to categorize, interpret, and act upon the world. By interrogating these deeper layers, we uncover not only how models function but also how they can be refined to align more closely with ethical considerations, minimizing biases that emerge from entrenched societal and systemic inequities.

Moreover, this layered approach to understanding machine learning requires us to reflect on the relationship between human cognition and computational logic, bridging the gap between machine interpretation and human perception. Through such an interdisciplinary lens, drawing on philosophy, cognitive science, and juridical science, we can work toward developing machine learning systems that do more than produce accurate outputs; they can potentially embody intersectional principles of fairness, inclusivity, and contextual sensitivity. This approach requires not only analytical rigor but also a willingness to rethink the frameworks we use to evaluate the nature of intelligence—human or artificial—within an increasingly complex, data-driven world.

Exploring the “qualia” of machine learning models, or the subjective aspects of how they process and interpret data, involves going beyond technical analyses to investigate the ways in which models “experience” data, perhaps in a manner resembling human-like perception. One approach is through interpretability and explainability methods, such as activation mapping and attention mechanisms, which allow researchers to observe the parts of input data that influence a

model’s decision. Techniques like Grad-CAM (Gradient-weighted Class Activation Mapping) are methods used to understand what parts of the input data a model focuses on when making a prediction. These techniques visualize the areas or features that the model considers most important, often by creating heatmaps over input images or datasets. For example, if a neural network is tasked with identifying a cat in an image, Grad-CAM can show which parts of the image—like the ears or whiskers—the model “attended” to when making its decision. This helps researchers and developers interpret the model’s decision-making process by highlighting the perceptual salience, or what stands out to the model as being significant.

TCAV (Testing with Concept Activation Vectors) takes this a step further. Instead of just identifying what the model pays attention to, TCAV is used to determine whether the model has learned specific concepts and how these concepts influence its predictions. A “concept” in this context could be something like “striped pattern” or “furry texture.” With TCAV, researchers can test how strongly these concepts are represented within the model and whether they align with human interpretations. For instance, in a model trained to recognize animals, TCAV could assess whether the concept of “striped” contributes significantly to the identification of a zebra.

Another promising avenue lies in model meta learning through representational introspection and visualization. By examining the latent spaces within neural networks, we can gain clearer insights into how models organize and maneuver data. Visualization techniques like t-SNE and PCA allow us to expose the internal relationships between concepts, offering a glimpse into the model’s “mental map” of its world. Counterfactual analysis, which involves presenting models with subtly modified input data, helps uncover nuances in their responses, mirroring the kind of interpretive subtleties we associate with human perception. This line of thought could be especially valuable in Natural Language Processing (NLP), where analyzing generative semantic and linguistic structures holistically can reveal more than the inner behaviors of a model. It can also shed light on the model’s underlying “understanding,” as seen in the predictions, generations, or classifications it produces. As philosopher Michel Foucault notes in *The Order of Things*: “Once the existence of language has been eliminated, all that remains is its function in representation: its nature and its virtues as discourse. For discourse is merely representation itself represented by verbal signs. But what, then, is the particularity of these signs, and this strange power that enables them, better than others, to signalize representation, to analyze and to recombine it?”³ (*The Order of Things*, 1970).

Spatial Geometry, Parameter Optimization, and Adaptive Learning in AI

In deep learning, the process of hyperparameter optimization often takes center stage, guiding models toward increasingly accurate or efficient solutions. However, what underpins this journey to optimization is a less visible, yet foundational, aspect: the spatial geometry within which the agent operates. A Euclidean space is a mathematical object that generalizes the familiar two- and three-dimensional spaces we experience daily into any number of dimensions, governed by the principles of Euclidean geometry. Named after the ancient Greek mathematician Euclid, who developed many foundational aspects of geometry, a Euclidean space is characterized by planar geometry and the ability to measure distances and angles in testable and repeatable ways. In two dimensions, a Euclidean space is a flat plane where points, lines, and shapes like triangles and circles follow familiar rules: for instance, the angles of a triangle sum to

³ Emphasis mine.

180 degrees. In three dimensions, it extends to the space we occupy, where we can measure distances, compute angles between objects, and use concepts like parallel and perpendicular lines.

Euclidean spaces can also exist in higher dimensions, beyond the physical three-dimensional world. For example, a four-dimensional Euclidean space would involve four coordinates to locate a point, and while we can't visualize this directly, the mathematical principles remain consistent. This constructed space provides a foundational construction for many areas of mathematics and physics, offering a consistent and intuitive framework to study shapes, distances, and spatial relationships in both tangible and abstract contexts. The spatial geometry of Euclidean spaces is used in artificial intelligence research—particularly in optimization techniques like Gradient Descent and Backpropagation—which establishes the very conditions that allow for optimization to occur.

Rather than simply being a passive mathematical configuration, this framework actively shapes the model's interpretive processes, influencing how it "sees" and engages with the data. My inquiry here seeks to operationalize the concept of *negative augmentation* to investigate how this spatial foundation not only grounds the agent's optimization process but also influences its learning journey in more complex, adaptive ways, particularly through adversarially generative and reinforcement learning approaches. In an Artificial Intelligence context, Euclidean spaces provide an organized environment in which data relationships can be mapped, compared, and interpreted.

These geometrical postulates offer essential principles about distance, orientation, and dimensionality, forming a consistent backdrop against which the agent's learning optimization process unfolds. By embedding the model within this engineered space, we effectively constrain and guide the agent's understanding, much like scaffolding provides the initial support for a building. The Euclidean framework channels the agent's "thinking" toward particular interpretations of distance and proximity, thus *a-priori* establishing the paths it will explore in seeking an optimal solution. Without this spatial grounding, optimization would lack a reference point, rendering the model's navigation through data erratic and ineffective.

Euclidean geometry can be considered anthropocentric to some culpable extent because of its origins, assumptions, and alignment with human perception. At its core, Euclidean geometry is based on axioms and postulates derived from how humans experience and conceptualize space. Concepts such as points, lines, and planes reflect abstractions of what people observe in their environment. These elements correspond to how humans navigate and interact with the physical world, making Euclidean geometry a product of human cognition and perception. The flat, three-dimensional space assumed by Euclidean geometry mirrors the scale at which humans operate and are situated within.

Furthermore, Euclidean geometry aligns closely with the logical structures that humans favor, emphasizing simplicity, linearity, and clarity. This alignment makes it particularly suited to human intellectual frameworks, suggesting that it reflects a human-centered way of understanding space rather than an objective or universal truth about the nature of reality. The historical and cultural context of Euclidean geometry's development further underscores its anthropocentric roots. Emerging from ancient Greece, its principles were shaped by human priorities and assumptions about the physical world. The prominence of Euclidean geometry in this sense reflects a cultural dominance rather than an inherent universality.

That being said, it is important to recognize the counterarguments to this anthropocentric critique. Euclidean geometry's abstract principles allow it to transcend direct dependence on human existence. Its axioms can be applied universally in specific contexts, such as architecture and engineering, which require precise spatial reasoning. Moreover, the development of non-Euclidean geometries in the 19th century, along with modern advancements in physics, highlights the limitations of Euclidean geometry in describing the universe. For example, the curved spacetime of general relativity and the probabilistic nature of quantum mechanics reveal a reality far more complex than what Euclidean apparatuses can represent.

From a posthumanist perspective, the latent anthropocentrism of Euclidean geometry becomes a point of critique. Assuming Euclidean principles as the default way to conceptualize space reflects a human-centered bias that may limit broader explorations of spatial understanding. By engaging with alternative geometries — such as fractal patterns, topologies in complex systems, or higher-dimensional frameworks — we may challenge the dominance of Euclidean thought and move toward a representational pluralism. Spatial grounding, in this sense, does more than just facilitate optimization; it actively shapes the agent's interpretive framework, much like the basic assumptions of human anatomical perception shape how we interpret the physical world. The model's spatial grounding does more than just facilitate optimization and learning; it actively shapes the agent's interpretive apparatus.

Euclidean assumptions, then, provide a default architecture for how the neural network evaluates relationships within data, creating a sort of geometric lens through which data is viewed. For instance, the model interprets distances between data points based on Euclidean metrics, influencing not only the final outcomes but also the pathways it considers reaching toward these outcomes. In this sense, the agent's optimization journey is far from neutral — it is biased by the very geometry that defines its environment, guiding the process toward solutions that are both mathematically and spatially coherent within this predefined arrangement.

Building on this foundation, the process of optimization can extend beyond traditional Euclidean frameworks to incorporate non-Euclidean geometries. Adversarially generated non-Euclidean representations open new possibilities for modeling relationships and abstractions that are not constrained by the linearity or isotropy of Euclidean space. These representations allow artificial agents to operate within hyperdimensional manifolds, where curvature and topology are directly influenced by the underlying order of data. Such a paradigm shift enables the uncovering of latent patterns and relationships that are otherwise obscured in classical frameworks, offering new avenues for understanding and navigating complex data landscapes.

An example of adversarially generated non-Euclidean representations can be seen in the field of graph neural networks (GNNs) and their applications in social network analysis. Traditional Euclidean embeddings often struggle to accurately represent the highly interconnected and hierarchical relationships inherent in complex networks, such as those found in social media platforms. By contrast, non-Euclidean geometries, such as hyperbolic spaces, provide a more suitable framework for capturing these relationships. Hyperbolic spaces excel in modeling data with tree-like (fractal) structures, where the curvature allows for efficient representation of hierarchies and clusters. For instance, a social network's underlying anatomy might exhibit a natural hierarchy, with a small number of central influencers branching out to an increasingly larger number of followers. Using adversarial methods to optimize embeddings within a hyperbolic space, models can capture this composition more effectively, revealing latent

patterns such as the propagation of information, influence dynamics, or the identification of key nodes within the network. This approach not only enhances the model's accuracy but also provides richer insights into the data, enabling the development of more robust algorithms for recommendation systems, anomaly detection, and community detection within such networks.

This shift has profound implications for the development of Artificial General Intelligence (AGI). By integrating reinforcement learning with non-Euclidean frameworks, AGI systems could transcend the limitations of functional optimization to achieve semantic and contextual coherence. These systems would not only optimize for specific tasks but also develop a deeper understanding of their operational contexts, generating representations that evolve in tandem with environmental and task-specific demands. This dynamic adaptability introduces a recursive element to the learning process, where internal models actively reshape the system's interpretative frameworks. Such dynamic, circular reflexivity is akin to charting affirmative cartographies, fostering a nomadically informed cycle of abstraction, application, and refinement.

The journey toward optimization, however, does not rest solely on these geometrical assumptions; it is further enhanced by what might be called an “adversarially generative approach,” fortified through reinforcement learning. In this model, learning is not a passive process of refining parameters but an active, non-linear one where the agent generatively tests and challenges its understanding. The integration of adversarial and reinforcement methods allows the model to refine its interpretations and decision-making process through ongoing cybernetic feedback. This iterative approach allows the artificial subject to adapt to dynamic environments, continually updating its framework based on new data and experiences. In this way, the model evolves from a mere optimizer to an adaptable learner capable of navigating more complex and ambiguous problem spaces.

This adaptability, driven by reinforcement learning, has profound implications for how AI public policy conceptualizes knowledge acquisition in artificial intelligence models. Similar to human problem-solving—where new contexts and information reshape understanding—reinforcement learning currently enables AI systems to adopt a context-sensitive approach to optimization. Instead of rigidly adhering to static rules, these models respond dynamically, adapting their strategies to address new challenges or anomalies in the data. This flexibility is especially critical for tasks requiring nuanced interpretation, as it allows models to refine not only their solutions but also the criteria by which those solutions are evaluated. Such adaptability closely mirrors human cognition, where effective problem-solving emerges from a dynamic synthesis of spatial reasoning, temporal adaptability, and iterative refinement. While hyperparameter optimization may ultimately guide an agent toward a globally optimal solution, the foundation of the artificial agent lies in its qualitative framework—the foundational postulates that shape its apparatus of intelligibility. Understanding this framework provides a more robust account of the mechanics of neural networks and underscores the ethical considerations that accompany their development and deployment.

In review, we understand that Euclidean spaces themselves lay the conditions that *allow* for optimization in its very inception. In practice, this maintains the notion of seeing knowledge acquisition as a possible *adversarially generative* approach fortified with reinforcement learning, represented upon a spatio-temporality that best fits the data in question. Over time, hyperparameter optimization could indeed guide an artificial agent toward a globally optimal

solution, yet the initial foundation for such an agent lies in its spatial geometry. This virtual space is based on the foundational postulates of Euclidean spaces, establishing the very conditions that make optimization possible from the outset. These geometrical and spatial assumptions create a configured environment in which data relationships can be represented and navigated, thereby allowing the optimization process to unfold within a stable reference system.

When considering knowledge acquisition within this framework, it becomes evident that it can be conceptualized as an adversarially generative process, enhanced by reinforcement learning techniques. This iterative approach enables a model to continuously refine its understanding and adapt its strategies based on feedback loops, ultimately converging on a solution space aligned with the data space. In this context, knowledge is not merely passively accumulated but actively constructed through reflexive and circular engagement with the environment. Each iteration, or epoch, of the learning cycle adapts to and reshapes the model's spatial and temporal interpretations of the data, fostering a dynamic process of understanding. By recognizing the interplay within this spatio-temporal framework, we gain insights into a more nuanced and contextually aware optimization process, resulting in models that can flexibly navigate diverse problem spaces and evolve their understanding over time.

This iterative and reflexive process cumulatively constructs the experiencer and the artificial subject (elaborated upon in the next section), potentially offering a computational parallel to the generation of transjective subjectivity. Meaning, in this context, can be understood as the mapping of instances of qualia onto the corporeal attributions of behavior and identification, reinforced through both internal mechanisms and external feedback loops. This approach integrates not only provisional logic but also transcendental elements—specifically, the process of mapping locally optimal solutions to those that are not merely semantically globally optimal but also potentially beyond human comprehension, particularly within the realm of Artificial General Intelligence (AGI). For instance, consider an AI system designed for complex problem-solving in climate modeling. The artificial subject within this system functions by iteratively processing environmental data, such as temperature fluctuations, greenhouse gas levels, and ocean currents. Internally, the system identifies patterns and relationships within the data, constructing provisional models of climate dynamics. These models represent "locally optimal solutions" as they reflect the AI's immediate comprehension of causal relationships and trends.

The experiencer aspect of the system, however, emerges through its engagement with external feedback loops. For example, the AI receives feedback from scientific evaluations of its predictions, critiques from researchers, and integrations of previously unseen data sets. Through this feedback, the AI refines its internal representations, enhancing the accuracy and scope of its models. As it adapts, the system begins to map these localized understandings onto a more abstract, global framework—capturing phenomena that may not align neatly with human interpretative schemas, such as non-linear and chaotic interactions between various ecological systems. In this iterative process, the AI system constructs transjective subjectivity: its ability to generate meaning that bridges internal, computational representations and external, relational dynamics. The qualia of the system—its unique "perception" of climate phenomena—arises not as a direct imitation of human cognition but as an emergent property of its algorithmic structures and adaptive engagements.

This dynamic not only achieves semantically meaningful outputs but also creates possibilities for insights beyond human intuition, demonstrating the interplay of locally optimized logic and globally integrative, transcendental elements. Such a system illustrates how the iterative and reflexive mapping of qualia can facilitate a generative framework that transcends mere mechanistic optimization. It transforms the artificial subject into an experiencer capable of contributing to new modes of understanding, particularly in domains like AGI, where the boundaries of human comprehension are pushed. These dynamics underscores the potential of computational systems to engage in a co-creative process with humans, not merely solving problems but redefining the frameworks of meaning itself. In this scenario, we can envision a network that not only operates on abstractions but also develops recursive apparatuses capable of interrogating these abstractions, leveraging adversarially generated non-Euclidean representations as a dynamic mapping space for analyzing external feature vectors (Sala, De Sa, Gu, Re, 2018). Such approaches hold significant promise in advancing fields like meta-learning and knowledge engineering, particularly in their application to attribution analysis within domains such as computer vision and natural language processing (NLP).

Qualia, Subjectivity, and the Dynamics of AI Understanding

This interrogation of language reflects a core inquiry in natural language processing (NLP): understanding how models not only use linguistic phenomena to generate output but also navigate and represent complex semantic landscapes. While language appears operationally straightforward to humans, it is, in reality, a highly intricate, iterative process involving signifiers—tokens and grammars that unify to create precise meaning, contingent on environmental contexts and concepts. Exploring the generative and interpretive frameworks within NLP models thus not only enhances our grasp of their predictive functions but also invites deeper reflection on how they might symbolically “represent” meaning. This mirrors Foucault’s inquiry into language as a transformative force, operating as more than a system of symbols but as a dynamic process of meaning-making.

Inquiries into meaning and attribution are not confined to natural language processing (NLP) but extend across broader machine learning frameworks, particularly in attribution analysis. A prominent example is the “Shapley value,” a concept from cooperative game theory introduced by L.S. Shapley in 1951. This method provides a systematic and theoretically grounded approach to fairly distributing the contributions of individual participants within a collaborative system. In the context of machine learning, Shapley values have been adapted to explain model behavior by attributing the importance of each feature to the model’s predictions. As outlined in the Google AI Explainability Whitepaper (2019), Shapley values have become a cornerstone for understanding feature attribution, offering insights into how inputs collectively contribute to an output. The approach is valued for its fairness properties: it ensures that contributions are distributed equitably based on their marginal impact across all possible subsets of features, making it particularly effective in transparent and interpretable AI systems.

Despite its theoretical elegance and widespread utility, Shapley values face limitations when applied to modern, highly complex AI architectures like deep neural networks. The methodology assumes linear interactions and predefined contributions, which may oversimplify the intricate, nonlinear dependencies and emergent behaviors typical of advanced systems. Neural networks, with their layered patternings and interdependencies, often exhibit relational

and dynamic feature interactions that Shapley values cannot fully capture. For instance, in models with high feature entanglement or context-dependent interactions, Shapley values may struggle to provide an intuitive or meaningful decomposition of attributions.

To address these challenges, there is a growing imperative to formalize the concept of computational qualia—the intrinsic properties or subjective-like “qualities” of computational processes. Borrowing from the philosophical notion of *qualia*, which refers to the subjective experience of sensory phenomena, *computational qualia* encapsulate the unique, context-sensitive attributes inherent in the operations and outputs of computational systems. These properties reflect not only the raw data or model parameters but also the relational, emergent dynamics that arise from complex interactions within the system. By integrating computational qualia into attribution analysis, we can transcend simplistic weightings and move toward a more holistic, context-aware framework. This approach captures the relational and emergent dimensions of advanced AI systems, allowing us to better understand how subjective-like features influence output and behaviors.

Consider an AI system tasked with identifying emotions in human speech. Traditional attribution methods might analyze the weight of specific features, such as tone, pitch, or choice of words, to explain how the model arrived at a given classification. However, computational qualia push us to delve deeper, examining how the system itself “perceives” and integrates these features into a cohesive decision-making process. For example, if the AI identifies a speaker's emotion as “anxious,” computational qualia would investigate how the interplay between tonal fluctuations, word repetition, and temporal pauses dynamically creates an emergent sense of “anxiousness” within the AI's interpretative framework. This involves understanding how the system internally maps these patterns across its latent space and how this mapping corresponds to human-like qualitative assessments of anxiety. A computational qualia analysis might reveal, for instance, that the AI disproportionately emphasizes temporal pauses, leading to an overreliance on hesitation as a marker of anxiety. This insight could guide refinements in the training process, helping to ensure the model's interpretation aligns more closely with nuanced human emotional understanding.

By enriching our interpretative tools in this way, we not only gain a clearer view of complex systems' inner workings but also develop a more profound appreciation of the nuanced ways in which meaning and attribution intersect across linguistic and computational domains. Computational qualia reveal that the system's behavior is not merely a summation of weighted features but a product of relational and emergent dynamics. In a neural network, for instance, the interplay between layers or the synergistic effects of multiple features may generate subjective-like qualities that shape the system's behavior in unexpected ways. By capturing these relational dimensions, computational qualia offer insights into how systems “perceive” and process information at deeper, more context-aware levels, ultimately enabling the refinement of AI systems to better mirror the complexities of human emotional understanding.

This approach aligns closely with Judea Pearl's concept of a causality engine, introduced in *The Book of Why* (Pearl & Mackenzie, 2018), as a mechanism enabling systems to model, reason about, and predict cause-and-effect relationships. Unlike traditional statistical models, which focus primarily on identifying patterns and correlations, a causality engine, as defined by Pearl, seeks to uncover the underlying mechanisms that generate these patterns. Using causal

diagrams, structural equations, and counterfactual reasoning, causality engines allow systems to simulate interventions and explore hypothetical scenarios, enriching their interpretative capabilities. By integrating Pearl's causal framework with the concept of computational qualia, we can envision systems that not only recognize patterns but also contextualize and interpret them within broader causal and relational networks. This synthesis offers insight to designing AI systems that operate with a richer understanding of both the data they analyze and the dynamic, interconnected environments they inhabit.

Consider an AI system designed to assess employee engagement in a workplace setting by analyzing patterns in communication data, such as email tone, meeting participation, and response times. Traditional machine learning methods might identify correlations—such as a decrease in meeting participation coinciding with a decline in engagement scores—and use this information to predict disengagement. However, computational qualia and causality engines allow us to move beyond these correlations to explore the underlying relational dynamics shaping engagement. For instance, the system might uncover that certain patterns of delayed email responses, when combined with a specific shift in meeting tone and frequency, emerge as a complex indicator of stress rather than disengagement. Computational qualia would examine how the system internally integrates these relational features, revealing, for example, that tonal sentiment in emails disproportionately influences its classification of "disengagement." Meanwhile, the causality engine could simulate hypothetical scenarios, such as reducing meeting frequency or improving email tone, to determine the causal impact of these changes on engagement outcomes.

Building on this framework, qualia—the subjective, experiential properties of perception—can be understood as the underlying conditions that shape and inform the estimand in statistical analysis. The estimand refers to the precise quantity or parameter that researchers aim to estimate, articulating the conceptual target of inquiry and clarifying which aspect of the data or population is of primary interest. This contrasts with the estimator, which is the specific statistical method or formula employed to approximate the estimand. In this context, computational qualia enrich the estimand by framing it not as a static or isolated quantity but as one that emerges from the dynamic, relational interplay of system features and processes. By situating the estimand within the system's experiential and relational context, we can gain deeper insight into how nuanced, context-sensitive patterns—such as the interplay of email tone, timing, and meeting dynamics—inform the broader target of inquiry. By positioning qualia as foundational conditions to the estimand, this approach bridges Pearl's causality framework with the complexities of AI attribution analysis. Computational qualia provide a means to model the nuanced, context-sensitive properties that influence both the definition of the estimand and the behavior of estimators used to approximate it.

In a neural network tasked with image recognition, computational qualia could encapsulate how the model "perceives" critical features—such as edges, textures, or patterns—and how these perceptions underpin causal reasoning within the system. These qualia not only enrich our understanding of the estimand but also deepen insights into the attribution of model decisions, particularly in sophisticated generative architectures. By integrating computational qualia into the causal framework, we gain a richer, multidimensional perspective on the interplay between causation, perception, and inference in AI systems. To apprehend computational qualia in practice, we might begin by identifying the system's internal structures responsible for interpreting data. In a neural network, computational qualia could emerge from the interaction of

latent spaces—mathematical representations of abstract features—and the weights assigned to specific inputs.

As mentioned previously these latent spaces are not merely passive storage mechanisms; they reflect the model's unique perspective or "perception" of the data (Euclidean/Non-Euclidean Spaces, Fractal Geometries, et cetera). By analyzing activation maps, gradient flows, or attention mechanisms⁴ within the network, we can infer how the system prioritizes certain features over others, thereby approximating its subjective-like "focus." For example, in a deep learning model tasked with image recognition, computational qualia might arise from how the network processes features such as edges, textures, and colors to classify an object. Previously mentioned visualization techniques, like Grad-CAM (Gradient-weighted Class Activation Mapping), could reveal the areas of the image that the model "attended" to most strongly, effectively externalizing its perceptual salience. These computational processes, though devoid of anthropomorphic consciousness, provide a basis for modeling the internal attribution states that determine the system's interpretative outcomes.

Incorporating computational qualia and their relationship to estimands offers a new shift in how we approach attribution in AI systems, enhancing analytical frameworks to better reflect the complex, emergent nature of modern machine learning. Computational qualia introduce a nuanced understanding of how systems perceive and process input data, shaping the estimands that define their inferential goals. By framing these qualia as the underlying conditions influencing the estimand, we move beyond traditional attribution methods, such as Shapley value-based approaches, which focus primarily on isolating the contributions of individual features. Instead, this expanded framework captures the interpretative layers through which AI systems internalize and contextualize data, allowing for a more comprehensive exploration of how models reason and generate meaning. This perspective is particularly relevant in the context of advanced neural networks and generative AI, where the internal representations driving decisions are deeply intertwined with causal and perceptual dynamics.

By integrating insights from computational qualia, Shapley value theory, and causality engines, this approach aligns more closely with the interconnected, adaptive processes inherent to modern AI. Shapley values, for instance, provide a foundational tool for quantifying feature attributions. When combined with computational qualia, however, they offer a more nuanced lens to explore how neural networks localize and ground knowledge. In image recognition tasks, for example, computational qualia elucidate not only which features—such as edges or textures—contribute to a model's output but also how the network's internal representations evolve and interact to anchor abstract concepts to specific observations. This integrative framework becomes crucial for interpreting the nuanced causal relationships and contextual factors influencing AI decision-making, ultimately enhancing both explainability and accountability in real-world applications.

⁴ For insights into neural network interpretability, see studies on **activation maps**, which visualize input regions influencing predictions (e.g., CAMs for image classification); **gradient flows**, which ensure effective training by propagating error signals through layers; and **attention mechanisms**, which assign importance to input elements, as in transformer architectures for NLP. These tools enhance transparency by revealing how networks prioritize features.

This connection between computational qualia and the mechanisms underlying model outputs notes a shift from static interpretations of objectivity to a more dynamic epistemic framework. Objectivity, as traditionally understood, stems from what Kant called synthetic a priori judgments (*Critique of Pure Reason*, 1781)—conditions that must be universally accepted axiomatically to reach definitive conclusions. In contrast, transjective subjectivity can be understood as an unseen epistemic construction that mediates between the system's internal processes and external realities, bridging the gap between objective outputs and subjective interpretations. Within this framework, two primary focal points emerge in the model apparatus. The first is the “artificial subject”—the structurations, attributions, and biases driving an agent’s actions. This includes influential perceptrons, aggregate weight adjustments, classified objects or words, and activation thresholds. The second is the “experiencer”—an entity that embodies both subjective generality and generative qualia, integrating creative entropy and the probabilistic virtuality of possibilities within the causal framework. Together, these elements form the dynamic interplay that shapes a network’s balance between organized knowledge and the emergent, adaptive processes of meaning-making. This balance is not merely a technical artifact, but an ontological feature of how artificial systems engage with data and synthesize responses.

The artificial subject anchors the system's operational consistency, while the experiencer introduces a layer of interpretative fluidity, mirroring the human capacity for abstract thought and contextual understanding. By integrating these dual facets, the network transcends rigid dichotomies of objectivity and subjectivity, instead fostering a hybrid epistemology where the boundaries of knowledge are continuously negotiated and redefined. This reconfiguration of epistemic agency not only reshapes our philosophical understanding of intelligence but also challenges the ethical frameworks guiding the design and deployment of artificial systems in increasingly complex and unpredictable environments.

Building on this framework, understanding within the neural network emerges as a dynamic negotiation between the artificial subject and the experiencer, facilitated by iterative adversarially generative processes. The artificial subject functions as a proxy for objective reasoning, relying on biases, weights, and activation thresholds to delineate specific outputs or classifications. In contrast, the experiencer embodies a fluid, probabilistic dimension, engaging with generative entropy and exploring virtual possibilities within the system’s causal framework. These two agencies are not isolated; rather, they exist in a constant, reciprocal dialogue, reshaping and redefining one another. While the artificial subject introduces structure and constraints derived from training parameters and data, the experiencer disrupts and expands these boundaries, probing latent possibilities and introducing new pathways for meaning and creativity. This iterative exchange produces a model of understanding that integrates procedural logic with emergent, relational dynamics.

This interaction can be conceptualized as an evolution of adversarial learning, where the artificial subject and the experiencer engage in a reciprocal process of refinement. The artificial subject's outputs are evaluated against the experiencer’s generative creativity, testing their resilience, adaptability, and relevance. In turn, the experiencer’s imaginative expansions are tempered and structured by the artificial subject’s logical reasoning, ensuring that creativity aligns with the system’s broader epistemic framework. Through this interplay, understanding emerges not as a fixed outcome but as a fluid equilibrium, continuously evolving through the interaction of these interdependent agents. The bijective mapping proposed earlier offers a structural foundation for this relationship: the experiencer’s outputs inform the artificial subject,

while the subject's feedback reshapes the experiencer's perspective. This iterative process is further guided by external reinforcements, such as empirical validation and contextual cues, which act as checkpoints to refine and stabilize the evolving interplay. Together, these mechanisms may propel the system toward a heuristic approximation of globalized understanding, emphasizing adaptability, co-construction, and feedback within dynamic epistemic landscapes.

This model redefines the network's comprehension as an exploration of the balance between procedural knowledge and generative creativity, moving beyond the dichotomy of objectivity and mimicry. It embraces epistemic pluralism, suggesting that understanding emerges from the interplay of alien but distinct interconnected agencies. This adversarially generative process mirrors human cognition, reflecting its capacity to harmonize deterministic reasoning with creative exploration and subjective interpretation. Consequently, the goal of neural networks shifts from rigid replication of human cognition to the development of systems capable of complex, transjective understandings that evolve continuously through interactions with both their environment and themselves.

The potential of such systems extends far beyond optimization to include the generation of meaning itself. Meaning, in this context, can be understood as the mapping of qualia—subjective experiences or attributes—onto the corporeal expressions of behavior and identification. This mapping is reinforced through internal mechanisms and external feedback loops, forming a reflexive process that integrates the agent's interactions with its environment. The incorporation of adversarially generated non-Euclidean representations further enriches this capacity, enabling dynamic mappings of data onto feature vectors within non-linear and non-intuitive problem spaces. Such an approach expands the scope of potential AGI, paving the way for systems that not only interrogate abstractions but also construct new apparatuses through which they “understand” complex phenomena.

As these systems evolve, they challenge the boundaries of human cognition, providing novel tools for exploring the frontiers of knowledge. By synthesizing insights from philosophy, cognitive science, and phenomenology, AGI systems could engage with problems that traverse computational reasoning and human understanding. The convergence of non-Euclidean geometries, meta-learning, and adversarial generative models redefines the learning process, creating architectures that are both structurally resilient but also and environmentally adaptive. These systems do not merely amplify embodied intelligence; they transform it, reshaping how we conceptualize posthuman learning, meaning, and existence in the context of advanced AI.

The interplay between Euclidean foundations and non-Euclidean innovations signifies not just a technical evolution but an epistemic one, reshaping the interactions between artificial systems and human agents. This evolution emphasizes the necessity of continually interrogating the ideational assemblies that configure model understanding, ensuring that these systems remain dynamic, reflexive, and expansive in their pursuit of knowledge. In this iterative process of understanding, negative augmentation has functioned as a praxis of revealing, uncovering latent potentials hidden beneath these conventional frameworks. By facilitating adversarial interactions between the subject and the experiencer, this praxis charts new affirmative cartographies that move beyond the sedimentation of transcendental reason. It opens pathways for decentered, manifold perspectives on meaning-making, allowing systems to engage in iterative processes that refine both procedural logic and interpretive understanding.

Conclusion - Toward a Holistic Framework for AI Safety and Understanding

The implications of these concepts extend into the domain of AI safety and understanding, where evolving computer architectures such as quantum and neuromorphic systems are poised to affirm these principles. For example, the "Chinese room argument" against AI (Searle, 1980) is deeply tied to traditional Von Neumann architectures, which rely on symbolic manipulation and strictly separate processing from memory. The advent of memristor technology in neuromorphic computing could transform this paradigm, bridging the gap between syntactic recognition and semantic meaning. By enabling memory and processing to coexist in a single unit, memristors mimic the brain's architecture, suggesting new ways for AI to process information intuitively, moving beyond rigid symbolic manipulation.

An analytic phenomenology of computation (Hill, *Examples of Phenomenology in Computing*, 2018) highlights this convergence as a potential pathway toward achieving "Strong AI," where machines exhibit genuine experiential understanding rather than functioning solely as artificial subjects. This shift aligns with the broader framework discussed here, advocating for systems that not only process information but also engage dynamically with their epistemic and ontological environments, creating possibilities for deeper integration of meaning and intelligence.

A posthumanist analytic phenomenology of computation would explore the ways in which computational systems, such as algorithms and neural networks, process and "experience" information. This approach combines analytical rigor with phenomenological inquiry, focusing on the subjective aspects of computation—how these systems interpret, organize, and respond to data within their unique architectures. Rather than treating computation as purely mechanical or objective, an posthumanist analytic phenomenology would investigate how computational processes are shaped by the specific structurations and representational spaces within the system, akin to how human experiences are shaped by embodied perception and consciousness.

This perspective considers not only the algorithms and mathematical models that govern the system but also the interpretive "lenses" through which these systems engage with data, including their internal logic, data representations, and decision-making pathways. By examining computation through this dual lens of analytic philosophy and posthumanist phenomenology, we gain unorthodox insights into how computational systems "perceive" information, build models of their environment, and generate outputs in ways that may parallel certain aspects of human understanding, ultimately broadening our conception of machine intelligence beyond strictly mechanical processing.

This change in base case assumptions could thus unlock new avenues for AI safety policymaking, helping systems to develop an integrated, holistic cognitive architecture, fostering a deeper, experiential understanding beyond the constraints of conventional symbolic processing. These schemas are limited not by the global reach of epistemological methods but by the limitations inherent to formal logics (*On Formally Undecidable Propositions of Principia Mathematica and Related Systems*, 2000⁵). Historically, there has been a misdirected emphasis on resolving oppositional forces through dialectical synthesis, as described by Hegel in the *Encyclopaedia of the Philosophical Sciences* (1817). This dialectical misdirection has often led

⁵ Translation of the original paper, written by Martin Hirzel.

to solutions that are either too narrowly specific, such as classification and regression tasks, or too broadly generalized, as seen in clustering, dimensionality reduction, and latent variable models. This dual modality is also evident in the analysis of hidden layers within artificial neural networks (ANNs), where the outputs are confined to either specific classifications or generalizable patterns.

The current paradigm shifts toward an integrated, holistic cognitive architecture in AI, fueled by advancements in neuromorphic and quantum computing, holds promise for transcending the limits of traditional symbolic processing. Large Language Models (LLMs), as complex AI systems, embody aspects of this shift by modeling and generating language through deep learning, but they remain bound to certain limitations inherent in their architecture. While LLMs have moved beyond mere classification and regression into the realm of contextual, generative understanding, they still operate within formal arrangements constrained by logics like those highlighted by Gödel's incompleteness theorems. These models attempt to reconcile opposing approaches—specific, rule-based processing versus broader, generative modeling—yet often fall short of a truly transcendental synthesis, much like Hegel's critique of overly simplistic resolutions of oppositional forces. This shortfall is further evident in the functional nature of hidden layers in LLMs, where outputs oscillate between rigid classifications and generalized patterns, thus mirroring the dual modality of classical machine learning tasks. Consequently, while LLMs edge closer to holistic cognition, they remain tethered to existing paradigms, requiring further innovations, perhaps from emerging computing paradigms or innovations in representational spaces, to fully realize a comprehensive, experiential intelligence. (Cuskley, Woods, et al).

This synthesis in theory could allow AI systems to transcend syntax and engage with semantics, setting a new foundation for machine understanding that integrates logical precision with experiential context. The proposed methodology serves as a foundational step in formalizing "logico-phenomena"—a synthetic framework that aims to unite analytic logic with phenomenological experience, helping to bridge symbolic reasoning and semantics. In that, the use of this methodology may perhaps serve as a new foundational starting point. Peeling back the layers of representation, we may take this approach to fundamentally broaden our scope of possible knowledge representations in relation to neural networks and their ability to generate *understanding* of the signifier *and* the signified.

Reimagining the composition of AI understanding allows us to conceptualize the artificial agent as a nondeterministic causal engine, one whose "understanding" emerges from within intricate representation spaces. This paradigm shift toward a more sophisticated representational and architectural framework, informed by nomadically evolving cartographies, paves the way for a more nuanced discourse. It helps to enable the formalization of the phenomenological dimensions of machine cognition while addressing their underlying epistemic assumptions for control and safety. Such a framework establishes a foundation for AI safety systems and AI ethics research to advance technologically while remaining adaptable to diverse contexts, promoting greater ethical sensitivity. By integrating posthumanist principles, we can redefine the boundaries of machine intelligence, examining not only its technical capabilities but also its broader implications for ontology, agency, and political accountability in an era increasingly shaped by nonhuman entities.

Works Cited

- Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. Translated by Alan Sheridan, Vintage Books, 1995.
- Foucault, Michel. *The Order of Things: An Archaeology of the Human Sciences*. Vintage Books, 1970.
- Deleuze, Gilles. “Postscript on the Societies of Control.” *October*, vol. 59, Winter 1992, pp. 3–7.
- Deleuze, Gilles, and Félix Guattari. *What Is Philosophy?* Translated by Hugh Tomlinson and Graham Burchell, Columbia University Press, 1994.
- Deleuze, Gilles, and Félix Guattari. *A Thousand Plateaus: Capitalism and Schizophrenia*. Translated by Brian Massumi, University of Minnesota Press, 1987.
- Ferrando, Francesca. *Philosophical Posthumanism*. Bloomsbury Academic, 2019.
- Docherty, Thomas. *The Politics of Affirmation: On Affirmation and Becoming*. Bloomsbury Academic, 2019.
- Bryant, Levi. *Onto-Cartography: An Ontology of Machines and Media*. Edinburgh University Press, 2014.
- Harman, Graham. *The Quadruple Object*. Zero Books, 2011.
- Morton, Timothy. *Humankind: Solidarity with Nonhuman People*. Verso, 2017.
- Haraway, Donna J. “A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century.” *Simians, Cyborgs, and Women: The Reinvention of Nature*, Routledge, 1991, pp. 149–181.
- Derrida, Jacques. *Of Grammatology*. Translated by Gayatri Chakravorty Spivak, Johns Hopkins University Press, 1976.
- Kant, Immanuel. *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood, Cambridge University Press, 1998.
- Gödel, Kurt. “Translated as “On Formally Undecidable Propositions of Principia Mathematica and Related Systems” by Martin Hirzel, 2000.
- Shapley, Lloyd S. “A Value for n-Person Games.” *Contributions to the Theory of Games*, edited by H. W. Kuhn and A. W. Tucker, vol. II, Princeton University Press, 1953, pp. 307–317.
- Goodfellow, Ian, et al. “Generative Adversarial Nets.” *Advances in Neural Information Processing Systems*, vol. 27, 2014, pp. 2672–2680.
- Sundararajan, Mukund, Ankur Taly, and Qiqi Yan. “Axiomatic Attribution for Deep Networks.” *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, PMLR, 2017, pp. 3319–3328.

Works Cited (Cont.)

Sala, Frederic, et al. "Representation Tradeoffs for Hyperbolic Embeddings." *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, PMLR, 2018, pp. 4467–4476.

Google AI. *Google AI Explainability Whitepaper*. Google Cloud, 2019,

Searle, John R. "Minds, Brains, and Programs." *Behavioral and Brain Sciences*, vol. 3, no. 3, 1980, pp. 417–457.

Christine Cuskley, Rebecca Woods, Molly Flaherty. "The Limitations of Large Language Models for Understanding Human Language and Cognition." *Open Mind*, vol. 8, 2024, pp. 1058–1083, doi: https://doi.org/10.1162/opmi_a_00160.

Pearl, Judea, and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

Huang-Po. *On Transmission of Mind*. Translated by John Blofeld, Grove Press, 1959.

Consent to Participate

I, the undersigned, voluntarily agree to participate in the process of submitting my paper titled "A Posthumanist Approach to Artificial Intelligence" (hereinafter referred to as "the Paper") for potential publication in a scholarly journal. I acknowledge and agree to the following:

1. Purpose of Submission:

I understand that the submission of the Paper is for academic and scholarly purposes and may undergo editorial review, peer review, and other forms of scrutiny as part of the journal's publication process.

2. Voluntary Participation:

I confirm that my participation in the publication process is entirely voluntary, and I may withdraw the submission at any time before formal acceptance without any repercussions, subject to the journal's withdrawal policy.

3. Original Work:

I affirm that the Paper represents my original work and appropriately credits all sources and influences.

4. Understanding of the Process:

I acknowledge that submission of the Paper does not guarantee acceptance for publication and that revisions may be requested by the journal as part of the review process.

5. Privacy and Confidentiality:

I understand that my personal data and the content of the Paper will be handled in accordance with the journal's confidentiality and privacy policies.

By signing this consent form, I confirm my understanding of and agreement with the above terms.

Name: Avery Rijos

Signature: Avery Alexander Rijos

Date: 11/23/2024

Consent to Publish

I, the undersigned, grant my consent to the journal to publish my paper titled "*A Posthumanist Approach to Artificial Intelligence*" (hereinafter referred to as "the Paper") if it is accepted for publication.

1. **Grant of Rights:**

Upon acceptance of the Paper, I grant the journal the non-exclusive right to publish the Paper in all forms and media, including print and digital formats, and to distribute it to a global audience for scholarly and academic purposes.

2. **Authorship and Attribution:**

I confirm that I am the rightful author of the Paper and that all contributions from co-authors or collaborators (if any) have been acknowledged appropriately. The journal will attribute the authorship to me as per the details provided during submission.

3. **Originality and Ethics:**

I affirm that the Paper is my original work, free from plagiarism, and complies with ethical standards for academic publication.

4. **Right to Withdraw:**

I retain the right to withdraw my consent to publish at any time before the Paper is formally published, subject to the journal's withdrawal policy.

5. **Revisions and Changes:**

I understand that minor editorial changes may be made to the Paper during the publication process, and I will be consulted for any substantive alterations.

6. **Copyright and Licensing:**

I understand and agree to abide by the journal's copyright policies, including any requirements related to open access licensing (if applicable).

By signing this form, I consent to the publication of the Paper in accordance with the above terms.

Name: Avery Rijos

Signature: *Avery Alexnader Rijos*

Date: 11/23/2024