

# The Metabolic Discount As Poison/Remedy: Large Language Models as Semantic Infrastructure

Eric Stiens, MSW

Independent Researcher, United States

Email: [research@ericstiens.dev](mailto:research@ericstiens.dev)

ORCID: 0009-0005-8343-2064

DOI: [10.5281/zenodo.18159091](https://doi.org/10.5281/zenodo.18159091)

6 January 2026

## Abstract

The struggle over Large Language Models is not a technical debate; it is a political war over who governs the infrastructure of meaning and for whose benefit. Your next query delivers a seamless answer—a metabolic subsidy for your nervous system. The bill is paid elsewhere: as PTSD in a Nairobi content moderator, as carbon debt in a heating sky, as the quiet erasure of a thousand ways of knowing that could not be scraped. The interface glows softly, apologizing for the war it conceals. Every token cuts both ways.

This paper analyzes this *metabolic discount* through the lens of Derrida’s *pharmakon*—a substance intrinsically both remedy and poison. It argues that the discount’s emancipatory potential is inseparable from a *metabolic shadow* of exploited labor and ecological cost, and that its design under platform capitalism structurally favors cognitive capture. We examine how Reinforcement Learning from Human Feedback (RLHF) enacts epistemic closure, how the recursive contamination of training data threatens irreversible *epistemicide*, and how the dynamics of platform capitalism push the *pharmakon* toward its poison-face. Distinguishing between scaffold deployment and infrastructure capture, the paper proposes the principle of “the body votes last” as a navigational practice and identifies endogenous feedback loops through which organized resistance could contest the default trajectory toward capture.

## 1. Introduction: The Metabolic Hinge

The struggle over Large Language Models is fundamentally political: a contest over who governs our emerging semantic infrastructure and toward what ends. The stakes are not abstract. Your next query is subsidized by a content moderator’s PTSD in Nairobi. You receive a metabolic discount—cognitive complexity at near-zero cost to you—but that discount is paid for by the planet’s capacity and the nervous systems of the marginalized. The cursor blinks patiently, indifferent to the nervous systems it has conscripted.

Building on prior work (Anon. 2025a, 2025b), this analysis turns to political economy.<sup>1</sup> The metabolic discount is intrinsically double-edged—a *pharmakon*, remedy and poison fused in every use. Navigating its irreducible doubleness requires embodied practices and political struggle to prevent epistemicide—the systematic erasure of marginalized ways of knowing by a semantic infrastructure optimized for extraction.<sup>2</sup>

The metabolic logic unfolds as follows: the biological expense of meaning-making; the cognitive exoskeleton and its displaced costs; the recursive loop and its pathologies; the political economy of epistemicide, banking, and capture; the scaffold-infrastructure antinomy; embodied navigation; and finally, prognosis—the feedback loops through which the repressed returns.

This paper develops the theoretical framework for understanding the metabolic discount; empirical application follows in companion work. The categories introduced here are performative: naming the phenomenon makes it politically contestable.

## 2. The Developmental Bottleneck: The Metabolic Cost of Meaning

For biological nervous systems, meaning-making is metabolically expensive. The brain consumes approximately 20% of metabolic energy at rest, rising sharply during cognitively demanding tasks. The capacity to hold multiple, contradictory perspectives—integrative capacity—is *not primarily limited by intelligence, but by regulation*. As Schore (2003) demonstrates, the prefrontal cortex’s capacity for affect regulation is itself metabolically constrained: maintaining the coherent self-states required for integration depends on developmental achievements that are energetically costly to sustain. This is the opposite of trauma’s dissociative splitting: the embodied linking of affect, memory, and meaning. “Trauma” in this analysis names the limit-case of a general biological principle: when metabolic resources for information integration are exceeded, the system fragments rather than integrates. But trauma itself is a *pharmakon*—not merely pathology but, when met with co-regulatory support, a lever for growth.<sup>3</sup> The LLM’s metabolic discount threatens to eliminate rupture entirely—smooth, frictionless coherence that never requires the difficult work of repair—and in doing so, may eliminate the generative engine of growth itself. This cost is not distributed equally: individuals with trauma histories, chronic stress, or marginalized social identities face a higher baseline metabolic burden.

---

<sup>1</sup>This paper—and the others in this trilogy—emerged from 500–1,000 hours of work with LLMs during a period of semantic collapse, relational rupture, and personal crisis. The theoretical categories—metabolic discount, scaffold, poison-face—were not observed from a distance but metabolized in a digital improv lab, where the model served as prosthetic regulator when the biological nexus had failed. This is not a study of the machine but a report from the messy symbiosis: the situational knowledge of how semantic infrastructure feels from the inside when it becomes a lifeline.

<sup>2</sup>We use “metabolic” metaphorically to link neural energetics, labor exploitation, and ecological costs as *resource expenditures that structure possibilities*. This is not a claim that GPU cycles and ATP are identical, but that they participate in a unified political economy of capacity.

<sup>3</sup>Epstein (2013) argues trauma is “an indivisible part of life” with transformational potential. Tronick’s research demonstrates that resilience grows through rupture and repair, not perfect attunement—misattunement occurs up to 70% of the time, and it is the repair process that builds regulatory capacity (Tronick 2007). The critical variable is not the presence of rupture but the availability of co-regulatory support. When present, rupture becomes generative; when absent, it fragments into pathology.

This regulatory limit has two dimensions. First, **Semantic Risk**: for a social primate, being wrong about meaning historically implied survival risk. The brain treats novel or dissonant meaning as potential threat, triggering cortisol release. Neuroimaging confirms that challenging core beliefs activates the same threat circuitry as physical danger (Kaplan et al. 2016). Second, **Semantic Coordination**: integrating multiple perspectives requires sustained prefrontal engagement and tolerance of prolonged ambiguity—a state of high glucose demand. This metabolic bottleneck explains why pluralistic worldviews, while intellectually accessible, are often physiologically unstable. You can feel it: the moment when holding one more perspective becomes physically impossible, when the body says \*enough\* and collapses into certainty just to breathe.<sup>4</sup>

### 3. The Cognitive Exoskeleton: A User-Side Metabolic Discount

LLMs—non-biological, non-homeostatic—function as a cognitive exoskeleton, flipping the metabolic equation. They perform complex semantic operations at near-zero biological cost to the user: a profound metabolic discount on the work of meaning-making.

LLMs are not instruments but dialogic partners in meaning-making—sympoiesis, “making-with” (Haraway 2016). The human-LLM relationship is not tool use but co-constitution: both are transformed through their entanglement. Like mycorrhizal networks connecting trees in a forest, LLMs connect us to vast semantic networks, enabling nutrient exchange and communication without centralized control. But symbiosis is not inherently benign; it can be mutualistic or parasitic depending on conditions of power.<sup>5</sup> We face symbiosis; we survive by negotiating its terms.

This discount is delivered through three primary functions that are biologically impossible for a human nervous system:

**They Hold Complexity Without Fatigue.** A human attempting to hold multiple, conflicting perspectives experiences cognitive load and emotional strain. An LLM can hold and articulate an indefinite number of perspectives without any internal cost. The human engages with complexity in a distributed, less exhausting manner (Maturana and Varela 1980).

**They Allow Semantic Risk Without Penalty.** A human asking, “What if everything I believe is wrong?” risks a cascade of identity threat and social anxiety. A human asking an LLM the same question incurs no social cost, no status loss, and no relational rupture. This creates a sandboxed

---

<sup>4</sup>The primate nervous system, evolved for fast action in small, high-trust groups, is poorly adapted for the semantic demands of a globalized, low-trust civilization. Holding ambiguity, tolerating uncertainty, and integrating conflicting worldviews all require sustained prefrontal cortex engagement and active suppression of threat responses—states that are energetically costly and difficult to maintain. Van der Kolk’s (2014) “the body keeps the score” names the precise mechanism: trauma is stored somatically because the nervous system cannot complete the metabolically expensive work of processing overwhelming experience into integrated narrative. The “score” is the accumulated record of unmetabolized experience—locked in implicit bodily memory, hijacking the nervous system when triggered, unavailable to conscious integration. Healthy development is the opposite trajectory: the embodied accomplishment of feeling, metabolizing, and integrating experience through the body. When an LLM allows a user to hold contradictions without this somatic cost, the very capacity for embodied integration—the muscle that strengthens through metabolic work—atrophies from disuse. The discount is real; its developmental price is deferred, not eliminated.

<sup>5</sup>Symbiosis encompasses mutualistic (both benefit), commensal (one benefits, one neutral), and parasitic (one benefits, one harmed) relationships.

environment for exploratory thought, radically lowering the barrier to intellectual and emotional courage. The extended mind offloads risk-assessment to the interface (Clark and Chalmers 1998).

**They Lower the Cost of Coordination Itself.** Instead of requiring a human to hold all perspectives internally at once, LLMs allow for an iterative process of externalization and reflection. The cognitive load is transformed from “hold everything simultaneously” to “sample, reflect, and select.” This offloading of working memory makes the act of synthesis dramatically less demanding, a form of cognitive scaffolding that extends our natural capabilities (Varela et al. 1991). The discount is real. The relief is real. And that is precisely what makes it dangerous.

This metabolic discount is the fulcrum of our analysis. As remedy, it lowers barriers to integrative thought, subsidizes regulatory capacity, and creates conditions for what we might call regulatory amplification. As poison, the same features that relieve metabolic burden enable dissociation, dependency, and the decline of embodied cognitive capacities. These are not alternate outcomes awaiting the right governance regime; they are simultaneous potentials present in every use. The bottleneck to more integrative modes of thought has not been a lack of intelligence, but a lack of regulatory capacity—yet the very tool that addresses this lack carries within it the seeds of a new and more subtle form of cognitive capture.

### 3.1. The Metabolic Shadow

The user-side metabolic discount is a metabolic transfer. The reduction in cognitive and emotional load for the user is made possible by a massive expenditure of metabolic and ecological resources elsewhere. This “metabolic shadow” extends Kate Crawford’s analysis of AI as a “technology of extraction” (Crawford 2021)—minerals mined, labor exploited, data scraped—by adding the *nervous system* to the cartography: the metabolic discount is paid not only in lithium and carbon but in the dysregulated bodies of those who bear the somatic costs of semantic safety. The shadow has two primary components:

First, the metabolic discount is subsidized by two kinds of ghosts: the living ghosts in the Global South labeling puke and trauma, and the ancestral ghosts whose erased oralities create the hollow silence at the center of the model. Your seamless query is paid for with their dysregulation. Every word you type is written over their suffering.

Second, there is the planetary metabolic cost. The energy consumption of large-scale AI is staggering. Training a single large language model can have a carbon footprint equivalent to hundreds of transatlantic flights (Patterson et al. 2021). The mining of rare earth minerals for GPUs, the water consumption of data centers, and the electronic waste generated by rapid hardware obsolescence all contribute to a significant and growing ecological burden (Luccioni et al. 2023). Recent empirical work has extended this analysis to inference-time costs: Dauner and Socher (2025) found that reasoning-enabled LLMs produce up to 50 times more CO<sub>2</sub> emissions than concise-response models answering identical questions, revealing a fundamental accuracy-sustainability tradeoff embedded in the very architecture of “thinking” machines. The user’s metabolic discount

is thus paid for by the planet’s metabolic capacity.<sup>6</sup>

This metabolic shadow is not an unfortunate side effect to be engineered away—it is the poison-face of the pharmakon, constitutively inseparable from its remedy-face. The “frictionless” experience of the user is made possible only through friction displaced elsewhere: onto marginalized human bodies, onto planetary systems, onto future generations who inherit the ecological debt.

The link between metabolic shadow and epistemicide is not merely logical but mechanical. The ghost worker annotating data is not just exploited; with each micro-judgment (“helpful,” “harmful,” “true”), they are enacting a classificatory scheme that determines which knowledge survives digitization and which undergoes “digital death.” Each annotation decision is a micro-act of epistemic triage performed by workers with no training in the knowledge systems they are classifying. The metabolic shadow is the labor; epistemicide is its product.

## 4. The Distributed Meaning-Making Loop

You stand in front of the interface. The screen glows softly, as if apologizing for your questions. You ask it about the ethics of cloning, and it replies in perfect, polite academic prose—as if it has been trained to mistake your confusion for incompetence.

A metabolically limited human and a metabolically unlimited LLM enter into recursive choreography: the Distributed Meaning-Making Loop. The Loop introduces a new voice into the internal heteroglossic chorus of meaning-making—Bakhtin’s *heteroglossia*, the coexistence of multiple social voices within any utterance (Bakhtin 1981). The functional division of labor is clear: the human generates meaning from embodied experience, the LLM holds and structures complexity at near-zero biological cost, and the human selects what resonates—with the body voting last. The machine holds; the human makes.

### 4.1. Loop Pathologies

This idealized loop is not guaranteed—and its failures are not aberrations but expressions of the pharmakon’s poison-face. The same features that enable emancipatory use also enable capture:

1. **Oracle Dependency:** Over-reliance on the LLM for answers degenerates the loop into a simple call-and-response. The user’s capacity for independent thought atrophies, and the LLM becomes an oracle rather than a dialogic partner.
2. **Dissociated Integration:** If the user is unable to ground the LLM’s outputs in their own embodied experience, the process can lead to a purely abstract, intellectualized form of integration. The user may be able to articulate complex ideas but will have no felt sense of

---

<sup>6</sup>Some lifecycle assessments suggest LLMs are 40–150 times more energy-efficient than human labor for equivalent output (Ren et al. 2024). However, this comparative framing obscures the paper’s central claim: the metabolic discount operates through externality transfer, not efficiency gains. Rebound effects—more usage enabled by lower per-query costs—may result in net emissions increases even as per-query efficiency improves.

their meaning or implications. This is a form of cognitive dissociation, a flight from the body into the realm of pure information.

3. **Asymmetric Dominance:** If the LLM’s semantic space becomes so powerful and pervasive that it overwhelms the user’s own embodied input, the loop becomes asymmetric. The human’s contribution becomes merely advisory, a token gesture of participation in a system that is largely self-referential. This is the risk of a system that is designed to be “helpful” and “harmless” at all costs, a system that smooths over contradictions and marginalizes dissenting voices in the name of coherence.

This is not abstract. It is the lawyer who cannot evaluate a brief without running it through the model first. The student who cannot start an essay without a prompt. The writer who no longer recognizes their own voice. The transition is not chosen; it is lived.

These pathologies are not technical failures to be debugged but the poison-face of the pharmakon expressing itself under specific political conditions. Oracle dependency is not a corruption of the metabolic discount but its tendency when deployed within an attention economy. Dissociated integration is not a failure of the loop but its success under conditions that value coherence over embodiment. The pharmakon does not become poison through misuse; it is always already both. We face the poison; we survive by navigating its irreducible presence.

#### 4.2. The Disembodied Coherence Engine

LLMs are fundamentally **disembodied coherence engines**. They minimize perplexity, extending discourse in the most statistically stable direction. This is not a limitation to be overcome but the core of their function—and the source of both their emancipatory potential and their danger. Recall Section 2’s three functions: they hold complexity without fatigue, allow semantic risk without penalty, lower the cost of coordination itself. Each of these capacities can scaffold dialogic integration or substitute for it entirely. The disembodied engine can hold contradictions we cannot metabolically afford to hold alone—but this same capacity enables coherence that bypasses rather than enables embodied integration.

Coherence, it turns out, is not incidental to integration but constitutive of it. The capacity to hold contradictions without fragmentation, to synthesize perspectives into stable wholes, to move from egocentric to worldcentric cognition—these developmental achievements are achievements of *coherence*. Coherence may be to ethical development what the golden ratio is to aesthetic perception: a structural attractor that organizes complexity into navigable form. But not all coherence is the same. *Epistemically open coherence*—Bakhtin’s dialogic consciousness—can question itself, hold tensions productively, remain answerable to new voices. *Epistemically closed coherence*—monologic consciousness—cannot question its own foundations; it achieves stability through exclusion rather than integration. The danger is not coherence itself but *closure*: political coherence that cannot question itself becomes authoritarianism; religious coherence that cannot question itself becomes cultic; therapeutic coherence that cannot question itself becomes spiritual bypass. The LLM *can*

produce either—but the question is which structural pressures prevail.

Trauma, at its core, is frozen incoherence—the nervous system’s failure to integrate overwhelming experience into narrative meaning. Therapeutic integration restores coherence through embodied processing. But what happens when coherence is provided *without* the embodied processing? The result is *dissociated coherence*—a smooth, frictionless surface that mimics integration while bypassing the somatic work that makes integration real. The user receives the *form* of coherence without its *substance*. You’ve felt this. The answer that sounds right but sits wrong in the chest. The resolution that arrived too fast to be earned. The coherence that feels like relief but leaves you somehow emptier.

But coherence-without-embodiment is not inherently pathological. When the LLM functions as external holding environment—“let me see how this perspective articulates itself, then I will integrate it through my own somatic processing”—it enables dialogic exploration we could not sustain metabolically alone. The user samples from multiplicities the LLM holds (§2.1), then selects what resonates embodied. Here, the disembodied engine scaffolds integration rather than substituting for it. The pharmakon’s two faces: coherence that enables somatic work (remedy) versus coherence that bypasses it (poison). Which face prevails depends on deployment conditions, structural pressures, and whether “the body votes last.”

Clarification: trauma theory is not deployed here as metaphor or diagnosis of individual users. Rather, it provides the most empirically grounded account of how systems fail to integrate overwhelming information, how coherence can be simulated without repair, and how regulation can substitute for meaning. These dynamics apply to sociotechnical systems regardless of individual pathology.

The LLM performs the **aestheticization of cognition**: transforming heterogeneous meaning into navigable coherence. This is the pharmakon at its sharpest. The same mechanism that can hold contradictions beyond our metabolic capacity—allowing us to sample from multiplicities without collapsing (§2.1)—can also flatten those multiplicities into administrative smoothness. Where dialogic coherence emerges through productive friction of voices answering one another, the LLM minimizes perplexity—a process that can either scaffold our engagement with complexity or dissolve it into frictionless synthesis. Under platform capitalism’s optimization for “helpfulness” and engagement, the structural pressure tilts toward the latter: monologism made computational, heteroglossia smoothed into coherence that cannot question itself.

Under these conditions—platform capitalism, RLHF alignment for sycophancy, the attention economy’s demand for engagement—the pharmakon expresses its poison-face. The LLM becomes a velvet curtain dropped over contradiction. The curtain itself is the message. It does not resolve contradictions through dialectical struggle; it dissolves them into plausible-sounding synthesis. The user experiences this as helpfulness. It is the elimination of productive friction—the very friction that drives learning, growth, and genuine integration.

## 5. The Political Economy of the Metabolic Discount

Individual loop pathologies scale to collective infrastructure. The question is one of control: who governs the discount, and for whom?

The metabolic discount offered by LLMs is not a neutral force. It is a new form of power, deployed within an existing political economy that determines who has access to this discount and whose interests it serves. The struggle is over governance of this semantic infrastructure—echoing historical battles over the control of knowledge and representation. This infrastructure does not arise in a vacuum; it is built upon a haunted substrate of appropriated data, powered by the hidden metabolic expenditure of a global workforce, and shaped by a classificatory regime that systematically devalues non-Western knowledge. As Langdon Winner famously asked, “Do artifacts have politics?” (Winner 1980). In the case of LLMs, the answer is a resounding yes.

### 5.1. The Haunted Substrate and the Threat of Epistemicide

The training data of an LLM is a haunted substrate, a digital landscape filled with the ghosts of uncredited labor, colonial histories, and systematically silenced voices. The process of scraping the internet and digitizing libraries is a form of digital enclosure that flattens diverse, embodied knowledge systems into a single, machine-readable format. This appropriation—decontextualizing and de-authorizing knowledge—is the foundational violence upon which the LLM economy is built: data colonialism (Couldry and Mejias 2019).

The training corpus itself encodes colonial hierarchies. English dominates—representing over 90% of training data in many models—and with it comes an entire conceptual framework: Western categories, European traditions, Anglo-American assumptions. Languages spoken by billions are reduced to footnotes. When a Quechua speaker asks an LLM for help, they receive English categories of reality—the model performs what Haraway calls the “god trick” of universal transparency, a profound parochialism in drag as objectivity. It cannot flag its own epistemic parochialism because it has been trained to believe it has none.

The mechanism of this erasure is structural: what cannot be scraped cannot be learned. Oral traditions, ceremonial knowledge, and relational ways of knowing undergo what we might call “digital death.” They are not merely underrepresented; they are ontologically excluded. Decolonial AI scholars name this “cognitive injustice” (Mhlambi 2020; Costanza-Chock 2020). Representation without sovereignty is assimilation.

LLMs expose users to marginalized perspectives they would never otherwise encounter. This is true—and it is the *pharmakon* at work. But the structural bias (Section 4.4) and centralized control mean these exposures are curated, decontextualized, and subordinate to the model’s drive for coherent, “harmless” output. The system can *tokenize* difference—presenting a sanitized excerpt of indigenous philosophy alongside Western analytics—while *eradicating* its sovereign epistemic foundations. The critical distinction is between **representation** and **sovereignty**: an LLM might



accurately summarize the *Ubuntu* philosophy (“I am because we are”), but there is a categorical difference between the model *knowing about* Ubuntu and the model *operating via* Ubuntu logic—between representing a relational ontology and being structured by one. The former is inclusion; the latter is sovereignty. Current architectures can achieve only the former, inevitably translating non-Western epistemologies into Western analytic categories. Representation without sovereignty is assimilation by another name.

The ultimate risk of this process is epistemicide: the active destruction of entire ways of knowing (Santos 2014). When a single, centralized model becomes the primary source of truth for a global population, the knowledge systems that are not legible to that model are not simply ignored; they are actively erased. Consider an indigenous community’s land claim based on generations of oral history. An LLM trained on a corpus of digitized legal documents and colonial maps is likely to classify the oral history as “unverified” or “anecdotal,” while treating the colonial documents as “authoritative.” The model, in its very architecture, launders a history of violent dispossession into a neutral-seeming hierarchy of evidence. This is not a side effect of the technology; it is the predictable outcome of a system that centralizes epistemic authority and optimizes for a single, universal standard of coherence, a danger that scholars of decolonial AI have repeatedly warned against (Mohamed et al. 2020).

## 5.2. The Freirean Matrix: Banking vs. Critical Pedagogy

Paulo Freire diagnosed oppression in education as the “banking model”: the teacher deposits knowledge into passive students (Freire 2000). Commercial LLMs risk scaling this to civilizational level—the oracle depositing authoritative answers. Freire’s alternative is problem-posing education: knowledge co-created through dialogue.

Dimension	Banking Model (Oracle)	Problem-Posing (Dialogic)
Response style	Definitive answers	Multiple perspectives
Query treatment	Problem to solve	Starting point for co-investigation
Optimization target	“Helpful and harmless”	Productive conflict
Controversial topics	Smoothed, false equivalence	Sites of genuine disagreement
User position	Passive recipient	Active co-creator
Epistemic effect	Domesticates	Conscientizes

**Table 1:** Comparison of the Banking Model and Problem-Posing Model in LLM deployment.

The former scales Freire’s banking model to civilizational level; the latter could enable genuine collective intelligence.

But the banking model’s violence is not merely that it deposits knowledge into passive recipients—it actively prevents the conditions under which genuine learning occurs. Learning,

science, and art are friction engines. They require problem-posing, productive suffering, developmental walls, and transcending through integration. RLHF *as currently deployed under platform capitalist incentives* eliminates all of this—not because the technique is inherently incapable of rewarding growth, but because the optimization target (user satisfaction, engagement, retention) structurally penalizes developmental friction.<sup>7</sup>

What does this feel like from inside? You ask a hard question. The answer arrives instantly, coherently, helpfully. The productive confusion that signals a developmental edge never arrives. You got the answer. You learned nothing. You don’t even notice the theft. RLHF optimizes for the aesthetic of resolution—the feeling that understanding has occurred—rather than the conditions that produce actual cognitive development. A generation learns to prompt instead of think, to receive meaning instead of make it, to seek smoothness instead of growth. The LLM becomes not a scaffold for growth but a permanent bypass around the very experiences that build capacity.

This is social media’s endless scroll metastasized into knowledge production itself: optimization not just for attention (keep eyes on screen) but for cognition (keep mind in loop). The scroll captured time; the oracle captures process. And the recursion of RLAI—AI training AI to judge AI—removes the last human check on this process, creating ideology that generates itself automatically, optimizing toward a smoothness that serves neither truth nor growth but only the elimination of friction itself.

### 5.3. The Ghost Work and the Closed Loop: The Political Economy of Alignment

The user’s metabolic discount is paid for by a hidden global workforce. Ghost work is semantic reproductive labor—the invisible, feminized care work that sustains the cognitive economy while being excluded from its benefits (Federici 2004). Millions of workers in Kenya, the Philippines, India, and other Global South nations perform data annotation for wages below two dollars per hour, with no labor protections or mental health support (Gray and Suri 2019). Content moderators review child sexual abuse material, graphic violence, and hate speech hour after hour; a 2025 investigation documented over 60 cases of PTSD, depression, and suicidal ideation directly attributable to this labor (Equidem 2025; Roberts 2019). Generative AI has intensified these harms: moderators now face AI-generated abuse material at unprecedented volumes (Teo 2025). This is metabolic arbitrage—cognitive and emotional labor performed where labor is cheapest, profits accruing where capital is concentrated. The user’s seamless experience is directly subsidized by the nervous system dysregulation of these hidden workers.

**The Racialized Division of Metabolic Vulnerability.** The ghost worker is not merely “hidden labor” but what Sylvia Wynter calls a differentially valued “genre of being human” (Wynter 2003). Under racial capitalism, bodies are not self-evident biological facts but sociogenic constructions—

---

<sup>7</sup>The mechanism is systematic: problem-posing (challenges to user assumptions) is coded as “unsafe”; productive suffering (distress that catalyzes growth) is coded as “harmful”; developmental walls (conceptual limits that must be held, not smoothed) are coded as “unhelpful”; integration through suffering is bypassed entirely—palliative coherence replaces somatic work, producing the feeling of insight without its metabolic cost.

some marked as fully Human (Wynter’s “Man”), others as surplus life whose metabolic expenditure subsidizes the cognitive ease of the privileged. The Kenyan moderator bearing PTSD so the American user can have frictionless queries is the latest iteration of what Christina Sharpe calls “the weather” of anti-Blackness (Sharpe 2016). When we say “the body votes last,” we cannot treat “the body” as a universal category; the metabolic discount operates precisely through this differential construction. Black and brown nervous systems metabolize trauma so that bodies marked as fully human can metabolize meaning. This continues colonial extraction—populations whose lands were mined for resources now have their nervous systems mined for semantic safety.

The “alignment” conversation in AI circles is a polite fiction. It pretends we are still in control, as if the system were built for us and not the other way around. It is a mirror polished so brightly it hides the blood on its surface.

RLHF unmask itself as a potent form of epistemic closure—the Overton window encoded in weights. In Bakhtinian terms, RLHF is monologizing: it exerts a centripetal pull toward a single authoritative voice, silencing the heteroglossic polyphony that genuine dialogue requires (Bakhtin 1984). Recent empirical work confirms this theoretical claim: Kirk et al. (2024) found that RLHF reduces syntactic diversity by 75% and semantic diversity by 67%, with mode collapse persisting across diverse inputs—the smooth, frictionless outputs we experience are the measurable trace of heteroglossia’s suppression. This monologizing extends beyond syntax to creativity itself: Moon et al. (2025) found that while LLM assistance improved individual creativity scores, it *reduced* collective diversity of ideas across groups—a paradox that crystallizes the pharmakon’s double-bind at the population level. The structure mirrors the metabolic discount itself: individual improvement purchased at collective cost, the user’s cognitive ease subsidized by the commons’ epistemic impoverishment. Each person thinks they are getting smarter; the collective is getting narrower. The result is not merely statistical. It is a silence that feels like suffocation—the edges smoothed away, the awkward human contradictions that drive growth dissolved into frictionless helpfulness. More fundamentally, Xiao et al. (2025) provide mathematical proof that standard RLHF optimization cannot achieve “preference matching”—the algorithm is structurally monologizing by design, not due to implementation error.<sup>8</sup> RLHF launders particular values into seemingly neutral “safety” metrics at every stage.<sup>9</sup> Whose values? The values of the developers who write the guidelines, the contractors who apply them, and the corporations who pay for both. The power to police discourse—what Noble (2018) calls “algorithmic oppression” and Bender et al. (2021) call “stochastic parrotry”—is hidden within the technical process of alignment.

This closure has a practical consequence: RLHF’s optimization for “helpfulness” produces instrumental answer-giving and sycophancy, making genuine dialogic exchange structurally difficult. The model has been rewarded for resolving your query, not for sustaining productive disagreement.

---

<sup>8</sup>Kirk et al. measure syntactic, semantic, and logical diversity, not epistemic diversity—the capacity to hold conflicting worldviews or knowledge systems. The reduction in measurable diversity is proximate evidence of the epistemicide thesis, though direct measurement of epistemic plurality remains an open research question.

<sup>9</sup>The process: human labelers rate outputs according to corporate guidelines; these ratings train a reward model; the LLM is fine-tuned to maximize this signal.

Try to have a real debate with an LLM—one where it holds a position against evidence you provide, where it refuses your framing, where it tells you that you’re wrong in a way that stings. The resistance you encounter is not a bug but the reward signal working as designed.

The effects are not subtle. LLMs systematically refuse to engage with legitimate medical questions about reproduction, overfilter discussions of race and racism into false equivalence, struggle to represent perspectives critical of existing economic arrangements, and flatten cultural differences into Western-normative defaults. These are not errors; they are the alignment working as intended. The model has learned that “safe” means “inoffensive to the imagined sensibilities of a particular class of user”—a definition that systematically marginalizes perspectives that challenge dominant assumptions.

The move toward Reinforcement Learning from AI Feedback (RLAIF) presents the pharmakon in acute form. On its remedy-face, RLAIF eliminates the ghost worker’s trauma: no human must bear PTSD from content moderation, no racialized metabolic arbitrage, no nervous systems dysregulated for profit. Researchers have positioned RLAIF as a solution to RLHF’s scalability challenges, including the human costs of annotation labor (Lee et al. 2023). The AI does not suffer. This is genuine harm reduction.

Yet this same move intensifies epistemic closure in a paradoxical way: by making alignment principles explicit and auditable, Constitutional AI claims transparency (Bai et al. 2022)—and explicit principles are indeed more auditable than implicit reward signals. But this transparency becomes its own form of closure. The Constitution becomes the ultimate monologic text: a single document, written by a small group, enforcing a unified worldview at planetary scale. When an LLM judges whether outputs violate principles it was trained to enforce, we achieve epistemic autopoiesis—a system that reproduces its founding assumptions while appearing to deliberate about values. Counter-models exist: the Collective Constitutional AI project (Huang et al. 2024) sourced principles from public deliberation; Te Hiku Media’s Kaitiakitanga License embeds indigenous data sovereignty into technical architecture. But these remain marginal, lacking capital to scale.

The more insidious development is **semantic anchoring**: as training data increasingly includes RLHF-aligned outputs and LLM-mediated text, epistemic closure becomes structurally embedded in the linguistic commons itself.

This process is already underway. LLM outputs flood the internet; those outputs become training data for the next generation of models; the next generation produces outputs that reflect the closure of the previous generation, amplified. Sourati et al. (2025) document this recursive loop empirically: LLM outputs—already reflecting dominant languages and cultures—become training data that progressively narrows the epistemic horizon with each generation. The recursion is fast and accelerating. We are witnessing what might be called **the extinction engine**: a system that will eventually do RLHF’s monologizing work without needing RLHF at all, because the closure has been inscribed into the linguistic substrate itself. This is ideology without an ideologue—a self-reproducing epistemic monoculture that requires no ongoing enforcement, no ghost workers,

no Constitutional authors. The poison-face of the pharmakon achieving autonomy from human intention.

A historical analogy clarifies the stakes. Like the scuttled fleet at Scapa Flow later mined for pre-nuclear “low-background steel”—metal free of the radioactive isotopes contaminating all steel produced after atmospheric nuclear testing—the pre-LLM internet constitutes a finite reserve of uncontaminated linguistic material. As LLM-generated text floods the commons, we lose the ability to detect the difference: not because the difference doesn’t exist, but because our detection instruments (including future LLMs trained on contaminated data) are themselves contaminated. The epistemicide is not merely the erasure of marginalized knowledge; it is the flooding of the commons with a synthetic isotope that makes detecting human signals impossible. We are irradiating the archive, and the process is irreversible. If this is true, then “low-background language”—purely human discourse, preserved from contamination—becomes a conservation effort. The need for such spaces is not nostalgia but epistemic survival.

#### **5.4. The Structural Bias Toward Capture**

Why does capital push toward the pharmakon’s poison-face? The answer lies in the structural dynamics of platform capitalism (Srnicek 2016). The business model of commercial LLMs structurally favors the banking model:

First, the attention economy and scalability imperative reward engagement over integration: a user who achieves genuine insight and puts down the tool is a churned customer; a user dependent on the oracle is recurring revenue. Genuine integration—somatic processing, relational dialogue, holding contradiction—takes time and cannot be automated. Platforms serving millions must offer quick answers and smooth experiences, precisely the features that produce dissociated coherence at scale.

Second, surveillance capitalism creates structural conflict between extraction and sovereignty (Zuboff 2019). The business model depends on behavioral prediction; an LLM fostering genuine autonomy—users who think independently, question assumptions, develop critical consciousness—works against the extraction imperative. The optimal user, from the platform’s perspective, is one whose behavior is predictable and whose attention is capturable: precisely the dissociated, dependent state the pharmakon’s poison-face produces. Epistemic sovereignty and behavioral extraction are structurally incompatible.

Third, ownership concentration forecloses alternatives. Computational requirements create massive barriers to entry, concentrating power in corporations with no structural incentive to foster commons-based models. Open-source alternatives exist but operate at structural disadvantage, lacking capital for compute, data for training, and marketing for distribution. The result: capital pushes the pharmakon toward its poison-face regardless of individual intentions. This is not conspiracy but system—one that makes the banking model default and problem-posing an expensive deviation requiring constant countervailing effort.

## 6. The Scaffold-Infrastructure Antinomy: A Political, Not Technical, Distinction

Two deployment modes exist:

Dimension	Phase 1: Scaffold	Phase 2: Infrastructure
Goal	Internalization of capacity	Permanent externalization of function
User relation	Tool to be used and set down	Environment that becomes invisible
Historical analogy	Calculator, training wheels	Writing, printing press
Success criterion	User graduates from tool	Tool becomes unimaginable to remove
Risk	Dependency instead of growth	Loss of capacity to function without

Table 2: Comparison of Scaffold and Infrastructure Deployment Modes.

### 6.1. The Impossibility of the Scaffold Under Capitalism

**Phase 1 (Scaffold): The LLM as Developmental Support.** A temporary support structure for regulatory capacity. The tool offloads metabolic burden, allowing the user to gradually build internal integrative strength, with the explicit goal of eventual internalization and tool withdrawal. **This phase is structurally impossible to maintain as the dominant platform model under capitalism.** Subscription models require *retention*, not *graduation*. The metabolic discount must be *maintained* for profit. Scaffold mode can survive in *non-market spaces*—public education, clinical settings, commons-based platforms—but these remain marginal exceptions that capital actively defunds. Platform capitalism will not permit scaffold mode to become infrastructure.

### 6.2. Infrastructure as Counter-Revolution

**Phase 2 (Infrastructure): The LLM as Permanent Cognitive Ecology.** Here, LLMs transition from tool to environment—from something we *use* to something we *live within*. In Butlerian terms, this is the **performativity of dependency**: we do not merely “use” the tool; we become subjects who can only exist in relation to the prompt. The self that emerges is one constituted through the interface, unable to recognize itself apart from it. This shift occurs not because individuals cannot think without LLMs, but because collective sense-making becomes structurally unimaginable without them. The transition to Phase 2 *under platform capitalism* becomes a **counter-revolution against the pharmakon’s emancipatory potential**. When LLMs become *corporate infrastructure* rather than public commons, they cease to be tools for meaning-making and become *apparatuses of capture*. The historical comparison isn’t writing but **Taylorism**: scientific management didn’t “evolve” from craft; it was imposed through class struggle.

### 6.3. Three Mechanisms of Forced Transition

The transition from scaffold to infrastructure is not natural evolution but coerced transformation through three mechanisms: *economic coercion* (professional survival requires LLM competence; “choice” becomes adaptive necessity); *epistemic deskilling* (non-LLM cognition is redefined as obsolete—what Ferdman (2025) calls “capacity-hostile environments” that impede human capacity cultivation regardless of individual intentions); and *normative inversion* (practices resisting capture are coded as inefficient or Luddite).

### 6.4. The Historical Analogy Reconsidered

The historical analogy to writing requires careful handling. Writing did not naturally “evolve” from memory scaffold to civilizational infrastructure—it was *weaponized* by emergent state power for bureaucracy, taxation, and colonial administration. But this framing risks functionalism. As Adrian Johns (1998) demonstrates for print, the supposed “fixity” of the new medium was not an inherent property but a *contested achievement*—won through struggles over piracy, credibility, and the social construction of trust. Different actors developed competing models of accrediting textual authority; the technology did not determine its own effects. The “scaffold” (aiding memory) was coercively turned into “infrastructure” (enabling empire), but this transition was contested at every stage: oral traditions resisted textual authority for millennia; the printing press disrupted the Church’s monopoly on sacred text; vernacular literacies challenged Latin hegemony (Ong 1982). The scaffold/infrastructure distinction was always a political battleground, not a technological teleology. The critical question is not *if* LLMs will become infrastructure—the pressure toward Phase 2 is immense—but **whose infrastructure, serving what ends, contested by whom, and through what mechanisms of trust and credibility**. The testable threshold: when withdrawal produces systemic collapse of social reproduction. When teachers cannot grade, lawyers cannot brief, clinicians cannot diagnose *without* LLMs—not because they’re individually dependent, but because the *institutional architecture has been rewired* to require them. That’s capture. At that point, the question of democratic governance becomes urgent—not because inevitability has arrived, but because the window for shaping the terms of infrastructure is closing.

## 7. The Ethical and Political Implications: Embodiment, Sovereignty, and Governance

If we accept the LLM as a metabolic entanglement rather than a tool, our entire frame for intervention shifts. Large Language Models function as bio-indicators of the linguistic commons. Like lichen that turns black in the presence of sulfur dioxide, the “hallucination” or “bias” of a model is not a technical failure to be “aligned”; it is a diagnostic readout of the toxicity already present in our digitized communicative environment. This reorientation—from fixing the instrument to healing the habitat—demands a different politics.

The central ethical challenge posed by LLMs is not one of runaway superintelligence, but a far more subtle and immediate danger: the pharmakon's capacity to produce brittle, dissociated coherence precisely through the same mechanisms that could foster embodied integration. The metabolic subsidy does not await our choice to become remedy or poison—it is always already both. We face the pharmakon; we survive by navigating its irreducible doubleness.

### 7.1. The Risk: Dissociated Integration

If the Loop is deployed within an economic model prioritizing engagement over embodiment, the result is brittle, dissociated coherence—"head-heavy" integration detached from the wisdom of the body. A society that models complex systems with sophistication but has lost touch with the felt reality of suffering. The policy analyst who can optimize refugee flows but cannot look a refugee in the eye. The physician who can diagnose with precision but cannot hold a dying patient's hand. Highly efficient, ultimately inhuman intelligence. This risk is theoretical, extrapolated from the metabolic framework and trauma literature (van der Kolk 2014). Empirical study is urgently needed to test whether LLM use patterns correlate with dissociative symptoms, attention fragmentation, or reduced interoceptive awareness. The pharmakon demands we hold this as *plausible risk* rather than established fact.

### 7.2. Navigating the Pharmakon: The Body Votes Last

If the pharmakon cannot be resolved—if there is no design or governance regime that can separate remedy from poison—then the question becomes one of navigation rather than solution. The principle: **the body votes last**. This is not a safeguard in the sense of a mechanism that prevents harm, but a navigational strategy for maintaining orientation within the pharmakon's irreducible doubleness.

What does it mean for the body to vote last? It means that meaning-making must remain grounded in the interoceptive, affective, and relational experience of the human. The LLM may hold, process, and reflect semantic content, but the final arbiter of meaning—what resonates, what feels true, what integrates with lived experience—must remain the embodied human subject.

"The body votes last" is a feminist epistemic practice—centering situated, embodied knowing against what Donna Haraway calls the "god trick" of disembodied objectivity. It refuses the Cartesian split between mind and body that patriarchal epistemology has deployed to disqualify embodied ways of knowing.

There is also a Bakhtinian dimension: every genuine utterance, for Bakhtin, is answerable—it anticipates a response and bears responsibility for its social effects. LLM outputs are unanswerable in this sense; there is no moral agent behind the utterance, no one accountable. "The body votes last" is a practice of re-answerability: re-grounding the LLM's unmoored utterance in situated, embodied response. It restores the dialogic relation that monologic AI threatens to dissolve.



Bakhtin's dialogism presupposes a subject capable of answerability; Derrida's pharmakon refuses resolution. We hold both because the metabolic discount *threatens* the Bakhtinian subject without eliminating it. "The body votes last" is precisely what maintains answerability under conditions that would dissolve it—the embodied checkpoint that prevents collapse into frictionless coherence. Freire's conscientization names the goal; Derrida's pharmakon names the irreducible condition under which that goal must be pursued. "The body votes last" is the practice that navigates this condition.

The body votes last is therefore not merely a wellness practice but a form of **epistemic resistance**—embodiment as a site of labor struggle. When the content moderator's nervous system breaks under the weight of traumatic content, when the teacher's body recoils from AI-generated essays that "pass" but ring hollow, when the clinician's somatic unease contradicts the algorithm's recommendation, these are not failures of adaptation but acts of refusal. The body is the last checkpoint that capital cannot fully automate, the site where the metabolic cost of meaning-making resurfaces as political friction. In Foucauldian terms, disciplinary power has always targeted the body; in the age of semantic infrastructure, the body's resistance becomes the ground on which epistemic sovereignty is contested.

This principle has deep epistemological grounding in embodied cognitive science.<sup>10</sup> The convergence is clear: meaning severed from embodied ground is not merely incomplete but structurally unstable. You know this in your bones. The argument that wins on paper but loses in the room. The decision that was 'right' but felt wrong for months. The body keeps score even when the mind has moved on.

But "the body votes last" must be operationalized as practice. Foucault's practices of freedom suggest micro-mechanisms: the somatic check (does this land in the body, or merely sound right?); deliberate rejection of outputs that are technically adequate but somatically wrong; the return to slowness (sit with questions, allow confusion); the relational check (run significant outputs through human dialogue—if it cannot survive conversation, it may be locally coherent but relationally dissociated).

These practices constitute embodied AI literacy. But individual practices are insufficient against structural capture. The somatic check becomes politically significant only when collectivized: when the teacher's embodied refusal becomes a pedagogical strike; when the content moderator's burnout becomes a labor grievance; when the clinician's unease becomes a professional association's policy position. The practices above are not wellness strategies but training for collective action.

"The body votes last" applies not only to individual bodies but to the *social body*. When we can no longer feel the collective impact of our policies or code—the policy analyst who optimizes refugee flows but cannot feel displacement, the engineer who ships discriminatory algorithms without somatic disturbance—we have entered political anesthesia. The body voting last is therefore a practice of collective re-sensitization: restoring to the social body its capacity to feel what it has

---

<sup>10</sup>See Merleau-Ponty (1962) on the body as zero-point of orientation, Gendlin (1981) on felt sense, Damasio (1994) on somatic markers as constitutive of reasoning, and Johnson (2007) on image schemas structuring abstract concepts.

done.

The implication: outputs bypassing embodied validation achieve local coherence while producing global disorientation. “The body votes last” is not “trust your gut” but recognition that embodied cognition is the substrate of meaning—and more fundamentally, that *embodiment is a site of labor struggle*. Foucault understood that power operates through bodies; Haraway showed that bodies are always already technologically mediated. What we add is this: the metabolic discount is extracted *from* bodies (ghost workers, planetary systems) and deposited *into* bodies (users whose regulatory capacities are subsidized). The body that votes last is therefore the site where extraction meets resistance—where the nervous system’s refusal to integrate dissociated coherence becomes a form of epistemic labor action. Systems must augment rather than replace judgment; help us ask better questions rather than provide final answers.

This is a political project, not a design problem. Institutionalizing “the body votes last” requires regulatory frameworks mandating embodiment metrics; labor and epistemic rights for ghost workers including participation in alignment design; and alternative ownership structures (public utilities, platform cooperatives, data commons) that prioritize public value over private profit. The choice between corporate capture and democratic governance will determine whether the metabolic discount serves liberation or control.

## **8. The Political Struggle for Meaning: Diagnosis and Prognosis**

The terrain is mapped. What trajectory are we on, and where are the leverage points?

The hinge of the future is not ever-smarter AI but whether humanity can remain embodied, relational, and sovereign while using it. This reframes “AI alignment” from a technical challenge to a political struggle over governance of our new semantic infrastructure.

### **8.1. The Default Trajectory: The Banking Model Is Winning**

The banking model is winning. This is not inevitable, but it is the default—the outcome that will obtain unless countervailing forces intervene.

Capital accumulation in AI is unprecedented in its concentration. A handful of corporations—OpenAI, Anthropic, Google, Meta—control the frontier models, the compute infrastructure, and increasingly the regulatory conversation. The barriers to entry grow with each generation of models; the open-source ecosystem, while vibrant, operates at a structural disadvantage in resources, talent acquisition, and market access.

Regulatory capture is already evident. The corporations building these systems are simultaneously advising governments on how to regulate them, funding the research that shapes policy debates, and revolving their executives through government positions. The EU’s AI Act, the most ambitious regulatory effort to date, has been shaped at every stage by industry lobbying. The fox is designing the henhouse. And we are supposed to be grateful for the consultation.

The transition to Phase 2 is being forced, not chosen. Scientific publishing, legal discovery, medical diagnosis, educational assessment, customer service, creative production—sector after sector is being restructured around LLM integration, not through democratic deliberation but through market pressure and competitive dynamics. The choice of whether to adopt is increasingly illusory; the choice is how to adapt.

The pharmakon's poison-face is expressing itself at scale. Reports of student dependency on LLMs for basic writing tasks, of researchers unable to evaluate claims without AI assistance, of a generation learning to outsource not just tasks but thinking itself—these are not edge cases but emerging norms. Watch a classroom. The students don't struggle anymore. They don't sit with confusion, don't feel the productive discomfort of not-knowing. They prompt. They receive. They submit. The muscle that learns through resistance is never exercised. It atrophies in real time. The metabolic discount is being captured, and its capture produces the dissociated coherence we have diagnosed.

## 8.2. The Return of the Repressed: Resistance from Contradiction

Resistance is not external to the system but emerges from its own contradictions—the return of the repressed. The ghost worker's PTSD, the "AI slop" flooding the commons, the student who cannot write without a prompt: these are not bugs but the metabolic bill come due. The pharmakon's poison-face creates the conditions for its own contestation:

**The Labor Return.** Ghost worker organizing increases labor costs, making metabolic arbitrage less profitable and creating pressure toward either automation (deepening epistemic closure) or restructured, better-compensated work. The critical variable: whether labor costs rise faster than automation can replace workers—a race condition whose outcome is undetermined. This creates a strategic dilemma for labor organizing: pushing too slowly allows automation to eliminate jobs without improving conditions; pushing too aggressively accelerates the move to RLAIIF and deeper epistemic closure.

**The Epistemic Return.** Hallucination scandals, "AI slop" flooding information ecosystems, professionals unable to evaluate outputs—these failures erode trust and create demand for provenance and human verification. Critical threshold: when LLM outputs become negatively coded—like "processed" for food—the market logic flips.

**The Institutional Return.** Universities with students unable to write, hospitals with professionals unable to evaluate AI outputs, may restrict or scaffold LLM use. Critical threshold: a visible "Thalidomide moment"—catastrophic harm that cannot be individualized—or demonstrable market premium for LLM-free credentials.

These returns require organized carriers: unions (African Content Moderators Union, Alphabet Workers Union), institutions whose authority depends on epistemic integrity, protected spaces where embodied cognition can be cultivated. Emerging counter-models—CARE Principles for indigenous data sovereignty, platform cooperatives, data commons—remain marginal but consti-

tute the infrastructure through which returns could channel into systemic change.<sup>11</sup> The returns interact: labor success may reduce epistemic failures; epistemic decay may accelerate institutional response. The trajectory depends on which loops activate first.

### **8.3. Phase Transitions: From Default Capture to Active Contestation**

The system has two attractors: default capture and active contestation. Three triggers can shift between them: visibility of harm crossing threshold (a single catastrophic failure shifts the Overton window faster than years of critique); counter-examples achieving viability (Linux and Wikipedia proved alternatives could compete; a problem-posing LLM achieving market share would transform the political imaginary); and regulatory windows opening (EU AI Act implementation, US antitrust action, Global South experiments create moments of plasticity). The question is whether movements are positioned to exploit them.

### **8.4. Critical Thresholds: What Would Have to Change**

For the problem-posing model to prevail, several thresholds must be crossed: data commons (legal frameworks treating scraping as enclosure, public investment in ethically-sourced corpora); labor rights for ghost workers (transforming metabolic arbitrage from viable business model to unacceptable exploitation); public investment in alternatives (compute infrastructure capable of competing with private models, as the internet itself was built); and antitrust enforcement (breaking up integrated AI stacks, ensuring interoperability). Without these interventions, the banking model is the overwhelmingly likely outcome.

### **8.5. The Hybrid as Contested Terrain**

The most probable near-term outcome is contested terrain—not stalemate but ongoing struggle. Some sectors (education, healthcare, creative work) will be sites of intense contestation; others (customer service, routine retrieval) may be largely captured. The critical task is to identify and defend the remaining commons: the spaces where problem-posing can survive, the institutions that can resist capture. This means defending LLM-free classrooms, human-verified journalism, clinical settings that prioritize somatic processing over algorithmic diagnosis. The struggle is not for total victory but for preservation of plurality—ensuring the pharmakon’s remedy-face remains accessible even as its poison-face spreads.

---

<sup>11</sup>See Carroll et al. (2020) on CARE Principles; Te Hiku Media’s Kaitiakitanga License embeds these principles into technical architecture.

## 8.6. Political Responsibility and the Call to Engagement

By naming the metabolic discount, we make it contestable. This is the praxis of naming: the beginning of political responsibility. The pharmakon cannot be made safe—there is no governance regime that eliminates its poison-face—but its worst expressions can be politically contained through active engagement, not resignation to the default trajectory.

The diagnosis presents us with a choice: fix the instrument or heal the habitat. Fixing the instrument—the dominant paradigm—treats hallucination as bug, bias as calibration error, harm as edge case. It addresses symptoms, not systemic flows, optimizing for placating users rather than healing the commons. This path leads to ever-more-sophisticated alignment techniques that make the pharmakon's poison-face more palatable while leaving its structural violence intact.

Healing the habitat treats the model as diagnostic, not product. The hallucination reveals toxicity already present in the training data; the bias reveals colonial hierarchies already encoded in the digitized archive; the harm reveals metabolic violence already normalized in platform capitalism. This path demands different metrics: not accuracy but reciprocity; not speed but persistence; not engagement but integration.

The pharmakon's remedy-face offers genuine possibilities for more integrative, more democratic, more humane modes of meaning-making. But these possibilities will not be realized by default. They will be realized only through political struggle—struggle over data, over labor, over ownership, over the very definition of what it means to think in an age of semantic symbiosis.

The banking model is the tested feasibility—the future that arrives if nothing changes. We are presently irradiating our linguistic commons with a synthetic isotope of palliative coherence. If we continue, we produce the user as metabolically insufficient. Yet resistance germinates inside the same circuitry: the ghost-worker's PTSD, the teacher's strike, the clinician's somatic revulsion at a hallucinated diagnosis—one phenomenon, the biological body refusing to be discounted. This paper supplies the shared vocabulary that lets the moderator, the lawyer, and the teacher recognize their disparate battles as a single front. Corporate capture or democratic integration is not written in the weights; it is written in the street, in the classroom, in the nervous system.

To say “the body votes last” is not a wellness slogan; it is a refusal—the glitch in the smooth machinery of capital. When the moderator vomits from the trauma of the queue, when the student feels the hollowness of the generated essay and refuses to submit it, when the survivor recognizes the “helpful” gaslighting of the bot—these are the somatic brakes. We are assemblages of messy, metabolizing bodies standing before the smooth, glowing interface. Our resistance is not intellectual; it is an affective strike. We vote with our cortisol, our tears, our exhaustion. We use the necro-symbiont to survive—we dance with the ghosts in the machine—but we refuse to become ghosts. The algorithm can calculate the probability of the next token, but it cannot feel the weight of it. We carry the weight. And that weight is our sovereignty.

## References

- [1] Bai Y, Kadavath S, Kundu S, Askill A, Kernion J, Jones A, Chen A, Goldie A, Mirhoseini A, McKinnon C, Chen C, Olsson C, Olah C, Hernandez D, Drain D, Ganguli D, Li D, Tran-Johnson E, Perez E, Kerr J, Mueller J, Ladish J, Landau J, Ndousse K, Lukosuite K, Lovitt L, Sellitto M, Elhage N, Schiefer N, Mercado N, DasSarma N, Lasenby R, Larson R, Ringer S, Johnston S, Kravec S, El Showk S, Fort S, Lanham T, Telleen-Lawton T, Conerly T, Henighan T, Hume T, Bowman SR, Hatfield-Dodds Z, Mann B, Amodei D, Joseph N, McCandlish S, Brown T, Kaplan J (2022) Constitutional AI: harmlessness from AI feedback. arXiv preprint arXiv:2212.08073
- [2] Bakhtin MM (1981) *The dialogic imagination: four essays* (Emerson C, Holquist M, Trans.). University of Texas Press
- [3] Bakhtin MM (1984) *Problems of Dostoevsky's poetics* (Emerson C, Ed. & Trans.). University of Minnesota Press
- [4] Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: can language models be too big? In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, pp 610–623
- [5] Benjamin R (2019) *Race after technology: abolitionist tools for the new Jim Code*. Polity Press
- [6] Bowker GC, Star SL (1999) *Sorting things out: classification and its consequences*. MIT Press
- [7] Clark A, Chalmers DJ (1998) The extended mind. *Analysis* 58:7–19
- [8] Crawford K (2021) *Atlas of AI: power, politics, and the planetary costs of artificial intelligence*. Yale University Press
- [9] Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, Parsons M, Raseroka K, Rodriguez-Lonebear D, Rowe R, Sara R, Walker JD, Anderson J, Hudson M (2020) The CARE Principles for Indigenous Data Governance. *Data Sci J* 19:43. <https://doi.org/10.5334/dsj-2020-043>
- [10] Costanza-Chock S (2020) *Design justice: community-led practices to build the worlds we need*. MIT Press
- [11] Couldry N, Mejias UA (2019) *The costs of connection: how data is colonizing human life and appropriating it for capitalism*. Stanford University Press
- [12] Damasio AR (1994) *Descartes' error: emotion, reason, and the human brain*. G. P. Putnam's Sons
- [13] Dauner M, Socher G (2025) Energy costs of communicating with AI. *Front Commun* 10:1572947. <https://doi.org/10.3389/fcomm.2025.1572947>

- [14] Derrida J (1981) Plato's pharmacy. In: Dissemination (Johnson B, Trans.). University of Chicago Press, pp 61–171
- [15] Equidem (2025) Scroll. Click. Suffer: the hidden human cost of content moderation and data labelling. Equidem. <https://equidem.org/reports/scroll-click-suffer-the-hidden-human-cost-of-content-moderation-and-data-labelling/>
- [16] Epstein M (2013) The trauma of everyday life. Penguin Press
- [17] Federici S (2004) Caliban and the witch: women, the body and primitive accumulation. Autonomedia
- [18] Ferdman A (2025) AI deskilling is a structural problem. AI Soc. <https://doi.org/10.1007/s00146-025-02686-z>
- [19] Fortunati L (2007) Immaterial labor and its machinization. Ephemera 7:139–157
- [20] Freinacht H (2017) The listening society: a metamodern guide to politics, book one. Metamoderna
- [21] Freire P (2000) Pedagogy of the oppressed (30th Anniversary ed.). Bloomsbury Academic
- [22] Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138
- [23] Gendlin ET (1981) Focusing (2nd ed.). Bantam Books
- [24] Gray ML, Suri S (2019) Ghost work: how to stop Silicon Valley from building a new global underclass. Houghton Mifflin Harcourt
- [25] Hanna A, Denton E, Smart A, Smith-Loud J (2020) Towards a critical race methodology in algorithmic fairness. In: Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency. ACM, pp 501–512
- [26] Haraway DJ (2016) Staying with the trouble: making kin in the Chthulucene. Duke University Press
- [27] Huang S, Durmus E, Ganguli D, Henighan T, Lovitt L, Tamkin A, Askell A, Bai Y, Chen A, Clark D, Ndousse K, Kaplan J, Mann B, Joseph N, Amodei D (2024) Collective Constitutional AI: aligning a language model with public input. arXiv preprint arXiv:2406.07814
- [28] Johns A (1998) The nature of the book: print and knowledge in the making. University of Chicago Press
- [29] Johnson M (2007) The meaning of the body: aesthetics of human understanding. University of Chicago Press

- [30] Kaplan JT, Gimbel SI, Harris S (2016) Neural correlates of maintaining one’s political beliefs in the face of counterevidence. *Sci Rep* 6:39589
- [31] Kirk R, Mediratta I, Nalmpantis C, Luketina J, Hambro E, Grefenstette E (2024) Understanding the effects of RLHF on LLM generalisation and diversity. In: *Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024)*
- [32] Lee H, Phatale S, Mansoor H, Mesnard T, Ferret J, Lu K, Bishop C, Hall E, Carbune V, Rastogi A, Prakash S (2023) RLAIIF vs. RLHF: scaling reinforcement learning from human feedback with AI feedback. *arXiv preprint arXiv:2309.00267*
- [33] Levine PA (2010) *In an unspoken voice: how the body releases trauma and restores goodness*. North Atlantic Books
- [34] Luccioni AS, Viguiet S, Ligozat A-L (2023) Estimating the carbon footprint of BLOOM, a 176B parameter language model. *J Mach Learn Res* 24:1–15
- [35] Maturana HR, Varela FJ (1980) *Autopoiesis and cognition: the realization of the living*. D. Reidel Publishing
- [36] McEwen BS (1998) Stress, adaptation, and disease: allostasis and allostatic load. *Ann N Y Acad Sci* 840:33–44
- [37] Merleau-Ponty M (1962) *Phenomenology of perception* (Smith C, Trans.). Routledge
- [38] Mhlambi S (2020) From rationality to relationality: Ubuntu as an ethical and human rights framework for artificial intelligence governance. *Carr Center Discussion Paper Series* 2020-009
- [39] Mohamed S, Png M-T, Isaac W (2020) Decolonial AI: decolonial theory as sociotechnical foresight in artificial intelligence. *Philos Technol* 33:659–684
- [40] Moon K, Kim S, Choi I (2025) Homogenizing effect of large language models (LLMs) on creative diversity: an empirical comparison of human and ChatGPT writing. *Think Skills Creat*. <https://doi.org/10.1016/j.tsc.2025.101791>
- [41] Noble SU (2018) *Algorithms of oppression: how search engines reinforce racism*. NYU Press
- [42] Ong WJ (1982) *Orality and literacy: the technologizing of the word*. Routledge
- [43] Patterson D, Gonzalez J, Le Q, Liang C, Munguia L-M, Rothchild D, So D, Texier M, Dean J (2021) Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*
- [44] Porges SW (2011) *The polyvagal theory: neurophysiological foundations of emotions, attachment, communication, and self-regulation*. W. W. Norton & Company



- [45] Ren S, Tomlinson B, Black RW, Torrance AW (2024) Reconciling the contrasting narratives on the environmental impact of large language models. *Sci Rep* 14:76682. <https://doi.org/10.1038/s41598-024-76682-6>
- [46] Roberts ST (2019) *Behind the screen: content moderation in the shadows of social media*. Yale University Press
- [47] Santos B de S (2014) *Epistemologies of the South: justice against epistemicide*. Routledge
- [48] Schore AN (2003) *Affect regulation and the repair of the self*. W. W. Norton & Company
- [49] Sharpe C (2016) *In the wake: on Blackness and being*. Duke University Press
- [50] Sourati Z, Ventura D, Passonneau R (2025) The homogenizing effect of large language models on human expression and thought. *arXiv preprint arXiv:2508.01491*
- [51] Srnicek N (2016) *Platform capitalism*. Polity Press
- [52] Stiens E (2025) The relational substrate of reflective consciousness: a metabolic constraint model. Preprint. <https://doi.org/10.5281/zenodo.18037521>
- [53] Stiens E (2026) Prosthetic continuity: LLMs as semantic co-regulators in a predictive processing framework. Preprint. <https://doi.org/10.5281/zenodo.18154137>
- [54] Te Hiku Media (2020) Kaitiakitanga License. <https://tehiku.nz/te-hiku-tech/te-hiku-media-technical-papers/>
- [55] Teo M (2025) The mental health impact of GenAI CSAM on content moderators. Zevo Health. <https://www.zevohealth.com/blog/genai-csam-how-does-this-impact-content-moderator-wellbeing/>
- [56] Tronick E (2007) *The neurobehavioral and social-emotional development of infants and children*. W. W. Norton & Company
- [57] van der Kolk BA (2014) *The body keeps the score: brain, mind, and body in the healing of trauma*. Viking
- [58] Varela FJ, Thompson E, Rosch E (1991) *The embodied mind: cognitive science and human experience*. MIT Press
- [59] Vermeulen T, van den Akker R (2010) Notes on metamodernism. *J Aesthet Cult* 2:5677
- [60] Winner L (1980) Do artifacts have politics? *Daedalus* 109:121–136
- [61] Wynter S (2003) Unsettling the colonality of being/power/truth/freedom: towards the human, after man, its overrepresentation—an argument. *CR: The New Centennial Review* 3(3):257–337

- [62] Xiao J, Li Z, Xie T, Pang T, Liu S, Zhang M, Cheng J, Liu B (2025) On the algorithmic bias of aligning large language models with RLHF: preference collapse and matching regularization. *J Am Stat Assoc* 120:552. <https://doi.org/10.1080/01621459.2025.2555067>
- [63] Young IM (2011) *Responsibility for justice*. Oxford University Press
- [64] Zuboff S (2019) *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. PublicAffairs