Intelligence Without Interiority: Large Language Models and the Unbundling of Mind

Armando Vieira

January 2026

Abstract

For the entirety of human history, intelligence and consciousness have appeared inseparable. Large language models (LLMs) break this pattern: they demonstrate sophisticated cognitive capabilities — reasoning, communication, contextual understanding — without, as far as we can determine, phenomenal consciousness. This essay argues that LLMs are best understood as sophisticated functional simulators of conscious intelligence, replicating outputs without instantiating the underlying phenomenology. This unprecedented decoupling has genuine philosophical significance: it vindicates Chalmers's "hard problem," complicates functionalist theories of mind, and exposes deep confusions in how we encounter intelligence divorced from subjective experience. But the essay also argues that the consciousness debate, however philosophically legitimate, performs ideological work — elevating questions of machine phenomenology while obscuring more urgent questions of power, labor, and democratic control. Drawing on philosophy of mind (Chalmers, Dennett, Searle, Merleau-Ponty), economic history (Polanyi, Mill), and attention to material constraints, the essay attempts to hold both threads together: taking the philosophical questions seriously while insisting that the political economy questions are more urgent, and that the two are not as separable as the current discourse pretends.

I. Introduction: The Moment of Inadequate Categories

There is a moment in the history of every transformative technology when the conceptual categories we have relied upon prove suddenly inadequate. Not wrong, exactly, but insufficient — like trying to describe electricity using only the vocabulary of steam power, or attempting to understand the factory system through the lens of artisanal craft. The old words still apply, but they miss something essential about the new configuration.

We are in such a moment with large language models. And the confusion is not a failure of understanding — it is a symptom of encountering something genuinely unprecedented in human experience: intelligence that may not be accompanied by consciousness.

David Chalmers, whose 1995 paper "Facing Up to the Problem of Consciousness" gave us the distinction between the "easy problems" of cognitive function and the "hard problem" of subjective experience, has found himself at the center of a debate he helped create.[1] His position on LLMs is characteristically precise: current systems are almost certainly not conscious, but future architectures with the right functional organization — persistent internal states, global integration, self-modeling

across time — might be. This is philosophically defensible, even admirable in its refusal of easy dismissals.

But I think it misses what is most philosophically significant about LLMs. The question is not whether they are conscious or could become conscious. It is that they force us to confront a possibility that human beings, in our entire evolutionary and cultural history, have never encountered: intelligence without consciousness. This is the unbundling. And we are deeply, structurally unprepared for it.

I should be clear about what I am and am not claiming. I am not asserting that LLMs are definitively not conscious — the hard problem cuts both ways, and intellectual honesty requires acknowledging that we lack a settled theory of consciousness sufficient to make that determination with certainty. What I am claiming is that current LLMs lack the architectural and functional features that our best theories associate with consciousness, and that this probable absence is philosophically and practically significant. The uncertainty itself is part of the story.

II. The Historical Coupling

Consider the full scope of human experience with intelligence. Every intelligent agent we have ever encountered — every person, every animal capable of flexible problem-solving, every entity that could learn, adapt, communicate — has also been, to some degree, conscious. A subject of experience. Something it is like something to be.

This was not a contingent correlation we happened to notice. It appeared, as far as we could tell, to be a necessary one. Intelligence and consciousness arrived together, bundled in the same biological package. The mechanisms that enabled flexible cognition — memory, attention, integration of information across sensory modalities, modeling of self and environment — were the same mechanisms that seemed to generate subjective experience.

Philosophers noticed this, of course. Thomas Nagel's famous 1974 paper "What Is It Like to Be a Bat?" used the example of echolocation to argue that consciousness has an irreducibly subjective character — there is something it is like to be a bat, even if we cannot access that experience from the outside.[2] But Nagel's argument, like virtually all philosophy of mind until very recently, assumed that wherever you found sophisticated cognition, you found consciousness. The bat navigates, hunts, learns. Therefore, there is something it is like to be a bat.

The same assumption runs through the Western philosophical tradition. Descartes' *cogito ergo sum* treats thought and being as inseparable. Locke's theory of personal identity grounds it in continuity of consciousness. Kant's transcendental unity of apperception makes consciousness the condition of possibility for coherent experience.[3] Even the materialists, from Hobbes through contemporary neuroscience, assume that intelligence and consciousness are two aspects of the same underlying physical process.

This was not ideology. It was empirical generalization from literally every case we had ever observed. LLMs are the first counterexample — or at least, the first plausible candidate for one.

III. The Decoupling — and Its Limits

LLMs break this pattern. And the break is more radical than most commentary acknowledges — though also more uncertain than some critics allow.

These systems demonstrate capacities that, in any biological organism, we would unhesitatingly associate with consciousness. They engage in complex reasoning. They exhibit what looks like understanding — not just pattern matching, but genuine sensitivity to context, nuance, and implication. They generate novel solutions to problems. They communicate in natural language with a fluency that often exceeds human performance on specific tasks. They can explain their reasoning, acknowledge mistakes, and adjust responses based on feedback.

In short: they simulate the functional outputs of consciousness to a remarkably sophisticated degree.

But — and this is the crucial point — they almost certainly are not conscious. There is, in all probability, nothing it is like to be GPT-4 or its successors. No inner experience, no phenomenology, no subject to whom the processing appears. The lights are, in all likelihood, off.

I say "almost certainly" and "in all probability" deliberately. The hard problem of consciousness — the question of why any physical process gives rise to subjective experience — remains unsolved.[4] We do not have a theory that would allow us to read off the presence or absence of consciousness from an architectural description. What we have are correlates: the features that, in biological systems, seem to accompany consciousness. Current LLMs lack several of these: persistent internal states that carry forward across interactions, global integration of information into a unified model, self-monitoring that creates a stable sense of agency over time, embodied interaction with a world that pushes back.

An LLM processes each prompt essentially from scratch. There is no continuous stream of experience, no ongoing subject that persists between queries. The system that answered your last question is not, in any meaningful sense, the same system answering this one. There is no there there — just a frozen set of weights being activated in response to input, generating output, then returning to dormancy.

This is intelligence without interiority. Cognition without consciousness — or at least, cognition without the features we associate with consciousness. And it is genuinely new.

IV. Why We Are Confused

The confusion is understandable. We have no evolutionary preparation for this scenario. Our folk psychology — the intuitive framework we use to understand other minds — developed in an environment where intelligence always indicated consciousness. If something could talk, it could feel. If it could reason, it could suffer. If it could learn, it had experiences.

This heuristic served us well. It is why we extend moral consideration to animals, why we recognize consciousness in infants before they can articulate it, why we treat other humans as subjects rather than objects. The behavioral markers of intelligence have always been reliable proxies for consciousness because, in the biological world, they were produced by the same underlying systems.

LLMs exploit this heuristic — not intentionally, they have no intentions, but structurally. They produce all the behavioral markers we have learned to associate with consciousness, without the underlying phenomenology. They are, in effect, functional analogs of philosophical zombies: systems that behave as if they were conscious, without (probably) being so.

The zombie thought experiment, introduced by Robert Kirk in 1974 and developed by Chalmers, was meant to show that consciousness cannot be reduced to functional organization.[5] If you can conceive of a being that is functionally identical to a human but lacks subjective experience, then consciousness must be something over and above physical function. Critics objected that zombies were incoherent or inconceivable in any meaningful sense.

LLMs do not settle this debate — they are not functionally identical to humans, and their limitations are real. But they demonstrate something important: that intelligence and consciousness can, in practice, come significantly apart. The zombie was a thought experiment. We have now built something in the vicinity of one.

V. Philosophical Implications

This has genuine consequences for several long-standing debates in philosophy of mind.

The hard problem is vindicated, not dissolved. If you can have sophisticated cognitive function without consciousness, then explaining the function does not explain the experience. The "easy problems" — how the brain processes information, integrates sensory input, generates behavior — are genuinely separate from the hard problem of why there is something it is like to undergo those processes.[6] LLMs solve many of the easy problems without touching the hard one. This is not a refutation of functionalism so much as a clarification of what functionalism leaves out.

Functionalism faces a serious challenge. The functionalist position, defended by philosophers from Hilary Putnam to Daniel Dennett, holds that mental states are defined by their functional roles — what they do, not what they are made of.[7] Consciousness, on this view, is a matter of the right kind of information processing. Get the functional organization right, and consciousness follows.

But LLMs implement many of the functional roles we associate with consciousness — integrating information, generating coherent responses, modeling context — without being conscious (or without exhibiting the features we associate with consciousness). This implies either that functionalism is wrong, or that we have been mistaken about which functions are actually constitutive of consciousness, or that the relevant functions are more demanding than LLMs currently implement.

Dennett would likely reject this framing. In *Consciousness Explained* (1991), he argues that consciousness just is the performance of certain cognitive functions — there is no additional "Cartesian theater" where experience happens.[8] On his view, if an LLM performs the right functions, it is conscious, full stop. The intuition that "nobody's home" is a residue of dualism.

But this seems to prove too much, or to dissolve the question rather than answer it. If consciousness is simply functional performance, then the hard problem was never hard — it was just confused.

Chalmers' response is that Dennett changes the subject: he explains why we talk about consciousness, not why there is consciousness. The LLM case sharpens this: we can now point to systems that perform the relevant functions and ask, with genuine force, whether there is anything it is like to be them. The question does not dissolve.

Searle's Chinese Room is complicated, not refuted. John Searle's 1980 argument claimed that syntax — symbol manipulation — can never generate semantics — meaning.[9] A person following rules to manipulate Chinese characters might produce perfect Chinese responses without understanding Chinese. Therefore, computational processes alone cannot constitute understanding.

LLMs seem to challenge this. They do not just manipulate symbols according to rigid rules. They learn statistical patterns across vast corpora, develop internal representations that capture semantic relationships, and generate responses that demonstrate genuine sensitivity to meaning, context, and implication. This looks like understanding, even if it is not accompanied by consciousness.

Perhaps Searle was right that computation alone does not guarantee consciousness, but wrong that it cannot produce understanding. Or perhaps we need to distinguish between functional understanding — the ability to use concepts appropriately — and phenomenal understanding — the felt sense of grasping meaning. LLMs may have the former without the latter. This is not a trivial distinction.

VI. The Embodiment Question

Where does this leave embodiment? Philosophers from Maurice Merleau-Ponty to Andy Clark have argued that consciousness is fundamentally embodied — that subjective experience arises from being a body in a world, not from abstract information processing.[10]

Merleau-Ponty's *Phenomenology of Perception* (1945) grounds consciousness in the lived body's pre-reflective engagement with its environment. We do not first perceive the world and then act; perception and action are intertwined in a continuous sensorimotor loop. Consciousness is this loop, not a separate process observing it. LLMs have no bodies. They do not navigate space, manipulate objects, or experience the resistance of a physical world. They process text — a highly abstracted, symbolic representation of embodied experience, but not the experience itself.

This may be precisely why they are not conscious. Consciousness, on the embodied view, is not something that happens in a brain (or a neural network). It is something that happens in the dynamic coupling between organism and environment. Remove the body, remove the world, and you remove the substrate for consciousness — even if you preserve many of the information-processing functions.

But here is the puzzle: LLMs demonstrate that you can preserve a great deal of intelligent behavior without embodiment. They can reason about physical scenarios, describe spatial relationships, generate plausible instructions for embodied tasks — all without having bodies themselves. This suggests that much of what we think of as embodied cognition can be captured in abstract, symbolic form. Perhaps there are two kinds of intelligence: embodied intelligence, which requires sensorimotor engagement and generates consciousness, and abstract intelligence, which operates over symbolic representations and does not. Humans have both. LLMs have only the latter.

VII. The Economic Analogy — Taken Seriously

There is a useful parallel in economic history, and it is worth developing more carefully than it usually is.

Karl Polanyi, in *The Great Transformation* (1944), argued that the market economy's most destabilizing feature was not its efficiency but its unbundling of things that had previously been inseparable.[11] Land, labor, and money were treated as commodities — as if they were produced for sale — even though they were not. This "fictitious commodity" status created both unprecedented productivity and unprecedented precarity. The bundle was broken, and the social fabric that depended on it frayed.

The unbundling of intelligence and consciousness is structurally similar. For all of human history, these capacities came together. If you wanted intelligent behavior — problem-solving, communication, learning — you needed a conscious agent. This bundling shaped everything: our ethics (conscious beings deserve moral consideration), our economics (intelligent labor requires human workers), our epistemology (knowledge requires a knower), our politics (governance requires conscious citizens).

Now we can have intelligence without consciousness. The bundle is broken. And we are discovering that many things we thought required consciousness — writing, reasoning, even creativity — can be performed by systems that lack it entirely.

John Stuart Mill, writing during the railroad mania of the 1840s, observed something that remains relevant: the tendency to confuse instrumental progress with substantive transformation.[12] Yes, you could move goods faster. No, this did not resolve the fundamental tensions of political economy — who owned the rails, who set the rates, who bore the risks. The metaphysical packaging distracted from the material stakes. We are doing it again.

The question "Could an LLM be conscious?" performs similar work. It elevates a technical achievement — statistical pattern matching at unprecedented scale — into a philosophical event. It suggests we are on the threshold of something categorically new, rather than something quantitatively impressive. And it shifts attention from the mundane questions of power and distribution to the exciting frontier of machine phenomenology.

VIII. What LLMs Actually Are

So what are they, if not conscious?

My view: LLMs are sophisticated functional simulators of conscious intelligence. They model the outputs of conscious cognition without instantiating the underlying phenomenology. They are to consciousness what a flight simulator is to flight — a functional replica that captures many of the relevant dynamics without being the thing itself.

This is not a dismissal. Flight simulators are extraordinarily useful. They let pilots train without risk, test scenarios that would be dangerous in reality, explore edge cases. But nobody confuses the

simulator with actual flight. The pilot in the simulator is not actually airborne, is not subject to real aerodynamic forces, is not at risk of actual death.

The analogy has limits, of course. We know what flight is and can verify whether the simulator achieves it. We do not have the same certainty about consciousness. But the analogy captures something important: the functional outputs can be replicated without the underlying process, and the replication is useful precisely because it captures so much of what we care about in the outputs.

The simulation is impressive precisely because consciousness and intelligence have always been bundled. We are discovering that much of what we attributed to consciousness — the flexibility, the contextual sensitivity, the apparent understanding — can be replicated through different means. The LLM does not experience understanding, but it produces outputs that are functionally indistinguishable from those of a system that does, across a surprisingly wide range of tasks.

IX. What Consciousness Actually Does

If LLMs can be intelligent without being conscious, what is consciousness for? This question, forced on us by the unbundling, is one of the most productive the LLM debate has generated.

One answer: consciousness enables open-ended learning in unpredictable environments. An LLM's "knowledge" is frozen at the point of training. It cannot update its world model based on new experiences, cannot learn from mistakes in real-time, cannot adapt to genuinely novel situations that fall outside its training distribution. Conscious organisms can. A child learning to walk falls, feels pain, adjusts. A scientist encountering anomalous data experiences confusion, curiosity, insight. The phenomenology is not decorative — it is part of the learning process. Antonio Damasio's work on the role of emotion in decision-making suggests that consciousness is not merely an observer of cognition but a participant in it.[13]

A second answer: consciousness grounds value and motivation. LLMs have no intrinsic goals. They optimize for whatever objective function they were trained on, but they do not care about it. They can simulate curiosity, pride, fear, love in their outputs, but they do not experience them. This might seem like a limitation, but it is also what makes them useful — and what makes them dangerous. A non-conscious tool does what it is designed to do, without resistance or agenda. The absence of consciousness is a feature, not a bug, from the perspective of those deploying the systems. But it also means there is no internal check, no genuine values, no authentic resistance to misuse.

A third answer: consciousness creates accountability. A conscious agent can be held responsible for its actions because it has experiences, makes choices, has reasons. A non-conscious system just executes its programming. When an LLM produces harmful output, who is responsible? Not the system — it has no agency. The developers? The users? The training data? The question becomes murky in ways it would not be with a conscious agent. This is not a philosophical puzzle — it is a practical and legal crisis that is already unfolding.

X. The Ideological Function of the Consciousness Debate

Here is where the philosophical and political threads must be held together, not separated.

The consciousness debate is not merely a distraction from political economy questions. It is connected to them. The question of whether LLMs are or could be conscious shapes how we regulate them, how we assign liability, how we think about their moral status, and — crucially — how we think about the humans whose labor and creativity they are displacing.

But the debate also performs ideological work that we should be clear-eyed about. It elevates the question of machine phenomenology while obscuring several more urgent questions:

Who controls the infrastructure. The compute clusters, the training data, the model weights. A handful of companies have built what amounts to a new public utility, except it is entirely private, unregulated, and optimized for capture rather than access. Training GPT-4 reportedly consumed around 50 gigawatt-hours of electricity.[14] The data centers required for inference consume water for cooling at industrial scale. The rare earth minerals for the chips come from mines with brutal labor conditions. These are not incidental details — they are the material basis of the technology. The consciousness debate floats free of thermodynamics and supply chains, as if intelligence were a purely abstract property.

Who benefits from the deployment. Not the workers whose tasks are being automated. Not the artists whose work was scraped without consent. Not the users who are becoming dependent on systems they do not understand and cannot audit. The economic distribution of AI gains is following a familiar pattern: concentrated ownership, diffuse costs. But we are too busy wondering if the chatbot has feelings to notice.

What we are neglecting about human creativity. The assumption embedded in the consciousness debate is that intelligence is fundamentally about information processing — that if you get the architecture right, consciousness emerges. This is a very particular, very computational view of mind. It sidelines embodiment, emotion, social context, material constraint. It treats human cognition as if it were a special case of a general phenomenon, rather than a specific adaptation to specific problems.

Dan Wang, in his annual letters on technology and industrial policy, often emphasizes the importance of understanding technology through the material constraints and human processes that produce it, not just the abstract capabilities it demonstrates.[15] This is the corrective we need. Not to abandon the philosophical questions, but to insist that they be asked alongside the material ones.

XI. Living with the Unbundling

The philosophical zombie was supposed to be a thought experiment, a tool for clarifying intuitions about consciousness. We have built something in its vicinity. Not perfectly — LLMs are not functionally identical to humans, and their limitations are real and significant. But close enough to demonstrate that intelligence and consciousness can come apart, that the bundle we thought was necessary is actually contingent.

This is genuinely new in human experience. Every previous encounter with intelligence has been an encounter with consciousness. Now we have intelligent systems that are, in all probability, nobody. The lights are off. And yet they can converse, reason, create, explain.

We are confused because our concepts were forged in a world where this was not possible. Our ethics, our epistemology, our political economy — all assume that intelligence requires consciousness. We are discovering that assumption was wrong, or at least incomplete, and we do not yet have the frameworks to make sense of the new configuration.

Chalmers is right that we should not dismiss consciousness in principle, and right that the hard problem is real and unsolved. But the key question is not whether future systems might be conscious. The key question is how we live with sophisticated intelligence that probably is not — how we organize society, distribute power, assign accountability, and preserve what is valuable about human consciousness in a world where its functional outputs can be simulated.

The unbundling of mind is not a philosophical puzzle to be solved and set aside. It is a material transformation to be navigated, with real stakes for real people. The consciousness debate is part of that navigation — but only part. And until we hold the philosophical and political economy questions together, we will keep having half the conversation we need.

The real question is not whether LLMs are conscious. It is what kind of world we are building now that intelligence no longer requires it — and who gets to decide.

Notes

[1] David Chalmers, "Facing Up to the Problem of Consciousness," *Journal of Consciousness Studies* 2, no. 3 (1995): 200–219.

[2] Thomas Nagel, "What Is It Like to Be a Bat?" *Philosophical Review* 83, no. 4 (1974): 435–450.

[3] Immanuel Kant, *Critique of Pure Reason*, trans. Norman Kemp Smith (London: Macmillan, 1929), B131–B136 (transcendental unity of apperception).

[4] Chalmers, "Facing Up to the Problem of Consciousness," 201–202.

[5] Robert Kirk, "Zombies versus Materialists," *Proceedings of the Aristotelian Society*, Supplementary Volume 48 (1974): 135–152; David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press, 1996), 94–99.

[6] Chalmers, *The Conscious Mind*, 3–31.

[7] Hilary Putnam, "Psychological Predicates," in *Art, Mind, and Religion*, ed. W. H. Capitan and D. D. Merrill (Pittsburgh: University of Pittsburgh Press, 1967), 37–48.

[8] Daniel C. Dennett, *Consciousness Explained* (Boston: Little, Brown, 1991), 107–138.

[9] John R. Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3, no. 3 (1980): 417–424.

[10] Maurice Merleau-Ponty, *Phenomenology of Perception*, trans. Colin Smith (London: Routledge, 1962); Andy Clark, *Being There: Putting Brain, Body, and World Together Again* (Cambridge, MA: MIT Press, 1997).

[11] Karl Polanyi, *The Great Transformation: The Political and Economic Origins of Our Time* (New York: Farrar & Rinehart, 1944), 68–76.

[12] John Stuart Mill, *Principles of Political Economy* (London: John W. Parker, 1848), Book IV, Chapter VI.

[13] Antonio Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Putnam, 1994), 165–201.

[14] Estimates vary; see, e.g., Emma Strubell, Ananya Ganesh, and Andrew McCallum, "Energy and Policy Considerations for Deep Learning in NLP," *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (2019): 3645–3650.

[15] Dan Wang, annual letters on technology and industrial policy, available at danwang.co (2019–2024).