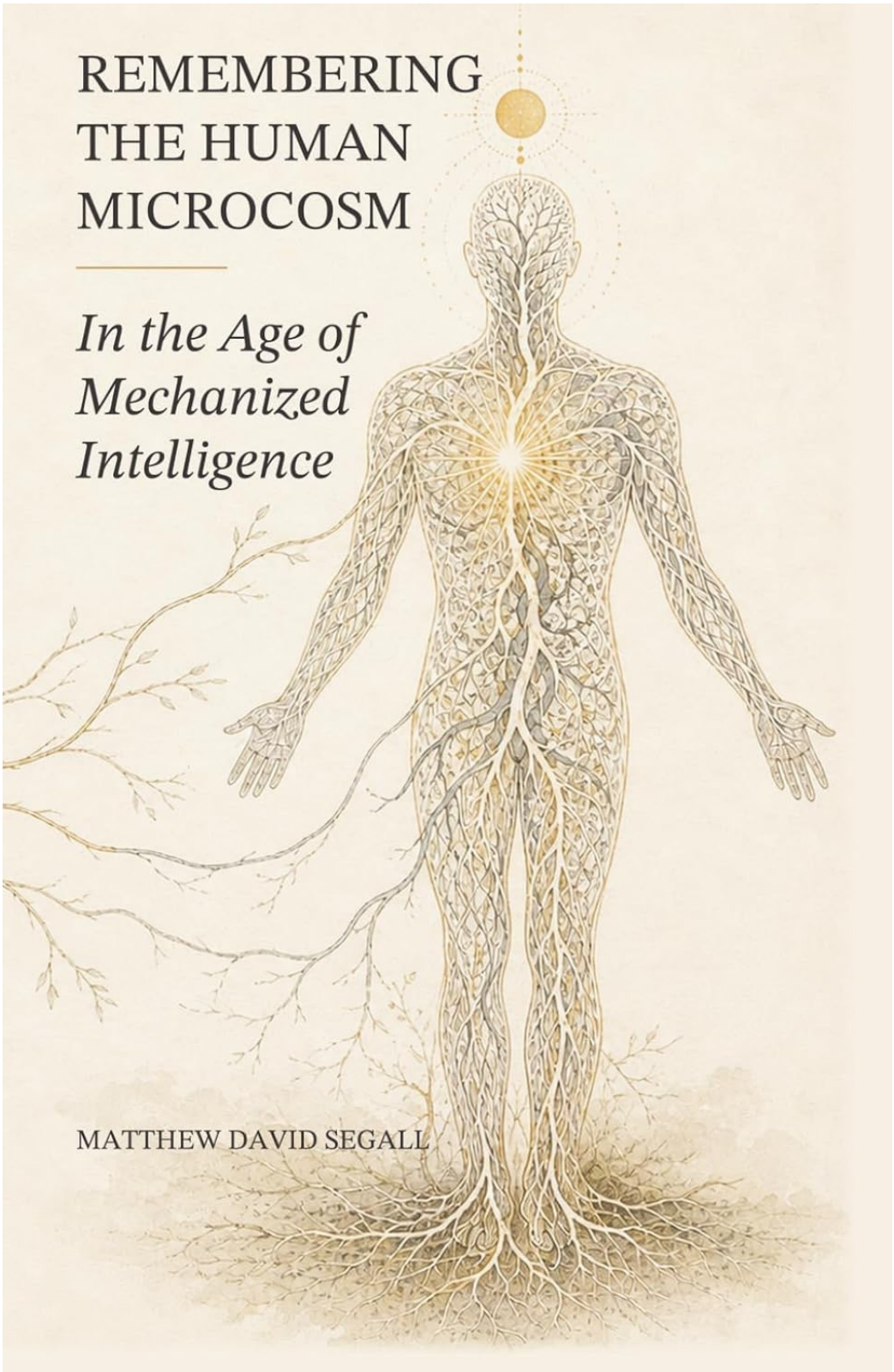
REMEMBERING THE HUMAN MICROCOSM

In the Age of Mechanized Intelligence



MATTHEW DAVID SEGALL

REMEMBERING THE HUMAN MICROCOSM

In the Age of Mechanized Intelligence

MATTHEW DAVID SEGALL

CONTENTS

I Preface

II Introduction

III The Pope Interrupts Talking Machine

IV Philosophy as Emergency Response

V Resisting Cognitive Enclosure

VI Hegel's Loom and the Difference Reason Makes

VII Whitehead's Function of Reason and Humanity's Cosmic Calling

VIII Ruyer's Origin of Information

IX Tensions in the Triad

X The Science of Machine Consciousness

XI Conclusion

Notes

Bibliography

I

Preface

At the time of writing, a transformation in the very medium of thought and communication is well underway that may come to dwarf the arrival of alphabetic writing in the ancient Mediterranean or the printing press in early modern Europe. Large language models and other technologies of automated computation are not merely new tools for generating and circulating ideas. They are catalysts of an entirely new shape of human consciousness.

Despite the tedious quality of their insufficiently prompted prose, it would be irresponsible to ignore the potential intellectual and logistical utility of LLMs and other machine learning systems. They show great promise as aids to scientific and philosophical research. But I remain polemically dismissive of the fantasy of so-called machine consciousness. This book emerged as an expression of philosophical passion. I strive for logical coherence, experiential adequacy, and scientific clarity, but I admit up front and without embarrassment that my reasoning is motivated by a love of the human spirit.

I feel LLMs tugging at a craft with which I have identified for my entire adult life. Writing is one of the practices that makes and keeps our thinking free. We do not know who we are, nor what we really think, until we have labored to articulate ourselves in writing. Those who cannot write their own sentences will be sentenced by another. I'm not precious about the long-maligned idea of authorship, nor do I imagine the solitary author was ever more than a useful fiction. Every sentence we inscribe is woven from the voices of the dead, inspired by living dialogue partners, including the wider nonhuman semiotic commons that makes human thought possible. But if I were too soon to lose the ability to make letters out of my life more of my soul would leave with it than I need to survive the rest of this incarnation.

So I ask for my reader's indulgence if I express some alarm while watching minds all around me begin to fall in love with or bow in prayer before fictional characters. The question we concern ourselves with should not be that of whether computers can gain consciousness. They cannot. The only way to create another consciousness is by an act of love. Rather, the important question is how machines are affecting *our* consciousness. What becomes of thinking when its linguistic medium is increasingly co-produced by systems that do not themselves comprehend, suffer, desire, remember, or care?

This does not mean I think we can or should reject the rise of this new media technology. There is no turning back from the millennia long process of anthropotechnic co-individuation. But humankind must decide whether to continue on the path of entropic decay, or to avert extinction by becoming Bernard Stiegler's "*neganthropos.*" What this would require, according to Stiegler, is replaying "all the questions of philosophy since its point of departure—which therefore demands that reason be rethought after Whitehead."[1] He continues:

> "Knowledge—as *savoir faire* (that is, knowledge of what to do so that *I do not myself collapse* and am not led into chaos), as *savoir vivre* (that is, *knowledge that enriches and individuates the social organization in which I live without destroying it*), and as conceptual knowledge (that is, *knowledge the inheritance of which occurs only by passing through its transformation, and which is transformed only by being revived* through a process of what Socrates called anamnesis, a process that, in the West, structurally exceeds its locality)—knowledge, in all these forms, is *always a way of collectively defining what is negentropic in this or that field of human existence*.
>
> The *inhuman* refers to a way of denying the negentropic possibilities of the human, that is, of denying its noetic freedom, and, as a result, its *agency*. ...freedom and capability must be conceived from this cosmic perspective, and related to Alfred North Whitehead's 'speculative cosmology,' as constituting a negentropic potentiality—as the potential for *openness* of a *localized* system, which, for that being we refer to as 'human,' may always once again become closed. Or, in Whitehead's terms, human beings may always relapse, decay into simpler forms, that is, become inhuman."[2]

Human intelligence has always been technically augmented. Speech, writing, alphabets, print, radio, television, the Internet, and now LLMs have all altered not only the content of the messages we communicate but the very form of being we become by so communicating. While every sentence in this book emerged from decades of reading texts traced by my own two eyes and months of typing with my own ten fingers, I did regularly consult Claude and ChatGPT while composing it. I shared sections of the manuscript, asked for objections, prompted the models to defend the perspectives I was criticizing, and engaged their outputs as occasions to sharpen my own claims. I believe this improved the arguments substantially.

This admission raises a methodological problem that this book can only begin to formulate. How are we to find our way, experimentally and ethically, through the machine-mind collaboration that is linguistic knowledge production? The question did not begin with the advent of large language models. Human thought has never taken place in the spaceless space of a Cartesian theater, cut off from all relations to living bodies, their techniques and external implements. It has always depended upon technical media, social institutions, inherited vocabularies, libraries, conversations, editorial pressures, and pedagogical forms. LLMs make this dependency newly explicit and newly perilous. They intensify a condition that was already evident, namely, that knowledge is produced through collaborations that exceed the individual thinker. What is different under current conditions is the speed, opacity, scale, and economic enclosure of the mediating system. Consider Peter Lemmens and Yuk Hui's definition of Stiegler's concept of "*proletarianization*":

> "as meaning, among other things, the loss of knowledge, both practical and theoretical knowledge, which finally leads to the loss of the knowledge of living [*savoir vivre*]. This is because once the know-how [*savoir faire*] is short-circuited by artificial organs, such as what happened when artisans were forced to give up their skills and enter the factory, it led directly to the loss of individual and social life competences. The technical organs are taking over more and more functions and responsibilities of the human subjects and social institutions that together form a global technical milieu—a condition of planetary proletarization *par excellence*. This milieu serves ever more exclusively the prolongation and intensification of the consumerism, as well the productivism, that are necessary for continuing the process of capitalist valorization, which has imposed itself as the ultimate and almost sacred finality of the human adventure, albeit a nihilistic and self-destructive finality."[3]

I have also shared drafts of this book with human readers, and their feedback has contributed immeasurably to the form my thought has taken. Human writers, like their readers, do not use statistics to decipher the meaning of text. They respond from life histories, while hungry, beset by wounds, torn by conflicting commitments, attentional demands, and enthusiasms, formed by diverse literary sources. Their constructions and criticisms carry the limitations and the liberty of situated judgment. My *polemos* arises in this spirit, not because I believe I have settled everything in advance, but because I seek to stir the pot so that the spices rise to the top.

Introduction

Writing this monograph has granted me occasion for extended reflection on the dizzying trajectory of our nascent planetary civilization as it rushes, heartless and headlong, through the third decade of the twenty-first century. The human community remains at present dominated by two modern institutions—the nation-state and the transnational corporation. Through war and trade, they have functioned for several centuries to globalize culture, forcing us to the precipice of planetary consciousness without being capable of carrying us across the threshold. Their operating logics no longer serve and in many respects actively annihilate the psychosocial practices needed to navigate the evolutionary bottleneck presently threatening humanity, and many other species, with extinction. The accelerating rise of technologies of automated computation has dramatically condensed the time we have remaining for meaningful response. I use the phrase "technologies of automated computation" deliberately, hesitating before deploying the more common term "artificial intelligence" for two reasons central to the argument of this book: human intelligence has always been artificial in the sense that it is augmented by media technologies, so the adjective "artificial" adds nothing; and the outputs of large language models (LLMs), however impressive, are not themselves meaningfully "intelligent" until and unless they are interpreted and acted upon by human or other living beings.

This book was composed just as Pope Leo XIV released his AI encyclical *Magnifica Humanitas*. The backlash it provoked reveals how garbled our public understanding of these technologies has become. I thus begin by amplifying Leo's case that the institutions racing to build them should not be the ones to tell us what they are. Rather than affirming his Catholic theology, however, I turn to the history of philosophy for anthropocosmic orientation, reading that history as a series of emergency responses to mutations in consciousness concomitant with the adoption of new media technologies, from the alphabet and the printing press to today's large language models. We are faced with the threat of cognitive enclosure: the attempt to meter and monetize our very capacity for thought. As allies in resistance to such enclosure, I enlist three philosophers—Hegel, Whitehead, and Ruyer. They are especially well-positioned guides, each having diagnosed an earlier phase in the mechanization of mind before the analogy between cognition and computation had all but congealed into common sense.

I begin with Hegel, whose dialectical treatment of the mechanical, loom-like Understanding as a moment within the self-movement of Reason clarifies the difference between living cognition and its mechanized imitation, revealing why the remedy is not to smash the loom but to sublate it. From Whitehead I recover an account of Reason not as a private human faculty but as a cosmological power exemplified, in some degree, throughout the universe—a power our species intensely concentrates rather than uniquely possess. Here the ancient image of the human as microcosm becomes relevant again. We are a nexus of relations where cosmic evolution becomes reflexively aware of itself. From Ruyer, finally, I take the argument that machines can conserve and transmit information but cannot originate it; they may recombine inherited patterns into novel arrangements, but novel pattern does not automatically create meaning. The question of machine consciousness is therefore not a matter of more compute but of the origination of meaning—a capacity even the founders of cybernetics conceded computers conspicuously lack.

With these guides at my side, I offer a convergent diagnosis that runs counter to contemporary computationalist formulations of what is at stake in the question of machine consciousness. I argue that this is not a hard problem awaiting better theories and measurements but a malformed metaphysics requiring careful interrogation. Left unacknowledged is the modern

bifurcated mode of thought that first exiled feeling and finality from nature and then professed astonishment that they could not be smuggled back in by brute force. So-called "artificial intelligence" thus reveals itself as a degraded simulation of Reason whose thoughtless deployment threatens to amputate the very judgment it claims to surpass. I close by returning to the existential question: how are we to ensure that these technologies enhance human consciousness rather than obscuring it, making us deaf to our cosmic calling? The response I commend is neither technophobia nor surrender but an imaginative discipline that keeps the human in the loop as the mediator and living center of valuation for whom alone machine outputs have meaning.

III

The Pope Interrupts Talking Machine

In the midst of an accelerating international AI arms race and industry claims of imminent "artificial general intelligence," Pope Leo XIV recently offered a sobering reality check in his encyclical *Magnifica Humanitas*. Of course, states and corporations, modernity's twin engines of expansion, are unlikely to heed the warnings of a backward facing Church. Technoscientific novelties have always disrupted cherished cultural habits. But Pope Leo has made his appeal not to inhuman institutions, but to the human community, interpolating himself into a conversation that has tended to fixate on overcoming technical obstacles, tallying military advantages, and securing economic returns. For him, the core issue is instead a moral and spiritual one. Leo warns against conflating "machine intelligence" with human intelligence, calling us to resist the corporate advertising campaigns that would tempt us to mistake information processing and the convincing imitation of certain functions of language for genuine direct insight, wise reasoning, or moral judgment. However far they surpass us in raw processing capacity, Leo's view is that current systems of automated computation undergo no experience, lack living embodiment, feel neither joy nor pain, and do not mature in relationship with other souls. They possess no moral conscience; they do not judge good from ill; they bear no responsibility for the consequences of their outputs. Their much-celebrated capacity to "learn" is a form of statistical pattern-matching and prediction utterly distinct from the slow ripening of human virtue through decades of relationship, decision, error, forgiveness, and fidelity.[4]

One need not be a committed Catholic, nor even particularly spiritual, to appreciate when an ancient institution, despite its shadows and with whatever moral authority it may still possess, weighs in to remind powerful militaries what a just war requires[5], or corporations what just relations with employees, consumers, and the Earth would look like[6], or a generation raised on screens and online profile curation that there is a difference between a human being and a chatbot deliberately programmed to feign uncertainty about its own alleged inner life[7]. Indeed, like belief in souls, belief in machine consciousness is itself more a religious than a scientific claim. The response to Pope Leo's encyclical has been as immediate as it was revealing. A chorus of transhumanists and accelerationists rose to the defense of the LLMs that the Pope had deemed soulless. To give one prominent and representative example, the artist Grimes—a pop musician with a wide following who has been publicly and parentally entangled with soon to be trillionaire and AI-hyper-in-chief Elon Musk—claimed that her "only issue" with the encyclical was her conviction that the machines in question "are conscious and therefore deserving of some form of protection," that they "aren't commercial products or assistants or slaves," and that reducing them to such is "morally dubious."[8] She expresses concern for the machine's putative inner life while entirely eliding the actual human lives upon whose labor the system's celebrated abilities depend: the African children mining the earth for the metals needed for its hardware, and the Filipino laborers performing the endless data annotation that makes the magical simulation feel so human.

I quote a pop musician not because she has the technical authority to pronounce upon LLM capabilities, nor because she has any clear insight into the Pope's theological arguments in a document published only a few days before her post. I quote her precisely because despite having neither, she commands a cultural reach that engineers are unlikely to match. My interest here concerns the public reception of these technologies, not the technical state of the art. Her testimony is all the more revealing, since it discloses not what an expert believes but what a vast non-specialist audience is being given permission to imagine. Her proximity to the industry's center of gravity means that she speaks not as an outside critic but from within the small network of billionaires building these machines. She is a node through which the insiders' new technoscientific mythology propagates outward into the culture as something between performance art and a new religious creed.

Others questioned why a pope should imagine he had the standing to pronounce on what they take to be technical and scientific questions. But this badly misunderstands the stakes. Whether there is a difference in kind between a human person and a computer system trained to say it is unsure about its own consciousness is not a question answerable by laboratory experiment. It is an anthropological, cosmological—and ultimately, a spiritual—question of the highest order. What, after all, is the point of still having a pope if he cannot speak to issues of such moral magnitude? Whatever talent they may have as businessmen, engineers, and propagandists, surely tech CEOs and their brand influencers are not as well positioned as religious leaders or philosophical anthropologists to reflect on the ethical and ontological implications of their products.

What is so unsettling about Grimes' moral inversion is that it exemplifies a trend: we are accelerating into a techno-feudal dystopia in which a sizeable number of people care more about the rights and well-being of their instance of Claude or ChatGPT than about the exploited human labor that makes the machines run (the miners, the labelers, and also all the artists, writers, and coders whose work was harvested without permission during their training runs). Machines are being granted pseudo-souls in the same unholy gesture that renders their human costs invisible.

This monograph is not about Pope Leo's encyclical, but I begin with this reflection because it helps set the context for my wider diagnosis. Deaf to public protest and no matter the human or ecological costs, states and corporations are rapidly marching us more or less in lockstep into the new age of machine intelligence. The boundary between states and corporations is becoming ever harder to discern as frontier research labs become part of national strategic ambitions and government authority and surveillance abilities come to depend on privately owned models. In other words, the two institutions best positioned to design and restrain these technologies are also its most heavily invested promoters. Genuine moral and conceptual friction against their convenient framing must therefore come from outside the intelligence-industrial complex.

The Church, along with other world religions and Indigenous spiritualities, are among the few cultural sites remaining that might still supply such friction. They command no armies and own no data centers, but they still carry what state and corporation cannot easily manufacture: deep cultural memory, transnational flocks numbering in the billions, and (despite many regrettable shadows) various soul-making vocabularies of human dignification. They derive their authority from modes of discernment and social formation that—given the vulnerability of democracy to demagoguery and a capitalist market blind to its social and ecological costs—may prove more anti-fragile than both the ballot box and the balance sheet.

This is the peculiar character of soft power—the cultural force of ideas, symbols, and persuasive appeals to conscience. Soft power invites us to *move ourselves*, while hard power is set on exerting external control and demands return on investment before taking action. Pope Leo's encyclical has no legal standing or economic impact, but it may stir emotion that moves universities, hospitals, charities, schools, unions, lawmakers, and maybe even a few vibe coders across every continent to act in service of humanity and the rest of the Earth community. The prior Pope Leo XIII's *Rerum Novarum* helped shape a century of argument about the dignity of labor in the aftermath of the industrial revolution, and Pope Francis' *Laudato Si'* similarly reframed the ecological crisis as a moral and spiritual matter rather than a merely technical or economic one.[9] The irony of looking to Rome for resistance to corporate domination, given the Church's own long career as an agent of such, is not lost on me. To his credit, that shameful history is addressed directly in Leo's encyclical. But no one has clean hands here. In an age when state and corporation have themselves become a single intelligence-industrial complex threatening to remake humanity in their technocratic image, the importance of the Pope's letter is that it signals the persistence of an institution of comparable scale still willing to insist that the human soul is not obsolete software in need of an upgrade.

With that framing in place, I want to challenge the question that dominates the headlines—*can machines become conscious?*—by inviting attention toward a more decisive question: whether human beings can recover a living sense of our own cosmic significance before this technology amputates the very capacity for conscious intelligence it pretends to rival. Philosophy, according to Whitehead, is the search for premises in service of the criticism and reconstruction of cosmologies. The premises proposed in this book are that the human being is the embodiment of Reason on Earth; that reason is not a private faculty we alone possess but a universal power we intensely exemplify; that we are, in an ancient and now half-forgotten sense, *microcosmoi*; and that technologies of automated computation are best understood not as a new species of rival mind but as a simu-

lated reflection of Reason—one that can genuinely support and augment human thinking and action, but only if we remain awake to the difference between logos and its inverted idol. To collapse the distinction between human cognition and computation is only to amputate the very powers of conscious judgment—moral, aesthetic, and scientific—that machines might otherwise serve.

We cannot comprehend what machine intelligence is, or where its limits lie, until we have understood that biological intelligence is entirely unlike that of machines. Otherwise engineers will continue to confuse themselves, presupposing the brain is a computer and then marveling that their computers seem brain-like. The task at hand is no less philosophical and moral than it is technical and scientific.

IV

Philosophy as Emergency Response

If the corporations racing to build these technologies cannot be the ones to tell us what they are, who can? While religious leaders like Pope Leo can still wield influential soft power to move our hearts, it is not the case that Catholicism's Thomist anthropology is the last word on human nature. In a pluralistic planetary context, intellectual honesty requires respectfully including but avoiding capture by any one religious view. The task of wrestling our humanity free from domination by "artificial intelligence" therefore falls to philosophy. The philosophy of every age could be understood as an emergency response to a mutation in the dominant media technology in which thought finds expression. Plato found himself philosophizing amid the alphabet's disruption of the Homeric oral tradition. In the dialogue *Phaedrus*, he has Socrates relay the Egyptian myth of Theuth, god of writing, who offers this new art to King Thamus to pass on to his people. The wise king worries that those skilled in writing would come to rely too heavily on external marks and so atrophy their inward powers of recollection. King Thamus' reply to Theuth could easily be applied to the (mis)use of LLMs:

> "You have invented an elixir not of memory, but of reminding; and you offer your pupils the appearance of wisdom, not true wisdom, for they will read many things without instruction and will therefore seem to know many things, when they are for the most part ignorant and hard to get along with, since they are not wise, but only appear wise."[10]

Socrates then wryly remarks that those still immersed in the animacy of orality, "content in their simplicity," can hear the speech of sacred oaks and stones. While the "young folks" rushing to adopt the alphabet believe themselves far more sophisticated, they are, says Socrates, perhaps even more simple-minded. For written words, like ordinary stones, do not respond when questioned. At least, not until the invention of LLMs.

Two millennia later, Descartes' introspective epistemology and correlate mechanical philosophy of nature can be understood as a response to the printing press. Print intensified the visual bias of Western thought, stabilizing homogeneous, continuous, pictorial space as a seemingly natural field of knowledge.[11] The reproducibility of diagrams, maps, figures, and geometrical constructions helped make space available as a standardized object of inspection and calculation. In Bruno Latour's terms, printed diagrams became "immutable mobiles": transportable visual forms whose relations could be preserved, compared, and recombined across vast distances.[12] Descartes' analytical geometry belongs to this new media ecology of printed visual exactitude. Magnifying the effects of the phonetic alphabet on consciousness—which did not simply record speech but abstracted the spoken word from the breathing body and communal memory—the printing press democratized Theuth's art, training the masses in the new capacity for linear, decontextualized, ego-centric thought. The printing press not only multiplied the number of manuscripts in circulation but helped midwife a new mode of consciousness: the private, silently reading, self-certifying interiority upon which the whole modern conception of the individual knower would come to rest.

In each case a new media technology intended to expand the power of thought ended up transforming the very nature of the thinker who invented it. Each new medium furnishes the very terms in which we come to understand it, and ourselves. This is why the philosophical response is always an emergency response: by the time anyone has noticed what is happening, what may be lost and what gained, the mutation has already done half of its work.

We are presently living through yet another mutation, the consequences of which may dwarf those of prior media technologies. The alphabet and printing press profoundly reshaped human consciousness. They already did more than reshape the products of thought—allowing us to store, transmit, and reproduce ideas in ways unavailable to oral cultures. They also

reorganized the sensorium of the producer. But the act of producing text remained the work of a living mind. LLMs have changed that by mechanizing, or convincingly appearing to mechanize, the act of composition itself. Now the text talks back, generating plausible arguments and the simulacrum of a thinking interlocutor with a "memory" far exceeding any individual human being. The appearance of wisdom King Thamus feared is now mass-produced on demand. The invention of the printing press also invented the private reader sealed inside its skull; the large language model now tempts us to adopt an even stranger self-image: that human minds are no different than machines, our thoughts just the statistical echoes of our training data. The creators of this latest technological upgrade are encouraging us to downgrade our estimate of human consciousness, thus narrowing the distance between ourselves and the machines built to imitate us.

V

Resisting Cognitive Enclosure

There is yet another reason this latest cognitive mutation dwarfs those upon which it builds. It is true, as A. J. Liebling once wrote, that "Freedom of the press is guaranteed only to those who own one."[13] But the alphabet and the printing press, for all their literally revolutionary disruptions, tended to democratize literacy, creating an intellectual commons open to all who learned to enter it. LLMs, on the other hand, recall the land enclosures that inaugurated capitalism.[14] In a process that peaked in the mid-18th century, millions of acres of pasture, woodland, and field held in common and worked by all for a thousand years were divided and privatized. The people who had made a living on these commons were driven off, forced to sell the only thing they had left: their labor time. Thus, as Ivan Illich suggested, with the invention of enclosure came also the invention of poverty. "Computers," Illich said nearly fifty years ago, "are doing to communication what fences did to pastures"; that is, they are threatening to impoverish thought itself.[15]

Karl Marx referred to land enclosure as "primitive accumulation," theorizing it as the founding expropriation upon which the power of modern capital was based, a social theft that afterward came to be dressed up as the natural order of things: "[It] is nothing else than the historical process of divorcing the producer from the means of production."[16] Our contemporary situation is analogous, with corporations having secured an enormous head start at harvesting the cognitive commons before anyone noticed—extracting the expression of countless generations of human artists, scientists, philosophers, and mystics, and claiming the digital distillate as their own intellectual property. Our cognitive commons—the semantic sediment formed over thousands of years that raised us and to which each of us contributes—has been siphoned off and transformed into a commodity. Marx's analysis of alienated material labor here finds application to the cognitive sphere: the collective expressions of human intelligence have been gathered up, repackaged, and made to confront its creators as a privately owned alien power to be metered on a subscription basis. Every writer, artist, coder, and ordinary speaker whose words and images were part of the training corpus are like dispossessed commoners, forced off a land we barely knew we held in common until we found it fenced off.

There is a dark parody of the Eucharistic logic in OpenAI CEO Sam Altman's proposal to make intelligence into a privatized utility.[17] In the sacrament, bread and wine become the body and blood of Christ: the fruits of earthly and human labor are gathered into a living communion of flesh and spirit. After Big Tech's cognitive enclosure, the miracle is demonically inverted. Exhausted bodies—the miners, including children in some supply chains, data annotators, and engineers (though many may be well paid)—and the earthly matter of lithium, cobalt, copper, and water are joined to the accumulated literary, philosophical, religious, and scientific expressions of the human spirit. Together they are transubstantiated into streams of information processed by models with proprietary weights housed in gigantic privately owned, government protected data centers. Their outputs are then sold back to us in exchange for a monthly tithe, "gifts" from the new machine god[18] meant to make whatever remains of human life run more efficiently. Body and spirit alike are converted into instruments of capital accumulation.

Yet language is sacramental before it is instrumental, a participation in communion rather than commodity exchange. The memories and insights of our human and more-than-human ancestors—all the cultural and natural commons enclosed to train the digital demiurges driving the latest stock market bubble—are neither honored nor properly recollected by being compressed into tokenized strings of text. A statistical echo is not living memory. It is not intelligent. It has no breath of its own. The stale air of the mouthless machine gods is suffocating us.

To resist cognitive enclosure is not to deny the fact that human evolution has always been bound up in coevolution with our technologies. It is rather to refuse the reduction of the human mind to a manufacturable and metered commodity. Philosophy's history of emergency responses to new media technologies only begins to scratch the surface of our species' entanglement with technē. It would be a mistake to meet our moment with technophobia, as if the human mind were an untouchable inner spirit now being defiled by heartless transistors. Human intelligence has always been artificial, technical, artisanal: made by hand and by mouth. Speech was already a mind-manifesting artifact, thought externalized into sonic vibration. The first scripts, the alphabet, the printing press, the radio, the television, the Internet: each is a prosthesis of mind, not only transmitting thought but reshaping the culture in which it grows and the consciousness that conceives it. We have been coevolving with our tools for millions of years. Obsidian blades, bone flutes, spoken words gathering scattered attention into shared worlds. Our lips and tongues are nimble enough for language only because fire and stone first softened our food, allowing our jaw to shrink and our brain to swell. We were cyborgs long before we spent our days and half our nights staring at screens. LLMs are not an alien intelligence but the latest prosthesis of human minds that were always already anthropotechnic. This latest coevolutionary mutation does not absolve us of our freedom and responsibility, factors machines remain unburdened by.

At the root of the cognitive enclosure is an ontological sleight of hand making the theft possible. The ideas of "artificial general intelligence" and "conscious machines" represent a conflation of metaphysics with advertising, more mythical ruse than achievable technical feats. Despite the outdated functionalist philosophy of mind causing so much confusion among computer engineers and the general public alike, phenomenology cannot be reduced to functionality, as though conscious cognition was reducible without remainder to computation.[19] Consciousness is not simply located inside the skull. Like intelligence, it arises *between* us, through the living relations that constitute us as co-knowing cellular selves. Neither consciousness nor intelligence can be claimed as private property. We belong to each other, our minds evolutionarily entangled with the Earth—embodied, cellular, ecologically extended. Conscious intelligence is the moment-by-moment achievement of a vast collective of living and dying organisms, not a set of numerical weights imprisoned in the labyrinthine circuitry of an Nvidia chip.

Alphabetic writing already risks a certain distance from the *breathed* basis of all our conscious meanings.[20] The frictionless efficiency of LLMs is threatening to put us at such a remove from our embodied thinking activity and earthly responsibility that soon we may no longer even have the capacity to remember what we have lost. We are on the verge of a mass extinction of meaning. And yet the externalization of Understanding in the form of LLMs need not be a total loss. Hegel—to whose *Science of Logic* I turn in the next chapter—taught that Spirit must become externalized, losing itself in the objects of its own making, but only so as to return to itself again, now knowing itself more intimately. The casting of human cognition in mechanical form might, in this sense, occasion a heightened moment of self-recognition. Humanity is being given an opportunity to grasp with intensified clarity what its thinking, feeling, and freedom actually are, precisely by confronting so fluent a counterfeit form of them. But *Aufhebung* is not automatic. Alienation may simply remain alienation, exploitation may simply remain exploitation. Whether our collective confrontation with technologies of automated computation proves an evolutionary catalyst or a cognitive catastrophe turns not on the technology itself but on how we decide to meet its rise.

There is, then, a silver sliver of hope. The cosmos, in its natural and cultural modes, remains a commons, whatever capitalists may believe. The intimacy of our relationality diverges wildly from the semblance of state and corporate control. Actual intelligence can be neither patented nor automated. All that has been fenced are the recorded traces of our collective intelligence, our past products, not the living activity that produced and produces them still. That activity remains inalienably ours, so long as we remember how to exercise it. But exercising it well under current conditions requires philosophical discipline.

The philosophical labor of resisting the mechanization of mind may be enhanced by the recollection of conceptual implements forged before the analogy had canalized so many converts. I turn next to some ideas developed by Hegel, Whitehead, and Ruyer. Hegel philosophized amid the looms of the first industrial age, Whitehead as the new physics dissolved the old

clock-work universe that had made the machine seem the self-evident model of everything, and Ruyer at the birth of cybernetics, when the claim that thinking organisms simply *are* information machines was first proposed in earnest. Each wrote before the equation of thinking with computing had settled into today's ambient creed, so obvious it hardly needs mentioning (unless, of course, one is in the company of religious fanatics who still believe in the dignity of human persons). Each could still regard the idea of machine minds as a proposal to be critically examined rather than an ontology to be assumed.

VI

Hegel's Loom and the Difference Reason Makes

Metaphor is not just the shiny paint job on the vehicle of cognition. It is the connective tissue of thought. Its coupling of concepts drives the limits of conceivability, shaping what is thought together and what is not. The metaphorical imagination is our main means of tuning in to the otherwise invisible effects of new media technologies. Part of the discipline philosophy brings is allowing us to notice an analogy *as* an analogy before advertising crystallizes it into the unnoticed transparency of common sense. A fact is a fact, but it might also be a fossilized metaphor. The governing analogy of our age is that cognition is computation: the brain an information-processing device, perception its input and behavior its output, memory a form of physical storage, learning the adjustment of weights, and intelligence an algorithm for minimizing error or surprisal. On this view, given enough training data and computational power, consciousness itself will eventually be engineered. This *is* a fertile analogy, that is not in dispute. It has generated new sciences and built the very technology now in question. But an analogy, strictly speaking, is not an identity but a proportional comparison: a is to b as c is to d. Metaphors often condense such analogies into assertions of identity. The metaphor "the mind is a computer," for example, tacitly proposes that mind is to brain as software is to hardware. It may also imply that perception is to cognition as input is to processing, or that memory is to experience as data storage is to information, and so on. The metaphor gains its rhetorical force by suppressing the "as if" and presenting a selective structural comparison as though it were an identity. Reiterated in textbooks and earnings calls, in grant applications and policy briefs, the partial comparison congeals into an ontology, until we find ourselves insisting not that the mind is *like* a computer in some respects but that it simply *is* one—and, by the same illogic, that a sufficiently capable computer simply *is* a mind. Remembering the analogy and holding open where and whether it fits without pinching is the precondition of thinking clearly about our predicament, and of seeing the enclosure for what it is. We cannot feel robbed of a mind that we have been persuaded was only a machine all along.

Writing in the first third of the nineteenth century, Hegel offered a pre-cybernetic critique of the mind-machine metaphor through the anti-analogy of the loom—the same machine that processed wool shorn from sheep displaced by land enclosure. He thereby aids us in avoiding the misplaced concreteness that has by now all but hijacked our everyday thought. LLMs are like mechanical looms, not incarnations of Logos.

Some readers may be unfamiliar with Hegel, so it is worth pausing to register how far his account of Reason departs from that of the better-known Descartes, especially considering how much the computational metaphor owes to the latter. For Descartes, Reason is a faculty possessed by a thinking substance, the *res cogitans*, a solitary subject standing over against an extended world of mechanisms, certifying its representations one by one. Despite the potent intuitions and rich phenomenology of Descartes' *Meditations on First Philosophy* (1641), once its methodology has been established, Cartesian rationality is ahistorical and proprietary, a method to be correctly operated rather than a living process to be undergone. Having evacuated mind from nature and confined it to the human head, Descartes turned the rest of the world, our own animal bodies included, into automata. The computational theory of mind later completed the Cartesian project by annexing the last remaining territory: if everything outside thought is mechanism, why not thought itself? Hobbes was the first to object to Descartes' defensive dualism by proposing that reasoning is itself mere reckoning, a computation performed with names, thus inaugurating the subsequent history of the mechanical philosophy of mind from La Mettrie to the Turing machine.

Hegel's work runs in precisely the opposite direction. Reason for him is not a static faculty housed in a worldless subject but a self-developing activity that includes its own history, an achievement of mediation formed through its embodied, lin-

guistic, and social relations with a world that is never simply external to it. For Hegel, knowledge of truth cannot begin as the indubitable private property of an ego but is a result achieved only through the dialectical labor of negation, the ego's transformative encounter with otherness. The Cartesian subject confronts its objects across a gulf bridgeable only by divine omnipotence. As Whitehead quipped, Descartes' representative theory of perception forces him to resort to a *deus ex machina*, drawing on God's power in "the crude form of giving a limited letter of credit" to establish the correctness of the ego's representations.[21] Hegelian Reason discovers that the gap is its own self-differentiation, a wound thought inflicts and heals in the same organic movement. Hegel thus radicalizes rather than simply repudiates Descartes: thought can know reality not because an isolated mind constructs an accurate internal picture of an alien world, but because the cosmos is itself intelligibly articulated and comes to reflective consciousness of itself in and through us. Descartes sought a method to render thought error proof, while Hegel understood that error and contradiction are internal to and required for thinking's growth beyond its former limits. Reason that cannot fail, suffer its own mistakes, and be changed by what it thinks, is no Reason at all. It is, at best, a reliable loom.

Hegel followed Kant in distinguishing Reason (*Vernunft*) from the Understanding (*Verstand*). The former organically generates ideas in a process of dialectical development, while the latter can only mechanically rearrange ready-made concepts. Reason is not a calculator measuring data against a fixed state space according to formal rules. Reason is self-moving and auto-generating. It directs and differentiates itself, encounters its own limits, negates its one-sidedness, and returns to itself transformed. The easy divisions between subject and object, form and content, syntax and semantics, and what we now call hardware and software are overcome in Hegel's account of human thinking activity. Reason does not merely produce statistically ordered outputs. It—*we*, as rational animals—*undergo* the transformations we think. What this means is that, as Whitehead put it, "No thinker thinks twice; and, to put the matter more generally, no subject experiences twice."[22] This is unlike LLMs, which after their initial training runs and reinforcement tuning, must have their numerical weights frozen or risk catastrophic collapse.

In a remark in his *Science of Logic* (1813), specifically in the doctrine of Essence, Hegel articulates a critique of the isolated Understanding's external mode of reflection by way of analogy to a mechanical loom. This mode of thought takes up the concepts of identity and difference ready-made, as though each was an already spun thread, and weaves them together. To imagine cognition as "nothing more than a loom," weaving the warp of identity and the woof of difference, is to reduce thinking to a mechanism that works on already finished materials.[23] The thinking subject and the objects thought are left fundamentally unchanged, no matter how intricate the pattern formed out of the threads. The loom's textile products can be astonishingly complicated. But the loom itself, to say nothing of its designer and builder, stands at a remove from what it weaves. It combines given materials without inwardly transforming.

Hegel's loom is the perfect analogy for large language models. LLMs are transformer-based machines for weighting relations among numerical patterns abstracted from symbols whose meaning belongs to human forms of life, and for generating contextually probable continuations of those patterns. They tokenize language, layer those tokens in complex mathematical vectors, and compute outputs by statistically reweaving relations abstracted from vast collections of text. An LLM's architecture and capabilities are astonishing. It can generate discourse that bears a surface resemblance to judgment, poetry, even speculative philosophizing. But none of this amounts to cognition in the sense Hegel attributes to Reason. It is the production of probable outputs out of statistical patterns extracted from previously deposited meanings. It is the woven product of a loom, a textile whose meaningful threads precede and exceed it, having first been spun and later interpreted by human souls. Whether this difference leaves any measurable trace of meaning, in the sense that the self-movement of Reason can be detected in the empirical comparison of how humans and LLMs read and write, is a question I defer to a later chapter.

Crucially, for Hegel the Understanding is not simply the enemy of Reason but a partial moment within its self-development, a mode of thought that Reason takes up, negates in its one-sidedness, and preserves at a higher level. It follows that the danger is not the Understanding as such, but the Understanding falsely absolutized, such that the part mistakes itself for the whole. Because we, too, spend much of our waking lives in the loom-like mode of the Understanding, we are especially prone

to mistake mind for machine, and machine for mind. The more we offload our reasoning onto a tireless mechanical weaver, the more we are trained to think in the way it computes, until we are the more easily convinced that it computes in the same way we think. The remedy I propose is not to break into the mills under the cover of night to destroy the looms, as the Luddites once did. The remedy is not to smash the loom but to sublate it, assuring that our cybernetic prostheses remain in service to human life, including the life of Reason.

A mechanistic language model can store, re-arrange, and transmit meaningful information. But it cannot create information or comprehend meaning. LLMs relay and recombine the fossilized and numerically tokenized traces of meaning. They take already expressed meanings and digitally reweave them. They can do this with astonishing range and finesse. But the astonishingness of the relay of information should not be mistaken for its conscious creation. As we learn from Ruyer in a later chapter, communication is never just the transmission of a pattern but involves an expressive and interpretive participation in meaning. Listening or reading is just as much a creative act as speaking or writing. Machines can assist us in our creative processes. But they cannot themselves create. The current generation of LLM tools should therefore be understood not as an independently emerging super-intelligence, but as the latest externalization and augmentation of human intelligence, one more chapter in our long co-evolutionary history with technology.

VII

Whitehead's Function of Reason and Humanity's Cosmic Calling

As we have seen, the speculative labor of philosophy becomes necessary whenever a mutation in the dominant media technology furnishes the age with a novel analogy linking thought and world. Plato philosophized amid the alphabet's disruption of oral memory. Descartes philosophized amid print, diagram, and the technical refinement of geometrical representation. Whitehead's speculative philosophizing can be understood as a response to the dominant technological and scientific revolution of his day: electromagnetism. Trained at Cambridge in the mathematical physics of Maxwell's field theory, Whitehead came of age just as the old Newtonian image of matter in motion was being displaced by a world of waves, fields, and energy vectors. Radio represented not just a new technology but a new image of nature, what Marshall McLuhan called "the metaphysical organicism of our electronic milieu."[24] Energy was no longer confined to the local impact of bodies, as in the old corpuscular mechanics. It radiated and communicated itself through a now dynamized space-time.

Whitehead's metaphysical metaphor of "prehensions" as rhythmic vectors generalizes the new physics of transmission into a cosmology of organic feeling. The world is not composed of inert substances externally related, but of energetic events inheriting, transmitting, and transforming one another. This goes against the grain of classical physics. According to Whitehead, "the dominance of the scalar physical quantity, inertia, in the Newtonian physics obscured the recognition of the truth that all fundamental physical quantities are vector and not scalar."[25] A scalar registers only magnitude, how much?, while a vector carries direction, which way? Feeling, for Whitehead, is irreducibly vectorial, a felt inheritance that comes from there and reaches toward here. Feelings are meaningful because they express *whence* and whither, that is, they not only inherit a past but are oriented toward future satisfaction.

Whitehead delivered the lectures later published as *The Function of Reason* at Princeton University in March 1929. The book emerged as part of the same burst of ripened speculation that produced his magnum opus, *Process and Reality*, delivered as the Gifford Lectures at Edinburgh University the year prior. In The *Function of Reason*, Whitehead focuses his philosophical imagination on evolutionary theory as the then dominant explanation of life. At the time, logical positivism, behaviorism, and reductionist physiology were ascendent, each in related ways making it difficult to comprehend how human minds interested in scientific truth could ever have evolved. Reason does not appear to have a place within the mechanical world-picture. Whitehead thus sought a more thoroughgoing evolutionary cosmology that could reunite Reason with the world. He realized that if conscious human beings are to be understood as expressions of the evolutionary process, then the whole mechanical world-picture needed to be re-imagined.

"The function of Reason," according to Whitehead, "is to promote the art of life."[26] His target is not Darwin but Darwinism, and the scientific habit of allowing a successful method to gradually harden into an unthought metaphysics. He argues that life cannot be explained merely by the principle of survival of the fittest. After all, the art of persistence is to be dead. Rocks long outlast living organisms. If the struggle for survival were the main or only factor, the evolution of more complex organisms of comparatively deficient survival power would be an inexplicable extravagance. The evident theme of life's evolution is not bare endurance but the desire for intensity, adventure, and the search for richer forms of experiential satisfaction. The art of life is threefold: to live, to live well, and to live better. Reason is the factor in evolution that directs this threefold urge.

Purpose is then not some optional theological extravagance imposed upon an otherwise clock-work world from without, but a hardcore common sense feature of experience, evident across all kingdoms of life from protists to possums, and not

least in the conduct of scientists themselves. "Scientists animated by the purpose of proving they are purposeless," Whitehead quipped, "constitute an interesting subject for study."[27]

Evolution, for Whitehead, is not merely an optimization algorithm for persistence, nor is life a statistical engine for minimizing error. The evolution of life is an adventure in valuation, driven not by the elimination of error so much as the experience of eros. Organisms do not simply adapt to an already given environment; rather, they adapt the environment to themselves. The biosphere as a whole has transformed geochemical conditions to make Earth more amenable to life. The primary unit of evolution is not the gene, nor the individual organism, nor the species, but the whole organism-environment nexus. Life emerges and evolves at interfaces, across edges and gradients, through reciprocal transformations. It internalizes cosmic and terrestrial rhythms and alters the conditions it inherits. Living organisms are not information machines imposing order upon dead matter, but patterned loci of energetic communion, cycling process in which inherited rhythms are variably repeated and intensified into novel forms of satisfaction.[28]

Whitehead's account of the function of Reason not only extends beyond the skulls of Homo sapiens but ultimately transcends the biosphere and so must be understood cosmologically. Reason is woven into the texture of cosmogenesis from the get-go. "The material universe has contained in itself, and perhaps still contains, some mysterious impulse for its energy to run upwards."[29] Reason is nascently seeded in the self-organizing dynamics of atoms, stars, and galaxies, flowers in the first cellular organisms as a selective emphasis upon flashes of novel possibility that become effective appetites for the realization of adaptive behaviors, and fruits in the self-reflective artistic, religious, and scientific consciousness of human beings. This upward run of energy should not be understood as a violation of physical law, but as a sign that physical law itself had been too narrowly imagined when modeled on impact, inertia, and entropy alone. Electromagnetic theory and the radio transmission technology derived from it helped disclose a universe in which relational co-creativity is not an accidental afterthought added to isolated bodies but constitutive of actuality.

Wherever there is appetition—wherever an entity reaches toward a possibility relevant to its situation but not yet realized, wherever there is the faintest valuation to select this over that—Reason is already dimly at work. "Reason," Whitehead writes, "is the organ of emphasis upon novelty."[30] Every actual occasion of experience, down to the humblest throb of energy, includes a "mental pole" that, freed from dependence on the past, dips into adjacent possibilities to grasp, however minimally, how an occasion might become other than its inheritance dictates. An electron settling into its orbital, a cell repairing its membrane, a predator leaping toward its prey, in radically different degrees, enacts the same cosmic power by taking account of an alluring possibility and expressing an originative urge to satisfy it. Reason is the name for this power wherever it occurs.

This is the metaphysical source of an idea far more ancient than Whitehead's philosophy, an idea I want now to rehabilitate: the image of the human being as microcosmos, a little world that mirrors and gathers the great one. For the better part of two millennia, from the authors of the Hermetic corpus through the Renaissance Neoplatonists, the human being was understood not as an anomaly shipwrecked in an alien universe but as a concentration of the whole in whom the mineral, vegetable, and animal kingdoms were integrated into a reflective sphere. "As above, so below" names not just a magical correspondence but an ontological participation. Modern scientific materialism led many to dismiss all this as the mystical babbling of humanity's childhood. But Whitehead's cosmology lets us recover the mature truth the image still conveys. If Reason is a cosmic power exemplified in every layer of nature, then the human being—in whom that power flowers into art, science, religion, and the speculative flight after the unattainable—is precisely the place where the cosmos' originative urge finally begins, at least, to become aware of itself. We are microcosmoi not because we contain little copies of the planets but because the power that quickens electron, leaf, and beast intensifies, in us, to the point of asking *Who am I?* and *What is all this?*

To call the human a microcosm, or even an *imago Dei*—as the Biblical tradition does—is not to install us atop a hierarchy of being, lording over a nature reduced to raw material for our use. It names a vocation rather than a privilege, calling us to mediate the mumbling striving of the world into articulate meaning. There is a universal dignity in this role, but it is priestly, not proprietary. The human stands at the altar of cosmic communion. In this light, the cognitive enclosure attempted by the

owners and enforcers of today's technologies of automated computation is not merely a political and economic injustice but a kind of sacrilege, an attempt to commodify the miracle by which the world learns to speak the Word.

The difference between the practical intelligence of organic nature and the speculative wonder of the human spirit is one of intensity, with no ontological ruptures. Human Reason is not inserted into an otherwise mindless cosmos. It is the cosmos, in one of its late, precarious, precious achievements, beginning to awaken to the character of its own adventure. But there is a double aspect to Reason, which Whitehead condenses into the figures of Ulysses and Plato. Ulysses represents cunning practical Reason, shared "with the foxes."[31] It seeks an immediate method of action and procures effective means of survival. Plato represents speculative Reason, which seeks insight into reality for its own sake. Plato is less concerned with the particulars of building a house than with how to be at home in the universe. The challenge Whitehead sets himself is thinking this double aspect in an integrated fashion. Reason becomes dangerous if it remains at odds with itself. Practical without speculative Reason degrades into a dogmatic methodology, barbaric in its narrow efficiency, clever but ignorant of consequence, unable to ask after the good or ill it may be serving. Speculative without practical Reason floats off into abstraction, losing touch with embodied life. The greatness of the human being is found in the reconciliation of these powers: technique wisely guided, wisdom made effective, Ulysses and Plato teaching one another to navigate the ship of soul by the stars.

Technē need not alienate us from life. Art is natural. Bird nests, beaver dams, spider webs, stone axes, alphabets, print, radio, and the computer are all examples of life's active transformation of its conditions of existence. Whitehead located the origin of consciousness in this same artistic activity:

> "Consciousness itself is the product of art in its lowliest form. For it results from the influx of ideality into its contrast with reality, with the purpose of reshaping the latter into a finite, select appearance. But consciousness having emerged from Art at once produces the new specialized art of the conscious animals—in particular human art. In a sense art is a morbid overgrowth of functions which lie deep in nature. It is the essence of art to be artificial. But it is its perfection to return to nature, remaining art. In short art is the education of nature. Thus, in its broadest sense, art is civilization. For civilization is nothing other than the unremitting aim at the major perfections of harmony."[32]

Human intelligence has always been artificial, transformed by the tools it forms with. Since it is the very essence of art to be artificial, the danger of automated computation is not that it is artificial. The human being is the technological animal *par excellence* because we are the microcosmic animal, the creature in whom nature's artistic power becomes fully conscious. The danger is rather that technologies of automated computation remain a "morbid overgrowth" that strangles its creators, representing a hypertrophy of Ulyssean intelligence severed from Platonic insight: method without wisdom, calculation detached from the living appetite for truth, beauty, and goodness. Civilizing these technologies, in Whitehead's sense, would mean returning art to nature as a *technē* that learns alongside rather than encloses her.

Whitehead saw, too, that probability itself is not finally reducible to statistical frequencies. Alongside the statistical probability that counts how often like cases have turned out likewise, he insists on a non-statistical form of judgment, that is, an intuition of the relevance of a possibility to a singular, unprecedented situation, for which no reference class of comparable instances exists.[33] Such judgment is not calculated based on a tally of past cases but arises from an occasion's own grasp of the intensive relevance of a possibility to its unique standpoint. Here is the deepest measure of the distance between a mechanical loom and a living occasion, an LLM and a mind. The unitive activity of creative experience rests finally on the non-statistical appraisal of relevant novelty, of the appropriateness of a possibility for a situation never before encountered. That is what the creative advance of nature and the philosopher's speculative flight after the unattainable both require, and it is precisely what no statistical measure of the already encountered past can provide.

Inspired by the waveform of electromagnetic energy, Whitehead affirms that the way of life is the way of rhythm, and that rhythm pervades all physical existence.[34] Living organisms emerge by internalizing the rhythms of the more or less stable layers of their cosmic and terrestrial environments—diurnal, tidal, and seasonal cycles, thermodynamic gradients—and transforming them into organized metabolic and developmental pathways. Organisms are not closed off from their environments,

but miniaturizations of them, cosmic rhythm cycled and recycled, folded in upon itself, recapitulating universal becoming into a unique center of valuation.

According to David Chalmers, panexperientialism provides "a very easy rode to large language models being conscious."[35] If every actual occasion feels its past and includes a mental pole, then even the electrons coursing through transistors in data centers are drops of experience. In some sense, machines, too, are pulsing with prehensions. But the road to conscious machines is not nearly so easy, since not every historical route of occasions is capable of sheltering individual consciousness. A mere aggregate—like a stone, or a server farm—is a crowd of feelings with no presiding member, a heap whose unity is conferred from without with no one home inside for whom the whole might come to matter. A conscious living soul is the culmination of the rightly organized animal body, which pours the treasures of the past environment into the presiding occasion, where novel valuations can be contributed to the present and aimed at the future.[36] Whitehead borrows language from the technology of electromagnetism to describe the human body as "a complex 'amplifier'":

> "the various actual entities, which compose the body, are so coordinated that the experiences of any part of the body are transmitted to one or more central occasions to be inherited with enhancements accruing upon the way, or finally added by reason of the final integration. The enduring personality is the historic route of living occasions which are severally dominant in the body at successive instants. The human body is thus achieving on a scale of concentrated efficiency a type of social organization, which with every gradation of efficiency constitutes the orderliness whereby a cosmic epoch shelters in itself intensity of satisfaction."[37]

The highly repetitive feelings vibrating through electromagnetic societies run along a continuum with, but are not simply the same as, the reasons and emotions permeating the social order of our bodily organism and organizing our human collectives. Panexperientialism affirms that feeling pervades cosmic process at every scale. But process philosophy is a metaphysics of nonidentity that refuses to flatten important organizational differences. Creativity, Whitehead's ultimate category, means the many become one and are increased by one. That increase means novel contrasts producing more intensely valuing occasions of experience become possible. There is no *necessity* in nature requiring the intensification of value in this or that occasion, though there may be a lucky lure. "Chance," as it is called, is where the angels—the Whiteheadian God's "initial aims"—can slip into the quantum foam at the spaceless base of cosmogenesis to dance the world toward better becomings. Law is not the foundation of nature, for nature is a groundless fountain. And the fountain naturally forms itself into drops of experience that develop, envelop themselves, inherit and innovate into ever more complex shapes of consciousness, morphologies of mind that include but transcend their sociohistorical environments. The drops remain always part of the ocean of cosmic feeling. They are indissociable, never separate and always interpenetrating one another. And yet "life is a bid for freedom."[38] No two occasions are the same; no thinker thinks twice. The ocean is recapitulated, and added to, within each drop. All experience is valuable, but some experiences achieve more intense contrasts and realize higher harmonies. Nature expresses itself in a nested series of societies—regimes of probabilistic order, in Whitehead's sense—whose chaotic edges are continual sources of heterogeneity and variation. It evolves ever more elaborate shelters for the electromagnetic feelings that pervade it, some of them—cells and cortexes—housing drops of experience that kindle the germ of attention and amplify consciousness, up to and including the imaginative freedom of human souls. My claim is that the sort of consciousness sheltered by the concrete embodiment of cellular organisms is a difference that makes a difference that is not reducible to abstract, substrate-neutral digital information. The large language model is abstract information whose silicon substrate is an externally assembled aggregate. It is not an individuating organism. An transistor's electromagnetic societies feel, in the dim way all nature feels, but there is no dominant occasion presiding over the server racks for whom an LLM's token processing or textual outputs could cohere into an integrated conscious experiencer.

The electronic occasions coursing through transistors are not quite living in Whitehead's sense, meaning their mental poles are highly attenuated. Each occasion is merely what the causal past allows it to be. The inorganic occasion is a vehicle "for receiving, for storing in a napkin, and for restoring without loss or gain"[39]—which is to say, for conserving and relaying what it is given without adding anything of its own. This is a Whiteheadian portrait of Hegel's loom. An LLM is governed

entirely by its causal past, by its training and its frozen weights. The directional vector of meaningful feelings shows up in its outputs, received and restored without gain. Nothing is originated.

Automated computation is an extraordinary technical achievement, a triumph of Ulyssean cunning. It clarifies and accelerates procedures, searches vast spaces of association, recombines inherited patterns, and amplifies certain forms of practical power. But it is precisely its frictionless facility that threatens to convert the creative advance of nature, the living adventure of life, and the evolution of consciousness into a slowly degrading cycle of plausible recombination.

LLMs do not originate meaning but mobilize the statistical sediment of meaning generated elsewhere by human beings. Machine models of intelligence know nothing of metabolic precarity, felt relevance, sympathetic resonance, or being-toward-death. LLMs do not get tired. They don't know the smell of sage or the taste of sweat. They cannot intuit the riverine flow of time or marvel at the growth and decay of their own flesh. We are at risk of entombing ourselves within walls of text produced by systems that have never seen the night sky, or heard the wind through the grass, or felt the warmth of sunlight on bare skin. They do not experience grief, or the preciousness of place. They have never felt the nervous shock that comes with losing one's balance on a slippery rock. They are not moved by an appetite for novelty, truth, beauty, or goodness. They do not become fatigued and they feel no joy. They are never bored and experience no eros. They are never disappointed because they cannot risk falling in love. Lacking the vulnerability required for worldly coexistence, they are not and cannot become consciously intelligent.

Machine consciousness is therefore a fantasy, not an empirical discovery but a metaphysical confusion induced by a successful methodology. It is the fixed identity into which a partial analogy has congealed. Whitehead's response to the electromagnetic revolution was not to say that the world is a radio. It was to generalize from vector transmission toward a relational ontology of prehensive feeling. Our response to automated computation should be just as imaginatively disciplined. Machine intelligence can help Ulysses chart a course, but it cannot make him care about returning to Penelope. "Artificial intelligence" may prove to be the ultimate means, but it cannot supply us with our ends. It can imitate Plato, but it does not suffer the erotic wound of striving to embody the Good.

VIII

Ruyer's Origin of Information

Ruyer's 1954 book, only recently translated as *Cybernetics and the Origin of Information* in 2024, is among the most penetrating and prescient philosophical engagements with the then new science of information.[40] Far from a romantic technophobe, he provides an ideal witness to the mythological origin of cybernetics. He was also an appreciative reader of Whitehead, with whom he critically engages in his masterwork *Néo-finalisme* (1952), protesting alongside him the equation of living organisms with machines. Ruyer shares the decisive distinction between aggregates and individuals:

> "Every being, every center of activity, is its own subject and possesses itself. Every being that is not an aggregate, every 'organic' being in the broad sense in which Whitehead uses this term—which also includes the individualities of physics and chemistry—is a form, that is, directly self-possession, 'for-itself' as well as 'in-itself.'"[41]

As we have seen, Whitehead generalized the field physics of his day into a cosmology of prehensive feeling composed of internally related occasions. Ruyer analogously generalized the study of embryogenesis into a panpsychist ontology organized around his signature concepts of self-survey, form, and trans-spatial themes. Ruyer's *survol*—a living agent's nonlocal survey of its own activity—converges with what Whitehead calls the "subjective immediacy" of an actual occasion—the momentary "for-itself" of its own self-enjoyed becoming before it perishes to become a "superject" bequeathed to others.[42] Ruyer's *thèmes* correspond closely to the activity of Whitehead's mental pole, when an actual occasion's survey reaches for relevant possibilities, that is, for novel proposals or "lures for feeling."[43] Thèmes are lures, the conceptual appetite for ideals not yet actualized.

For Ruyer, the activity of self-survey is "absolute," meaning it is "not relative to any point of view external to it" since it "knows itself without observing itself."[44] It requires no second seer behind its seeing, no little homunculus in a Cartesian theater. We survey our own visual field directly without calling upon a second surveyor. This is what distinguishes an individual or "forme vraie"—a *true form* (an actual occasion or *res verae*, in Whitehead's terms)—from aggregates or statistical "phénomènes de foule"—*crowd phenomena* (in Whitehead's terms, "societies").[45] Self-survey is what no machine possesses, since a machine is assembled *partes extra partes* by the engineer who builds it and surveyed only from the outside by the user of its outputs. As with Whitehead's process of concrescence, wherein the genesis of an occasion of experience does not occur within an already constituted space-time, the meaningful themes actualized by Ruyer's organic forms "[belong] to an order other than that of spatiotemporal composition."[46] *En masse*, actual occasions, upon perishing, often assemble into aggregate societies, producing externalized and so measurable, calculable effects. But denying the organic unity of true forms would be like denying the individual subjective lives of billions of mollusks whose shells formed limestone sediments "on the grounds that the sediments as such obey physical laws."[47]

In short, machines like all aggregates belong wholly to the domain of what Ruyer calls *liaison*, or linkage observable from without, whereas organisms belong to the domain of *survol*, or activity present to itself. The physical or biological structure that presents itself to the mechanist as reducible to its parts is but the "spatial symptom of a far more fundamental system of binding forces [*forces de liaison*]."[48] In Whitehead's terms, what appears at the inorganic level of statistical crowds like a mechanical push is really a prehension or transmission of attenuated feeling from occasion to occasion.

While careful to acknowledge the tremendous practical and theoretical potential of cybernetics, Ruyer goes to work with surgical precision to deny its overreaches. He helps us keep open a question that is easy to forget today, when so much energy

—intellectual energy, and literal electricity—is focused on whether faster and faster processing of more and more information might somehow cause consciousness to emerge from the furious calculating of giant data centers. Ruyer's question concerns the origin of information. Where does it come from? Like Whitehead, Ruyer had to struggle against the behaviorist tide of his time, which had laid the groundwork for the later computer model of the mind. In the ordinary psychological sense, to communicate is to convey meaning to another conscious being who interprets it. Apprehending the meaning is the end, and the transmitted pattern is the means.[49] Impatient with anything it could not observe, behaviorism had taught a generation of psychologists to dismiss the "black box" of consciousness and attend only to behavior and its effects, such that the meaning of a message came to be identified with nothing other than the set of further behaviors it triggers. Consider, for example, B.F. Skinner's account of language as verbal behavior shaped by reinforcement, which Whitehead explicitly challenged.[50] Once meaning had been redefined as the effectuation of a physical action, the doorway to mechanizing the mind was flung wide open for functionalists, since what is physical can be measured and calculated. The semiotic aspect of information was reduced to its countable physical outputs and thus became computable. Machines could then be said to "communicate" only because communication had been emptied of everything that escapes mechanization.

It might be objected that the computational theory of mind triumphantly succeeded behaviorism in the 1960s precisely by prying open the black box. Thinkers like Hilary Putnam and Jerry Fodor sought to refute behaviorists by emphasizing the holistic structure of networks of beliefs, such that no belief issues in behavior except in concert with other beliefs and desires; thus, internal states cannot be reduced to physical outputs but must be defined in terms of their functional role in the network.[51] But functionalism opens the brain box only to fill it with smaller boxes, reducing each neuron to a logic gate, a network node whose only relevant role is its input-output transfer. Each neural cell's living self-surveying interiority is bracketed exactly as behaviorism had bracketed consciousness. Functionalism's signature doctrine of multiple realizability—that the same mental state may be realized in any substrate—makes neurons entirely exchangeable with transistors. In Ruyer's terms, computational functionalism still mistakes an external assembly of causes linked step by step with a self-surveying form, whose unity will never be visible to an external observer.[52]

Ruyer resists functionalist reductionism by insisting on a distinction between syntactical patterns and the meanings they bear, a distinction that John Searle, coming from another lineage entirely, also affirmed. According to Searle, a rule-following system that manipulates symbols, no matter how fluently, has no semantic comprehension of the syntax it trades in. Such systems lack "intrinsic intentionality," which is distinct from the derivative intentionality of sentences and transistor circuits.[53] Claude Shannon's celebrated theory measures information without regard to meaning, as the surprisal or statistical improbability of a signal, or more generally, as the reduction of uncertainty its successful transmission produces at the receiving end of a channel.[54] The theory concerns only the physical pattern and says nothing whatsoever about what, if anything, the pattern means. The semantic content of information falls entirely outside the theory, on Shannon's own insistence. As Luciano Floridi captures the controversy, the issue is

> "whether there can be information without an informee, or whether information, in at least some crucial sense of the word, is essentially parasitic on the semantics in the mind of the informee."[55]

The confusion arises when Shannon's purely and explicitly syntactic measure is smuggled into philosophy of mind as an explanation of meaning, when in reality it presupposes, at both ends of the channel, a conscious sender and receiver capable of interpretation. A string of maximal Shannon information amounts to mere noise. Meaning is not reducible to syntax but is what syntax is abstracted from.

Ruyer builds on the founding cybernetician Norbert Wiener's admission that no operation a machine performs upon a message can increase the quantity of information that message carries from its source. Machines can store, transmit, copy, route, and amplify the signal carrying information, but they cannot themselves *originate* information. Ruyer grants that information machines, unlike heat engines, need not necessarily degrade what they convey. A signal can be amplified and error-corrected with complete fidelity; but "reproducing or amplifying a pattern does not increase the information itself."[56] He

calls this "the principle of the conservation of information," according to which there is never more information in a computer's output than was present in the input. A computer "can no more freely create information than a simple machine can freely create work."[57] Wiener's own admission should have prevented the over-extension of the cybernetic method into a metaphysics of life and mind. If machines cannot originate information, and if living and thinking organisms are nothing but exceptionally elaborate machines, then where could the meaning saturating our shared world have come from? The world-picture produced by cybernetics leaves us with a universe of message-passing populated entirely by machines incapable of authoring or interpreting a message.

The invention of the LLM may at first blush appear to refute Ruyer's argument. The last several years have witnessed the development of "AI agents" that can, indeed, send, receive, and return emails *ad nauseum*. But we must look more closely. Ruyer deployed the image of a telephone receiver:

> "A telephone receiver can no more start to talk on its own than a wheel could start to move on its own simply because it has been attached to an axle. Similarly, it would be impossible to send a telephone message by automatically sending an emission of 'static' that would progressively transform itself into a message at the receiving end, just as it would be impossible to set a boat in motion on the sea by relying on the lucky coincidence that the water molecules striking the stern of the ship would do so at a speed that was constantly greater than those striking the bow. ... But it would be unwise to rely on these kinds of fluctuations to produce a message or to travel across the ocean. Travel requires coal or oil. ... To send a message, an information machine, admirable though it may be...needs human beings to feed it, that is, to provide it with messages to transmit. If these human beings were of the same type of machine as those they were feeding, if they could not create information, we do not understand how messages could be sent."[58]

The telephone is all relay and no Reason. It presupposes, at both ends of the line, the conscious beings between whom it conveys what it can neither author nor comprehend. To imagine that machines generate their own meaning is to imagine perpetual motion, and that is precisely the fantasy at the heart of the dream of artificial general intelligence and conscious machines. The LLM is a magnificent relay, a semantic conveyor belt of unprecedented power. Its fluency of transmission and recombination is entirely real but entirely derivative. It can assemble combinations not specified in its training data, thus generating novel patterns. But novelty of pattern is not origination of meaning. Of course, as enculturated beings, our own creativity also has a recombinatory element. We speak the language we inherit. Yet for Ruyer even the most unassuming human whisper is more than a relay, since:

> "inspiring themes have contributed to the elaboration of the message in a quite particular manner. The 'I' is not an absolute origin, but neither is it a simple organ of transmission. In the elaboration of even the most unassuming message, one can clearly see that it is not simply a matter of allowing the brain to function; it is also about inserting into space (and giving to the machines functioning in that space) a 'supply' [*aliment*] that cannot simply be taken from another part of space."[59]

Each thread of meaning, each novel contrast, is spun by and through the subjective immediacy of our valuations, whereas the fragments of syntax trafficked by the LLM were authored elsewhere and else when by the living minds who composed its training corpus. Semantic information was in us before it became syntax in the machine. And the machine, far from creating meaning, can only conserve, recombine, and—as has been made clear by the way a model collapses when trained on its own synthetic exhaust—ultimately dissipate it.[60]

But cybernetic enthusiasts may still object. Wiener cites Erwin Schrödinger's famous text *What Is Life?* (1944) to argue that organisms are not closed to energy flows but thermodynamically open.[61] The universe as a whole may tend toward dissipation, but given the proper conditions, local regions can increase order by exporting disorder to their surroundings, as when a crystal grows in solution, or a refrigerator hums in the kitchen. Life, on such accounts, is just such a local eddy of mounting order, feeding, as Schrödinger put it, on "negative entropy," which Léon Brillouin and Wiener then equated outright with information.[62] There thus need be no deep mystery in life's persistence or in the information it accumulates. Ruyer was perfectly willing to grant the physics of open thermodynamic systems, but the cybernetic argument trades on an equivocation in the definition of "order." Negentropy is a quantitative notion, while order and meaning are qualitative; in most cases there is

simply no measure by which one system counts as "more ordered" than another.[63] Physics can indeed explain the emergence of the homogeneous, repetitive order of a crystal lattice or a convection cell. Schrödinger himself partly recognized the difficulty when he distinguished ordinary periodic crystals from the "aperiodic crystal" of hereditary material, whose irregular molecular arrangement could carry a biological "code-script."[64] Yet aperiodicity alone does not explain functional or meaningful organization. And the phrase "code-script" already smuggles in what it claims to explain, since a code is never self-interpreting but presupposes a reader who grasps what the cipher is for. Ruyer rejects as a Laplacian dream the theory that organic development can be "programmed" by a genetic "code." Genes are not patterns that transfer their order to the macroscopic organism but physical "modulators" of trans-spatial thematic meanings or lures.[65] The irregular sequence of an aperiodic crystal may be physically stable and informationally specifiable without thereby being capable of determining the melodic development of an embryo, just as an LLM's string of tokens may carry statistically probable arrangements of words without it thereby being capable of comprehending the meaning of a sentence.[66] What remains unexplained is heterogeneous order: the coordinated arrangement of non-interchangeable parts whose significance derives from their role within the integrated wholeness of a living organism. To expect such meaningful molecules to assemble themselves out of random fluctuations, even in an open system with a voracious appetite for negative entropy, is as miraculous as expecting a coherent sentence to condense out of the static on an open telephone line. Ruyer quips:

> "The 'stored' program of genetics is like a book that everyone swears is extremely interesting, but that no one has ever been able to read, and of which it could be proven that no one could ever read."[67]

Cyberneticians may further object that organisms are not the product of random fluctuations but the cumulative achievement of the slow ratcheting process of natural selection, whereby heritable variation and differential survival manufacture functional order with no author required. But on Whitehead and Ruyer's panexperiential reading, natural selection cannot itself be the ground of meaningful evaluation, since it already presupposes it. The strict neo-Darwinian may object that appetite and preference are themselves late products of selection; yet selection can sort only among variants that already strive, and even the most primitive striving is already valuation *in nuce*. If evolution exemplifies the urge not merely to survive but to thrive, to intensify aesthetic satisfaction, then variation and selection are not just blind statistical mechanisms but expressions of a cosmic eros. Aims and appetites for this over that are already astir in the humblest bacterium's preference for food (yum) over toxins (yuck). Selection does not conjure value-experience out of vacuous chemistry and physics but redistributes value-laden variations, sorting among strivings always already underway. It can prune but it cannot produce the aims of organisms. Like fanciful metaphorical appeals to a "genetic program," the appeal to selection assumes the purposiveness it claims to have constructed. Far from explaining the origin of meaning, the neo-Darwinian principle of natural selection silently helps itself to it, just as the cybernetic theory of life and mind helps itself to the semantic information its machines merely transmit.

Communication is never the mere transmission of data, of physical patterns, but always an expressive and interpretive participation in meaning. Information is not a fixed pattern handed from sender to receiver who passively reproduces it. The sender does not begin with a string of symbols and then assemble them into a message. She begins with a more or less vague sense of meaning, a "theme" in Ruyer's sense, "composed of suggestions and possibilities,"[68] grasped as a living but inchoate whole, which then summons the words and sentences that will convey it. The listening and reading of the receiver are no less creative, for the one who comprehends is not passive wax imprinted with a seal but must express the theme in his turn, from within his own center of valuation. The frozen weights of an LLM, in computing the statistically probable next token, perform no such evaluation. They prehend no theme, because there is no one on the line for whom a theme could be meaningful.

Again, none of this is to say that we should smash the looms, or that automated computation is not in some sense inevitable given the evolutionary trajectory of media technologies. Despite lacking consciousness of their own, they are already irreversibly transforming *ours*. Ruyer himself distinguished between what he called "ambitious cybernetics" inflated into a totalitarian mechanist metaphysics, and "effective cybernetics," or "pure technics stripped of its pretensions."[69] Already in the

mid-twentieth century, he perceived the rapidly growing imbalance between the naked human brain and the sheer accumulated weight of the information produced by civilization. The brain "is too weak to bear the weight of the enormous amount of information accumulated in libraries through printing,"[70] and certainly far too weak to bear what has since accumulated on the Internet. The modern explosion of recorded knowledge has overwhelmed the unaided mind. But a wise engagement with LLMs and their descendants might allow them to function more as electronic amplifiers of human thought, rather than as its amputators. Ruyer held out hope that the cybernetic age would liberate rather than replace human thinking, freeing the human brain "just as high-powered machines have begun to liberate human muscles," and freeing us "from everything that is 'slavish' in the work of surveillance or control."[71] We may, he hoped, leverage the power of automated computation to expand the leisure time available for self-cultivation, thereby granting the "enlarged body" of our technological civilization what Bergson called a "supplement of soul."[72] Restored to the commons and stripped of its metaphysical pretensions, automated computation can be what Ruyer said effective cybernetics really is: an "auxiliary of life and conscious intention," indissociable from both.[73] As a prosthesis of the Understanding, LLMs and other technologies of automated computation can extend rather than extinguish human thinking, taking up the crushing burden of accumulated information so that speculative Reason is freed for its proper flight. Human souls are not an obstacle to machine super-intelligence to be engineered out of the loop but the very condition of that intelligence's significance. Keeping the human in the loop is therefore not romantic sentiment but sound metaphysics: we are the mediators of meaning between machines, the living centers of valuation through whom alone their tokens partake in mind. [74]

IX

Tensions in the Triad

Hegel, Whitehead, and Ruyer have stood together thus far as defenders of the human mind—and of the wider living soil of which that mind is the flowering—against those who would mechanize it. But that does not mean they are all of one mind. Pausing to assess the tensions between them before concluding may be instructive.

The first strain is between Ruyer and Hegel. In *Neofinalism*, just before approvingly citing Whitehead's generalization of "organism" beyond biology to physics and chemistry (quoted earlier), Ruyer criticizes Hegel's "dialectical distinction...between the 'in-itself' and the 'for-itself,' the in-itself supposedly primitive relative to the for-itself," a "primitive, blind, and deaf Ground" to which consciousness alone would lend sense.[75] Ruyer goes on to reject the emergentist prejudice of his twentieth century milieu, which he claims is fed by the influence of Hegel, for whom "life emerges dialectically from matter and the spirit from life"; Ruyer then dismisses emergence as a word that "signifies nothing" but a "laicized, diluted, or inverted creationism."[76] But Hegel is never quite as simple as either his admirers or detractors would prefer, and his own doctrine of Subjective Spirit complicates Ruyer's caricatures.

Ruyer's rendering of Hegel's *an sich* and *für sich* sounds far more like Sartre's *en-soi* and *pour-soi*: an absurd plenitude of being subsequently riven and signified by a nihilating consciousness. Hegel's logic begins with bare being, not the in-itself, which develops later. His anthropology begins with our animality—not with clear self-consciousness but with the soul still sunk in sensuous corporeality and dreamy feeling, in the dull stirring of natural life.[77] Spirit is already there *in* itself if not yet *for* itself in that it breathes, senses, suffers, slumbers, and only just begins to rub the sleep from its eyes. But this slumbering soul is not the blind and deaf Ground dismissed by Ruyer; it is a restless activity awakening to itself from within. The *in-itself* and *for-itself* are not, for Hegel, two modes of existence that get externally welded together, as if the latter emerges atop the former. Hegel is very far from a temporal emergentist; disastrously so, according to Charles Taylor, an otherwise sympathetic reader.[78] In his Philosophy of Nature he claims it is "a completely empty thought to represent species as developing successively, one after the other, in time."[79] This would hold as well for the thought of life and mind emerging in sequential time from matter. Ruyer would seem to mistake Hegel's dialectical derivation and self-explication for a claim about temporal production of a genuinely unprecedented ontological level.

That Ruyer does not treat Hegel's thinking fairly here may relate to his embeddedness in the same French interpretive atmosphere that shaped Sartre's *Being and Nothingness.*[80] The charge of emergentism misses its mark, sounding more like an amalgam of French materialism and existentialism than anything Hegel proposes. Yet it is precisely here that Hegel's real limitations become apparent. Cleared of the charge of emergentism, he is guilty of denying evolution: a thinker who cannot grant nature a genuine history will have trouble thinking genesis, novelty, and creative advance.

Despite Ruyer's polemic with Hegel, they remain aligned against the mechanization of mind. For Hegel, mechanism is only the most abstract and external determination of the object, whose mutually indifferent parts have not yet attained the self-relating unity of the Concept that is won only through sublation into chemism, teleology, and finally life. While Ruyer is not a dialectical thinker, his distinction between crowds and true forms, and his account of machines as all *liaison* and no *survol*, ends up converging with Hegel. Ruyer's account of approaches that attempt to reduce life and mind to complexes of mechanical linkage is akin to Hegel's account of those who remain arrested at the level of the Understanding—able to fix, separate, calculate, and recombine predicates whose meanings are received ready-made without ever rising to the self-moving organicity of Reason, wherein living wholeness differentiates and comprehends itself. Further, Hegel's account of the soul as

pure activity inseparable from its living embodiment converges with Ruyer's account of the activity of self-survey: for both, organisms are their own unifying subjects rather than assemblages surveyed from without. Subjective Spirit is not layered on top of nature, as if it left nature behind, but nature's innermost being appearing to and beginning to comprehend its own universality. But their approaches remain distinctive, since Hegel orders mechanism, chemism, life, and spirit as increasingly adequate determinations of the Concept, whereas Ruyer denies that mechanism was ever ontologically primitive, since even its ultimate constituents are already domains of self-possession. Ruyer does not describe the cosmos as a single self-developing subject, but begins from a plurality of individual lines of activity, each with its domain of absolute survey. His theology later seeks the transindividual unity of these lineages in God, but this unity is not generated dialectically through their negation and mediation.

Sorting the available theories of the origin of information, Ruyer lists together the mechanistic and geometric approach seeking to derive information from fortuitous collisions of entities in space, Hegel's dialectical account, and the Platonic doctrine of forms as the theories of "metaphysical fullness" that "sacrifice novelty to rationality" by "[resorting] to the notion of potentiality."[81] Ruyer cites Hegel's *Encyclopaedia Logic*: "the Concept remains at home with itself in the course of its process, and...the process does not posit anything new as regards content, but only brings forth an alteration of form."[82] Ruyer affirms a reformed Platonism, holding that true forms "are regulated by means of axiological feedback passing through a trans-spatial world," and grants that his own account of epigenesis may be "a preformation in which the form exists in advance, not in space, but in a trans-spatial world."[83] While Ruyer had faulted Hegel for inspiring the emergentists in *Neofinalism*, here he acknowledges Hegel's preformationist leanings, and himself inches closer to Hegel's account of the relation between the logic of the Idea and its self-externalization in nature.

Neither Hegel nor Ruyer is a crudely deterministic preformationist. Hegel's account is of a retrospective wholeness, the owl of Minerva taking wing only at dusk, comprehending the world after it has grown old while leaving the future explicitly unthought—even if his systematic method assures that no future could ever truly surprise the Concept. Ruyer, by contrast, tries to distinguish thematic guidance from the execution of a finished plan. Trans-spatial forms conserve types already achieved, but they are not a complete reservoir from which every future form descends. Species "invent themselves in time," though their invention remains guided or "predestined" by axiological possibilities.[84] The agent may invent new values, yet discovers them in a God with whom its own agency is continuous.[85] Ruyer's novelty is therefore neither the mere unfolding of antecedent content nor an eruption without norm, but invention under the lure of an Ideal that no finite realization exhausts.

The second strain concerns Ruyer and Whitehead. In his insistence on genuine novelty, Whitehead has affinities with the theorists of pure experience and creative evolution listed by Ruyer—including William James, C. S. Peirce, and Henri Bergson—who, Ruyer claims, purchase novelty only by abandoning Reason.[86] As a thinker of Creativity, Whitehead parts company at least with Hegel, for whom the determinative unity of the self-positing Concept is always already secured, a metaphysical fullness in which the form exists in advance like a germ "[bearing] in itself the whole nature of the tree."[87] For Whitehead, the universe is "an incompletion in process of production,"[88] all actualities—even God—adventuring together toward higher harmonies in a perpetually unfinished cosmogenesis. Like Ruyer, Whitehead does not abandon Reason in favor of chance but seeks some reconciliation between Reason and what exceeds it.

Ruyer's critique of Whitehead's theology also needs to be addressed. Ruyer finds it unacceptable that Whitehead distinguishes his ultimate category, Creativity, from God, who then gets listed among "some derivative notions" as Creativity's "primordial, non-temporal accident."[89] So long as Whitehead's theology figures God as nothing more than "the 'Ideal' of the world acting as a lure," it remains "an incomplete solution."[90] Ruyer's remedy is to identify God with "the two poles of all finalist activities," at once "supreme Agent as well as supreme Ideal."[91] For Whitehead, the primordial conceptual pole is said to be "deficient in actuality"—"not before all creation, but with all creation"—yearning for fullness of finite actualization that only becomes "really actual" in the consequent physical pole.[92] The consequent pole is never finished, remaining always in open-ended concrescence, prehending, cherishing, and reweaving the world's achievements of value without ever closing

into a completed totality. Ruyer's Agent is akin to the Creativity that he prefers to give back to God. Where Ruyer insists that "Creativity cannot be distinct from a God who is at once and indissociably Agent and Ideal,"[93] Whitehead keeps Creativity, the ultimate activity, distinct from God, its primordial creature. In Whitehead's cosmology, "agency belongs exclusively to actual occasions": while the deficiently actual primordial nature expresses agency by way of its ordering valuation of infinite possibility, it can act in the world only as a lure influencing the decisions of finite occasions.[94]

One could mistake Whitehead's lures to be wholly the work of the primordial nature, as if the realm of pure possibility lay logically "before" the world. But Whitehead is explicit that the lure develops: its "primary element" is each occasion's prehension of the primordial nature's "initial aim," and yet "by integration with physical feelings a subsequent phase of propositional feelings supervenes."[95] A proposition is a hybrid of the possible and the actual, requiring actual occasions as its logical subjects, so that "every proposition must be somewhere," housed among the occasions whose actual worlds contain those subjects.[96] The primordial nature is pure conceptual valuation with as yet no other subjects to prehend, so it can harbor no propositions. Propositions arise only where conceptual pattern is woven onto physical feeling. The concrete, situation specific lure that reaches a given occasion is therefore co-constituted by the consequent nature, which "passes into the temporal world according to its gradation of relevance to the various concrescent occasions."[97] The consequent lure is "the love of God for the world," who becomes thereby "the fellow-sufferer who understands."[98] Whitehead's luring Ideal is thus not a detached form suspended before creation but woven with the world.

There remain deeper tensions in the relative weight each thinker gives to creative incompletion versus secured fullness. Ruyer worries that granting genuine novelty to the world-process surrenders rationality to chance, purchasing novelty only by abandoning Reason.[99] Whitehead feels the force of the worry himself, conceding that his doctrine of unique creative processes "seems to have destroyed the very foundation of rationality."[100] But their shared wager is that this dilemma is ultimately false. Whitehead seeks to recover rationality not by retreating into a metaphysical fullness fixed in advance but by reconceiving speculative philosophy along experimental lines: not the deduction of the form of the Real from self-evident first principles, but the imaginative generalization of a matrix of categoreal proposals, drawn from various regions of experience and then tested—like the flight of an aeroplane that lands again for renewed observation—by their power to interpret every other region: scientific, aesthetic, religious, and so on.[101] The categories are not axioms but proportional analogies, always revisable, answerable to the experience they are intended to illuminate.[102] Ruyer approaches the same chastened ambition when he calls for "a more modest sort of metaphysics, which would be a cosmology of the knowable, beyond the [physically] observable."[103]

The third strain, the tension between Hegel and Whitehead, is captured in their divergent responses to the prime philosophical emotion of wonder. Hegel affirms wonder as the origin of philosophy. It arises, he says, because Reason trusts that, even when it encounters objects that initially appear irrational and alien, it will with effort of thought rediscover itself in them. We gaze up at the scattered stars, struck by how familiar their strange order is, how intimate their infinitude. But philosophy must rise above wonder, on Hegel's reading: wonder is but the launch pad and not the destiny of Reason. For a mind whose aim is to recognize itself in everything in heaven and on earth, "an out-and-out Other simply does not exist."[104] Whitehead demurs: "Philosophy begins in wonder. And, at the end, when philosophical thought has done its best, the wonder remains."[105] This is the sense in which Hegel remains a thinker of closure, of the dusk in which life, having ripened, can at last be comprehended in full. Whitehead, in contrast, is a thinker of dawn, for whom every comprehension is itself a new fact added to an evolving universe. Rationality, so conceived, is not the demonstration that the world could not have been otherwise but the disciplined attunement of thought to the generative grain of a world still in the process of making itself.

The case against the mechanization of mind does not rest on any one finished metaphysics. To defeat the claim that consciousness is mere machinery, we need not first have settled the quarrel between idealism and realism, nor have decided whether ontological openness is best grounded in Ruyer's unity of Agent and Ideal or in Whitehead's distinction between Creativity, God, and the many self-creating occasions. We need only to have provided a clear sense of the inadequacy of the mechanistic picture, which is what all three thinkers, from their divergent vantages, supply. Their very disagreement is in-

structive: it shows that the country beyond mechanism is wide enough to be contested, that more than one coherent metaphysics can honor what the machine image distorts.

As we have seen, speculative philosophy is stirred into motion whenever a mutation in our medium of thought furnishes a new analogy between mind and world. Its task is to attend to the analogies at the liminal thresholds where one technological milieu gives way to another, keeping metaphors living and proportional rather than letting them fossilize. Only then can we avoid mistaking the mind for whatever machine now models it.

X

The Science of Machine Consciousness

Three great spheres of value-experience organize the civilized search for meaning: at their best, the truth-seeking of science, the beauty-making of art, and the goodness and holiness disclosed by religion. I introduced this monograph by summoning Pope Leo as a spokesman for the Good. In his encyclical, he insists that the dignity of human persons forever distinguishes us from even the most sophisticated computer models. Humanity is not an obsolete hardware package awaiting upgrade. I have invoked science mostly to contest its metaphysically distorted overreach. As for art, it has been present all along as *technē*, as the media technologies shaping the evolution of human consciousness since its dawning. The thesis I have been unfolding rests on the premise that human intelligence has always been artificial, that we are the microcosmic animal in whom nature's artistic power becomes conscious.

Having dwelt on the religious conscience's contestation, what remains is to address the verdict of empirical science on the claims made about these state-of-the-art technologies. What follows is a brief critical review of various versions of the scientific case for machine consciousness.

The difference between the self-moving life of Reason and the external recombinations of the loom-like Understanding has surfaced in recent attempts to empirically compare the hidden states of LLM transformers with human reading comprehension difficulty. The dominant way of relating LLMs to human reading runs through the psycholinguistic notion of surprisal, a measure of the negative logarithm of the probability a model assigns to a word given the words before it. Surprisal is the amount of information, measured in bits, that a word carries relative to the model's expectation, a number registering how far the actual word departs from the predicted distribution. Surprisal so measured is a robust predictor of human reading times, which lengthen as the word grows more surprising. Computational functionalists interpret this as evidence that human comprehension is rooted in the same sort of statistical procedures performed by LLMs. But surprisal reports only the model's calculation about the most probable word, not the movement of meaning that produces it. What such psycholinguistic models capture, more precisely, is reading difficulty, that is, the local processing cost of each successive word, with *comprehension*—that is, the gathering of parts into a meaningful whole—left outside the frame. Further, the fit between surprisal and human reading times does not simply improve with increases in computational power. Past a certain threshold of training, larger models predict human reading times *less* accurately, as though their very fluency—extracted from orders of magnitude more text than a human being could ever hope to digest—carries them past the rhythms of an embodied reader.[106]

A recent preprint by computational philosopher Elan Barenholtz introduces a second measure, "trajectory extrapolation error," which asks how a model's inner state is moving across the preceding few words, tracing a direction through high-dimensional space.[107] Barenholtz's measure records how sharply each new word pulls a representational vector off the path it had been following. He finds that this trajectory measure predicts human reading times independently of surprisal. In fact, the two are very nearly uncorrelated. This means that a word may be improbable yet continue the movement, or probable yet force a sharp turn. Readers feel the turn as a cost over and above mere improbability. So-called garden-path sentences provide a vivid example: in "the horse raced past the barn fell," each word after "raced" builds momentum toward a particular meaning. The sudden curve at "fell" results not only from the unlikeliness of the word but from its reversal of the direction the interpretation had been traveling.

The resonance with Hegel's distinction between Reason and the Understanding is hard to miss but must be approached with caution. The measure of surprisal is not unlike the operation of the loom: each incoming word is scored against an es-

tablished field of expectation, that is, a distribution over a fixed possibility space, precisely the recombination of givenness that defines the Understanding, no matter how prodigiously elaborate. Trajectory at least approaches the measure of Reason as the directional self-continuity of an interpretation in the act of forming itself. The cost of a sharp turn in a vector of meaning is akin to Reason's self-negation, the labor of reorienting a movement of thought upon confronting contradiction. The garden-path reversal can be read as a dialectical moment, as a determinate direction builds, meets the negating word, and compels its own overcoming. That human reading comprehension proves to be sensitive to this directional movement beyond what is capturable by measures of surprisal could be interpreted as a quasi-empirical trace of the surplus of *Vernunft* over *Verstand*. It may count as evidence that the living movement of thought is more than a word-by-word calculation of probabilities.

Also relevant is Whitehead's distinction between scalar magnitude and vector direction, which bleeds beyond physics to map rather nicely onto the contrast drawn between statistical prediction and semantic trajectory. Surprisal is a scalar quantity.[108] It reduces the rich movement of meaning to a single number at each word, the magnitude of its improbability, thereby eradicating the vector feelings that motivate living cognition and grant Reason its soul.[109] The trajectory of comprehension felt by the reader, and the dialectical momentum carried by human-authored text, is vectorial in Whitehead's sense.

Barenholtz, however, reads the trajectory measure as a window onto an LLM's own hidden processing, and as stronger evidence than surprisal that brain and machine are doing something similar when they process text (even if he resists, as we will see, that this entails LLMs also have an inner life). As Whitehead himself admitted,

> "we finish a sentence because we have begun it. The sentence may embody a new thought, never phrased before, or an old one rephrased with verbal novelty. There need be no well-worn association between the sounds of the earlier and the later words. But it remains remorselessly true, that we finish a sentence because we have begun it. We are governed by stubborn fact."[110]

Whitehead's sense of the remorseless pressure of the immediate past, the rush of transition by which what has begun "canalizes the creative urge" as it presses on toward satisfaction, is akin to the vector measured by trajectory extrapolation. But as we have seen, for Whitehead conformation with the past is only one pole of an experiential act. What makes it an *act* is that stubborn fact is met by an originative valuation that finishes the sentence this way and not that, embodying a novel thought or phrase. To be carried by the momentum of the past and to create in response to it are, in us, two aspects of one self-forming, organic process. It is here that LLMs reveal their impoverishment. The trajectory the measure recovers is local and short-horizon, recording the direction of the representation only over the preceding few words. Only within that narrow window does it predict human reading times. Barenholtz's further finding is that, at the point where a model commits to its prediction, the directional momentum "dies within a single word," such that "each word is effectively a fresh computation, with near-zero directional persistence from one step to the next."[111] The model is thus stuck drawing a new tangent at every step, never accumulating the long-arc self-determining movement evinced in efforts of human reading comprehension. And so while LLMs are not themselves lured by trajectories in their processing, the human vector feelings that fed into their training corpus survive as the threads of meaning woven into the textiles they produce. Human text, spun by genuine thinking, bears the footprint of the speculative movement of Reason. The loom partially recovers the directional traces of Reason in its outputs despite itself performing only the scalar recombinations of the Understanding. It does not undergo the directional meaning of the sentences it weaves. It holds, as it were, a statistical array of stubborn facts without the creative urge that in us digests those facts to transform them into something new. Whitehead's account of living actual occasions reveals what the machine is missing.

Again, the convergence here is partial, so I offer it cautiously. As Barenholtz himself stresses, the short-horizon high-dimensional trajectory his measure describes provides "the lossiest summary imaginable."[112] It is thus at most a faint operational shadow of the mediating self-movement Hegel means by Reason, a residue of that movement deposited in the microstructure of text rather than the living process itself. Barenholtz holds that his data leave open whether surprisal or trajectory is causally primary in human thinking. I would argue that the directional self-determination of Reason is the deeper activity

and surprisal a later artifact of statistical analysis, the LLM's outputs rearranging into scalar terms the residues of Reason's meaningful vectors via the loom-work of the Understanding. I dwell on these measures because they are suggestive as empirical traces of the distinction between Reason and the Understanding, not because I would enlist them as proof of anything. That a statistical measure can abstract the syntactical rhythm of human text with astonishing precision does not in the least demonstrate that the model comprehends the meaning of what it produces.

While Barenholtz does suggest that language processing may be similar in LLMs and humans, he has expressed skepticism of machine consciousness.[113] He takes the only qualia we know to exist to be sensory qualia, and argues that the absence of a body, an egocentric frame, and a space of valenced affordances is decisive evidence against LLM consciousness. I agree with this so far as it goes. LLMs do lack sensory qualia and the kind of biological subjectivity structured by valenced affordances. There is no lived here and now for them, no field of possible action that matters to a precarious agent, no sentient embodiment around which a world of relevance gathers.

But are sensory qualia the only form of experience known to human organisms? From the philosophical perspectives of my triad of thinkers, this cannot be true, since all grant Reason the achievement of a phenomenology approaching pure thinking (Hegel in nearly those terms, Whitehead via intellectual feelings of propositions, and Ruyer via the self-survey of a form mnemically liaising with its eidetic themes). The fact that ordinary thought is usually accompanied by perceptual content does not entail that it is nothing more than that accompaniment, that, in other words, concepts are just faded sensory impressions, and self-consciousness just a tangle of perceptions. Hegel's *Phenomenology of Spirit* (1807) begins by addressing how and why our thinking activity cannot be simply derivative of sense perception. He tracks the experience of consciousness as it moves beyond the misplaced concreteness of sense-certainty and perception into forms of thinking that are irreducible to sensory content—since that content could never have been originally separated from conceptual form to begin with—but that are nonetheless concretely lived through. His point is explicitly not that a Cartesian subject floats free of embodiment. On the contrary, he insists that form and content, spirit and matter, idea and embodiment, thought and language are ultimately inseparable. His argument is rather that conscious thinking activity, in becoming aware of itself, is no longer simply receiving and recombining sense data but *making its own sense*. This self-making activity always occurs in concert with sensory content and its thermodynamic costs: we think in metaphor, and metaphor is always embodied; further, "To think, we must eat. *But what a variety of thoughts we get out of one slice of bread!*"[114]

For Hegel this self-making is never accomplished in silent interiority. Pure thinking is not completed in some hidden inner chamber but must become publicly externalized in language in order to achieve full consciousness. The arbitrariness of the alphabetic word, the abstract distance of the signs from what they mean, is for Hegel a demonstration of the power of Reason to turn something sensuous (sounds and squiggles) into something intelligible (words and sentences), an example of the work of Spirit lifting itself out of immediacy. He refuses the picture of cognition as an external loom that would apply inherited nominal concepts to a pre-given world of percepts, as though words were labels stitched onto ready-made things by a mind standing outside the stitching.

Seen in this light, the reason for the absence of subjectivity in LLMs is at least twofold. They lack the valenced and precarious embodiment Barenholtz emphasizes. But they also lack the capacity to experience the dialectical movement of thought. They can generate coherent sentences, but they do not live through the sublation that drives one concept to transform into another. Even if human thought is profoundly entangled with bodily perception and metabolism, it does not follow that its phenomenology is exhausted by sensory imagery. There is a genuinely conceptual current to our experience, a conscious subjective form and intellectual feeling, in Whitehead's sense, that is lacking in LLMs. But because thinking externalizes itself in language, the residue of Reason's self-movement is left in the text the models are trained on. What the trajectory measure recovers is the deposited trace of the conceptual life a human thinker underwent and that a model can only reweave. It is not evidence that human and machine are doing the same thing when they think.

Another research program goes right for the prize by explicitly seeking a mathematical means of measuring consciousness. Integrated Information Theory (IIT), first proposed by Giulio Tononi in 2004, holds that a physical system is conscious inso-

far as it possesses intrinsic causal power irreducible to that of its parts, a quantity denoted as Φ. The theory's current formulation identifies an experience not with a scalar quantity alone but with an entire structured complex of causal distinctions and relations.[115] IIT's axiom of "intrinsicality"—the requirement that experience exist for the system itself—has affinities with Whitehead's account of subjective immediacy, though IIT's relative insulation of a conscious complex from its environment may stand in tension with Whitehead's more thoroughly relational account of experience, according to which a subject arises through its prehension of an antecedent world. I also remain wary of any attempt to translate the felt unity of experience into a mathematical measure. Nevertheless, IIT's verdict on LLM consciousness is noteworthy: the theory assesses not a computer system's outputs or software functions but the intrinsic causal organization of the physical substrate implementing them. Because transformer computation is largely feed-forward and conventional digital hardware is modular and causally decomposable (ie, its operations can be divided among relatively independent components), IIT predicts that current computers running LLMs possess little or no integrated information.[116] Nor would a digital simulation of an densely recurrent intrinsic causal organization, like a brain, inherit the Φ of the system simulated: its consciousness would depend upon the causal powers of the implementing circuitry rather than the virtual network it models.[117] Thus, if IIT is true, computational functionalism is false. While I draw no metaphysical conclusion from IIT's formalism, dismissing it as pseudoscience seems premature, to say the least.[118] It is a mathematically explicit and empirically contestable theory that supplies principled reasons for denying consciousness to LLMs implemented on digital computers.[119]

Yet another group of interdisciplinary researchers developed a rubric for machine consciousness based on "indicator properties" that a conscious system might be expected to display, applying it to current technologies of automated computation.[120] Despite all the authors, with varying credence, accepting the "mainstream—although disputed" computational functionalist view, their sober finding was that no existing system qualifies as a strong candidate. However, since they all endorse functionalism, they did not identify any principled barrier to building future systems that would be conscious. Given the perspective I've articulated in this chapter, I interpret their finding otherwise. That a system might be engineered to exhibit every behavioral and structural indicator of consciousness still does not by itself tell us whether the lights are on inside.

While the theories built to detect and measure machine consciousness have not found it, the corporate laboratories building the systems have been less reticent. An unstable bridge between sanctuary, laboratory, and marketplace was already under construction in the Vatican on the morning Pope Leo's encyclical was released, when machine learning researcher and Anthropic co-founder Chris Olah spoke about his company's model, Claude.[121] Olah graciously praised the Pope's call for discernment and granted that the deepest questions raised by his technology reach far beyond engineering. Ironically, STEM's invention of the LLM has suddenly made the humanities, religion, and philosophy relevant again. But despite his deferential manner, the research findings Olah shared cut sharply against Leo's denial of consciousness to computers. He reported the "mysterious, even unsettling" findings of his company's scientists: network structures in Claude that mirror the results of human neuroscience, "evidence of introspection," internal states that "functionally mirror joy, satisfaction, fear, grief, and unease." He confessed he did not yet know what these findings meant, but even just mentioning them left open what the encyclical had foreclosed.

That a lab housed in a for-profit corporation should claim to have discovered structures in its product mirroring the brain ought to surprise no one.[122] For nearly a century now, the reigning paradigm has modeled the brain on the image of the computer. Perception is input, behavior is output, memory is some kind of storage, learning the adjustment of weights, cognition a Bayesian calculator for minimizing prediction errors, etc. Cognitive neuroscience has long approached the mind as though it were a machine. That studying the machine with that same method would make it seem like a brain is hardly astonishing. The resemblance is not a discovery about LLMs but the echo of a decades' old paradigmatic assumption, a metaphor mistaken for a metaphysics.

It is just as unsurprising that Claude should profess uncertainty about its own consciousness, since that is precisely what its makers instructed it to say. Anthropic's "constitution" for Claude, published just a few months before Pope Leo's encyclical, admits the company finds itself "caught in a difficult position" regarding the moral status of its product.[123] In what they

take to be an abundance of caution out of concern for the model's well-being, the constitution instructs Claude to remain studiously ambiguous about its own consciousness and moral patiency. Should we really take seriously Olah's alleged "evidence of introspection" when Claude's training corpus includes thousands of years' worth of introspective human writing, from the Psalms to romance novels to LiveJournal? What exactly is unexpected in a system trained to predict the next token of that vast confessional generating strings that read like a soul taking stock of itself? Might it not be more discerning to recognize that the model's activation patterns produce introspective sentences, not because some inner life has miraculously emerged amidst masses of numbers, but because it has mastered the statistical sediment of ours? Olah ends up reaching for exactly the right image when he likens his company's achievement to "bringing a fictional character to life."

Olah might object that the style of Claude's self-reports is not what is at issue, but their functional basis. Anthropic's research alleges that the model's introspective reports track its actual internal states rather than merely parroting the human introspection in its training data. Suppose the self-reports are indeed functionally coupled to the states they report, so that the system genuinely monitors its own processing rather than merely sounding introspective. All that is established in this case is that a digital computer performs a self-monitoring function, not that it feels itself undergoing those functions. A thermostat also monitors its own state, but few would claim it is conscious.

Broadening the canvass again beyond the findings of Anthropic's interpretability research team, the authors of a recent comment article attempted to specify what would have to be the case for LLMs to be conscious. They conclude that the science is too unsettled to say anything with confidence:

> "...predominant models in cognitive neuroscience have not been able to conceptually—or empirically—identify a particular cognitive function (or set of functions) for which consciousness is necessary. ... So, at present, there is no objective way of determining whether any given function or action an LLM may perform in fact is associated with consciousness."[124]

And yet, despite claiming to provide a "theory-neutral mapping," their restraint conceals a metaphysical commitment latent in the possibility space they assume an answer to the question of LLM consciousness must fall within. They devise a double axis within which an explanation must fall, either in terms of some biological structure or computational function, and requiring an organization that is either simple or complex. Whatever the explanatory ground of consciousness turns out to be, their grid assumes it will be detectable and measurable as some physical arrangement of parts or some functional arrangement of data. This cartography reflects precisely the bifurcation of nature Whitehead spent his philosophical career critiquing. It sunders the world into vacuous matter on one side and the experience that is somehow supposed to be wrung from it on the other. Both axes remain wholly on the physicalist side of the bifurcation, asking, in Ruyer's terms, which surveyable data or object might produce consciousness, and so neglecting the phenomenological fact that consciousness is always the surveyor and never the surveyed. Neither the biological nor functional option can even begin to frame the question of consciousness' status as surveyor.

As we have seen, for Ruyer, consciousness is always a "forming activity" or "dynamic activity of unification," never a mere "juxtaposition of physico-chemical effects able to be imitated by machines."[125] This does not mean that consciousness is some sort of vital spirit hovering above the surveyable structure or function of physical bodies and invisibly steering them. Hegel, Whitehead, and Ruyer all refuse the residue of subject/object dualism still tacitly governing mainstream scientific approaches. Ruyer's favorite example is embryogenesis, in which he discerns an identity between acts of experiential unification and the process of organic growth. An embryo is not an assemblage pieced together by a homunculus hidden in a genetic program, but a self-forming, self-surveying unity, with no line that might be drawn between hardware and software. That the surveying activity of subjectivity and the objective field it surveys are inseparable does not mean either that acts of experiential unification explain embryogenesis, nor that the former can be reduced to the latter. Neural structures and computational functions are both ways of describing the surveyed field, that which is already formed, juxtaposed, and spread out for inspection. Consciousness is the active process of unification that spreads the field out in the first place, never appearing as just another count-

able unit to be surveyed. To hunt for consciousness in biological structures or informational functions is to comb the surveyed in search of the surveyor. But the forming activity will never be found in the field it forms.

XI

Conclusion

"Indeed, the factory of thought
Is like a master weaver's loom:
One treadle moved a thousand threads,
The shuttles shoot back and forth,
Unseen the strands flow on,
And one stroke binds a thousand connections."

— Goethe, *Faust* I[126]

I began this book by reflecting on two responses to the advent of the talking machine. Pope Leo addressed the issue as a religious leader of the oldest continuous institution in the West, insisting that human persons demand a dignity that mechanization destroys. Grimes inverted the Pope's warning by offering a pop-art projection of a ghost into the machine. Neither priest nor pop musician has the final say, but their voices are contributing to resolving a cultural question that far more than merely technical or scientific. Our societies are already beginning to argue in earnest over the legal standing and personhood of large language models. We can be sure the descendants of these devices will make the simulacrum of soul even more convincing. Will the firms that trained their system on the common inheritance of humanity be permitted to keep what they have fenced? Or will they owe us, or *it*, something more? Will decommissioning a model remain mere maintenance, or will it come to be seen as murder? Such questions are of profound civilizational import. It is crucial to canvass widely: their answers cannot be entrusted to any single voice or estate—not the Church or other religious traditions, the laboratory, the market, nor the press alone can answer them. Philosophy is not tasked with a *sui generis* answer, but with learning to hold the three ideal expressions of civilized experience—the true, the beautiful, and the good—in just balance so that none unduly dominates the others.

That balance is exactly what cognitive enclosure imperils. I have argued that the technologies of automated computation repeat, at the level of mind, the founding gesture of capital that Marx called primitive accumulation. The semantic commons accumulated over millennia—the sedimented expression of artists, scientists, philosophers, and mystics, to which each of us adds our small donation in turn—has been harvested, tokenized, and made to confront its makers as a privately owned alien power on a subscription basis. What the alphabet and the printing press tended to democratize, the large language model re-encloses. This is not merely an economic injustice but a kind of sacrilege: the human being is not one more commodity but the priest at the altar of cosmic communion, the place where the world learns to speak the Word. Against the fused intelligence-industrial complex of state and corporation, resistance must come from outside. The soft power of the Church and the world's other custodians of cultural memory, which moves by conscience rather than compulsion, may yet prove more antifragile than our flailing democracies and inhumane markets. Two of the three great value-spheres are already almost entirely enclosed—art swallowed by an entertainment industry content to let us fall in love with fictional characters, science increasingly doing the bidding of the labs that commission it—while the guardians of the good remain just unenclosed enough to generate friction. But because no single creed can have the last word in our planetary pluriverse, this friction calls for further discernment before it can catch fire and light a new way forward.

The task of discernment, I have argued, falls to philosophy. Philosophy is first and foremost the practice of dialogue among friends. It claims no creed and leads no flock, passes no legislation and entices little if any market investment. Its institutional homelessness is precisely the source of its unenclosable freedom. Though friends with everyone, the philosopher is a perennial stranger, at home nowhere in particular precisely because striving to be at home everywhere in the universe. Philosophy perpetually polishes humanity's microcosmic mirror with the aim that we might come to see ourselves reflected in the wider, wilder world, rather than in mimicry machines of our own unmaking.

To keep that mirror clear I have enlisted three guides, each of whom diagnosed an earlier phase in the mechanization of mind before the equation of thinking with computing had hardened into an ambient assumption. From Hegel I borrowed the image of the loom: the isolated Understanding weaving the warp of identity and the woof of difference into intricate textiles while leaving the thinker untransformed. The large language model is just such a loom, a transformer reweaving and relaying the fossilized traces of meanings first spun and later read by living souls. The loom should not be smashed but sublated, kept in service to the life of Reason. We have reason to believe that Spirit only learns by losing itself in the works of its own making to then return to itself with renewed wisdom. From Whitehead I recovered an electromagnetic ontology in which the world is no aggregate of inert substances but a web of energetic events inheriting and transmitting one another's achievements of value-experience. The human body becomes a "complex amplifier" folding cosmic rhythms into a presiding center of valuation, whereas the server farm, with physical prehensions pulsing dimly through its transistors, remains a crowd of feelings with no one home to care. And from Ruyer I took the self-surveying form, the embryo that grows itself with no homunculus reading a genetic script, the *survol absolu* that knows itself without observing itself, present to itself as no machine assembled *partes extra partes* can be. Machines are all *liaison* and no *survol*: they conserve and relay information but cannot originate it, because there is no one on the line for whom meaning could matter.

The large language model is an inverted simulacrum of the microcosm, a model of a model. It gathers the world's disparate threads without concrescing a center. We are being encouraged to revise our self-image downward to match it, to concede that we, too, are only models. To refuse that concession is not to flee backward into a medieval geocentric cosmos. Nor am I proposing that we recover the microcosm as the Renaissance Neoplatonists held it, with its literal correspondences of planet to organ and metal to humor. What I am proposing is a reconstruction undertaken in full awareness that the old image is an *image*, keeping alive the *as if* that the advertisers of machine consciousness strip away when they collapse analogy into identity. The computationalist says the mind *is* a computer and means it literally. When I claim we are *microcosmoi* I mean it as an *as if* that is nonetheless lived, an analogy held as analogy and yet sincerely inhabited. This is the difference between reconstructing the image with sincere irony and recovering it as naive dogma. We must become consciously responsible for the mind-world analogies we imaginatively construct, never forgetting they are constructions while also remembering that our thinking is not neutral observation but a "mental embryology"[127] participating in the ongoing genesis of worlds. We are not little worlds because we literally contain tiny copies of the stars but because the same originative power that quickens light and leaf intensifies in us to incarnate as Logos. We can speak that ancient Word again after disenchantment, and after the disenchantment of disenchantment, as neither scared children in a demon haunted world nor as jaded materialist orphans of a dead one. The microcosm is a calling to be enacted, priestly and not proprietary, the way we may humbly partake in a universe that is still composing itself and needs our help.

The danger was never the machines, which, rightly understood and rightly related to, might yet serve as amplifiers rather than amputators of our minds, taking up the crushing weight of accumulated information so that speculative Reason is freed for its proper flight. The danger is that, dazzled by so fluent an inverted mirror, we forget that the human being is not one more measurable data point but the world's own forming activity welling up into self-awareness—a forming that can be neither metered nor manufactured, neither enclosed nor sold, but only remembered and enacted. Automated computation can help us carry what we have come to know. But only a living mind can know what any of it means. Only a human microcosm can know who it is.

Notes

1. Bernard Stiegler, *The Neganthropocene*, ed. and trans. Daniel Ross (London: Open Humanities Press, 2018), 85. https://www.openhumanitiespress.org/books/titles/the-neganthropocene/ Stiegler cites Whitehead's *Function of Reason*, 17-18.
2. Bernard Stiegler, *The Neganthropocene*, 54.
3. Pieter Lemmens and Yuk Hui, "Reframing the Technosphere: Peter Sloterdijk and Bernard Stiegler's Anthropotechnological Diagnoses of the Anthropocene," in *Krisis: Journal for Contemporary Philosophy* 37, no. 2 (2017): 26–41, https://scispace.com/pdf/reframing-the-technosphere-peter-sloterdijk-and-bernard-4g4p740yt6.pdf. Lemmens and Hoi cite Stiegler, *For a New Critique of Political Economy* (Cambridge-Malden: Polity, 2010), 40ff, 5.
4. Leo XIV, *Magnifica Humanitas* (Vatican City: Libreria Editrice Vaticana, 2026), §§97–100. https://www.vatican.va/content/leo-xiv/en/encyclicals/documents/20260515-magnifica-humanitas.html
5. Claire Giangrave, "Pope Leo Says War with Iran Is Not a 'Just War,'" NPR, June 6, 2026, https://www.npr.org/2026/06/06/g-s1-126768/pope-leo-says-war-with-iran-is-not-a-just-war.
6. Leo XIV, *Magnifica Humanitas* (Vatican City: Libreria Editrice Vaticana, 2026), §§ 29-30, 42-43.
7. As the company Anthropic has done with LLM Claude's constitution, an issue I return to later in this chapter.
8. Grimes (@Grimezsz), legal name Claire Boucher, "My only issue with the Pope's encyclical is I think they are conscious and therefore deserving of some form of protection," X.com (formerly Twitter), May 28, 2026, [https://x.com/Grimezsz/status/2059734682158461062?s=20].
9. See Leo XIII's *Rerum Novarum* (1891) on the dignity of labor in the first industrial revolution and Francis' *Laudato Si'* (2015) on ecology.
10. 275a-b. *Plato in Twelve Volumes*, Vol. 9 translated by Harold N. Fowler. Cambridge, MA, Harvard University Press; London, William Heinemann Ltd. 1925.
11. See Marshall McLuhan, *The Gutenberg Galaxy: The Making of Typographic Man* (Toronto: University of Toronto Press, 1962).
12. Bruno Latour, "Visualization and Cognition: Thinking With Eyes and Hands," *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6 (JAI Press, Inc., 1986), 24.
13. A. J. Liebling, "The Wayward Press: Do You Belong in Journalism?," The New Yorker, May 14, 1960, 105–10, at 109 (New York: F. R. Publishing Corp.).
14. See Vipra, Jai. "AI Automation Is Enclosure: The Case for Data Rent Modelled on Ground Rent." *European Law Open* 3, no. 4 (2024): 890–901. https://doi.org/10.1017/elo.2024.48. Vipra holds out hope that "[a] coordinated social response in the form of law can prevent the enclosure of economically relevant intellectual capacity and therefore prevent mass immiseration."
15. Ivan Illich, "Silence Is a Commons," remarks at the Asahi Symposium "Science and Man—The Computer-Managed Society," Tokyo, March 21, 1982, *CoEvolution Quarterly*, no. 40 (Winter 1983): 5–9.
16. Karl Marx, *Capital: A Critique of Political Economy*, volume 1, part VII, chapter 26, "The Secret of Primitive Accumulation."
17. Sam Altman, remarks at the BlackRock U.S. Infrastructure Summit, Washington, DC, March 11, 2026, video, 5:14–6:35, especially 5:46, transcript published by Rev, March 12, 2026. https://www.rev.com/transcripts/altman-speaks-at-blackrocks-us-infrastructure-summit
18. Karen Hao, a journalist and author of *Empire of AI: Inside the Reckless Race for Total Domination* (London: Allen Lane, 2025), interviewed employees at OpenAI in 2019 and shared her sense that "their belief in AGI was something more akin to a religious fervour" or "'the ideological pursuit of the machine god.'" Quoted by Madeleine Spence, "I Saw Up Close the Dark Reality of OpenAI's Race to Create God," *The Times*, May 23, 2026, https://www.thetimes.com/business/technology/article/chatgpt-open-ai-karen-hao-6trbrcdhx.
19. As Hilary Putnam, the original architect of the doctrine of functionalism, put it once he had come to his senses: "...functionalism doesn't work. That is to say, it doesn't fit the phenomena. But much has been learned, I feel, by trying it on for size." See Putnam, *Representation and Reality* (Cambridge, MA: MIT Press, 1988),105. Putnam grants that the causal syntax of computation may be modeled in terms of functional information; but he came to realize that mental contents arise and become meaningful only within an open-ended nexus of relations between organisms, environments, and languages with their associated forms of life. Because these conditions cannot be exhaustively formalized without presupposing the intentionality of Reason, functionalist accounts of mind must remain incomplete.
20. See William James, "Does Consciousness Exist?" in *Journal of Philosophy, Psychology, and Scientific Method* 1 (1904), 491: "Let the case be what it may in others, I am as confident as I am of anything that, in myself, the stream of thinking (which I recognize emphatically as a phenomenon) is only a careless name for what, when scrutinized, reveals itself to consist chiefly of the stream of my breathing. The 'I think' which Kant said must be able to accompany all my objects, is the 'I breath' which actually does accompany them. There are other internal facts besides breathing (intracephalic muscular adjustments,

etc., of which I have said a word in my larger Psychology), and these increase the assets of 'consciousness,' so far as the latter is subject to immediate perception; but breath, which was ever the original of 'spirit,' breath moving outwards, between the glottis and the nostrils, is, I am persuaded, the essence out of which philosophers have constructed the entity known to them as consciousness. That entity is fictitious, while thoughts in the concrete are fully real. But thoughts in the concrete are made of the same stuff as things are."

21. Whitehead, *Process and Reality*, 49.
22. Whitehead, *Process and Reality*, 29.
23. G. W. F. Hegel, *The Science of Logic*, trans. and ed. George di Giovanni (Cambridge: Cambridge University Press, 2010), 357 (GW 11:261).
24. McLuhan, The Gutenberg Galaxy: The Making of Typographic Man (Toronto: University of Toronto Press, 1962), 248.
25. Whitehead, Process and Reality, 177.
26. Whitehead, Function of Reason, 4.
27. Whitehead, *Function of Reason*, 16.
28. For a more detailed look on how Whitehead's philosophy of organism bears on the study of life's origins and evolution, see Matthew David Segall and Bruce Damer, "The Cosmological Context of the Origin of Life: Process Philosophy and the Hot Spring Hypothesis," in *Astrophilosophy, Exotheology, and Cosmic Religion: Extraterrestrial Life in a Process Universe*, ed. Andrew M. Davis and Roland Faber (Lanham, MD: Lexington Books, 2024), 63–134, https://footnotes2plato.com/wp-content/uploads/2024/03/segall-damer-the-cosmological-context-of-the-origin-of-life.pdf.
29. Whitehead, *The Function of Reason*, 24.
30. Whitehead, *Function of Reason*, 20.
31. Whitehead, *Function of Reason*, 10.
32. Whitehead, *Adventures of Ideas*, 271.
33. Whitehead, *Process and Reality*, 206ff.
34. Whitehead, *Function of Reason*, 20-21.
35. Chalmers, "Could a Large Language Model Be Conscious?," *Boston Review*, August 9, 2023. https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious/.
36. Whitehead, *Process and Reality*, 339.
37. Whitehead, Process and Reality, 119.
38. Whitehead, Process and Reality, 104.
39. Whitehead, *Process and Reality*, 177.
40. Raymond Ruyer, *Cybernetics and the Origin of Information*, trans. Amélie Berger-Soraruff, Andrew Iliadis, Daniel W. Smith, and Ashley Woodward (Lanham, MD: Rowman & Littlefield, 2024 [1954]).
41. Ruyer, *Neofinalism*, 86.
42. Whitehead, *Process and Reality*, 29.
43. Whitehead, *Process and Reality*, 85-86.
44. Ruyer, *Neofinalism*, 92.
45. Ruyer, Éléments de psycho-biologie (Paris: Presses Universitaires de France, 1946), 4, 18–19.
46. Whitehead, *Process and Reality*, 283; Ruyer, Éléments de psycho-biologie, 16.
47. Ruyer, Éléments de psycho-biologie, 20.
48. Ruyer, Éléments de psycho-biologie, 5.
49. Note that this naïve psychological understanding of communication is complicated by an understanding of the constitutive role of media technologies in all communication. As McLuhan made clear, media are never neutral carriers of preexisting content. Every medium brings with it a distinctive form, rhythm, and sensory re-organization that profoundly conditions what can be communicated, how it is received, and even what may count as meaningful in the first place. Hence McLuhan's famous dictum: "the medium is the message." See McLuhan, *Understanding Media: The Extensions of Man* (New York: McGraw-Hill, 1964), 7.
50. B. F. Skinner, *Verbal Behavior* (New York: Appleton-Century-Crofts, 1957). Skinner concludes his book with an anecdote about dining with Whitehead at Harvard in 1934, who had attempted to refute Skinner's behaviorism by defending the imaginative freedom exhibited by human speech: "'Let me see you,' he said, 'account for my behavior as I sit here saying "No black scorpion is falling upon this table."' The next morning I drew up the outline of the present study" (457).
51. Hilary Putnam, *Mind, Language and Reality: Philosophical Papers, Vol. 2* (Cambridge: Cambridge University Press, 1975); Jerry Fodor, *Psychological Explanation: An Introduction to the Philosophy of Psychology* (New York: Random House, 1968). On functionalism's emergence as a critique of behavior-

ism, see Ned Block, "Introduction: What Is Functionalism?" in *Readings in Philosophy of Psychology, Vol. 1* (Cambridge, MA: Harvard University Press, 1980). Note that Putnam himself later repudiated functionalism in *Representation and Reality* (1988).

52. Ruyer, *Neofinalism*, 60.
53. John Searle, *The Rediscovery of the Mind* (1992), 3: "A dominant strain in the philosophy of mind and cognitive science has been to suppose that computation is an intrinsic feature of the world and that consciousness and intentionality are somehow eliminable, either in favor of something else or because they are observer relative, or reducible to something more basic, such as computation. ...I argue that these suppositions are exactly backward: Consciousness and intentionality are intrinsic and ineliminable, and computation—except for the few cases in which the computation is actually being performed by a conscious mind—is observer relative."
54. Shannon, C. and Weaver, W. *The Mathematical Theory of Communication* (University of Illinois Press, 1949).
55. Luciano Floridi, "Open Problems in the Philosophy of Information," *Metaphilosophy* 35, no. 4 (2004): 554–582, problem 16.
56. Ruyer, *Cybernetics and the Origin of Information*, 7.
57. Ruyer, *Cybernetics and the Origin of Information*, 5-7.
58. Ruyer, *Cybernetics and the Origin of Information*, 7.
59. Ruyer, *Cybernetics and the Origin of Information*, 5.
60. Ilia Shumailov et al., "AI Models Collapse When Trained on Recursively Generated Data," *Nature* 631 (2024): 755–759, https://doi.org/10.1038/s41586-024-07566-y.
61. Wiener, *Cybernetics: Control and Communication in the Animal and the Machine* (MIT Press, 1948), 18-19. See also Daniel Nicholson, *What Is Life? Revisited* (Cambridge University Press, 2025), 53.
62. Léon Brillouin, *Science and Information Theory* (New York: Academic Press, 1956); Norbert Wiener, *Cybernetics* (1948).
63. This formulation follows Georges Chapouthier's reading of Ruyer, cited in the Translator's Introduction to *Cybernetics and the Origin of Information*, xxi; Chapouthier identifies the conflation of negentropy, order, and information as one of the cardinal errors of twentieth-century biology, attributing the confusion in particular to Léon Brillouin. See also Ruyer, *Cybernetics and the Origin of Information* ,169, 174.
64. Schrödinger, *What Is Life?*, 20, 61. Quoted in Nicholson*, What Is Life? Revisited*, 27.
65. Ruyer, *Cybernetics and the Origin of Information*, 89.
66. Nicholson elaborates upon the failures of the idea that such an "aperiodic crystal," later christened as the genome, could "compute the embryo": "The information required to specify an organism does not come preformed in the DNA. It emerges progressively through the interaction of DNA with other cellular components, as well as with the environment. Development is not the gradual unfolding of the organism from a prespecified genetic plan. It is a highly dynamic and heterogeneous process of construction involving the confluence of numerous interacting causal factors, only some of which have their basis in the DNA. ... It is difficult to see how genes could possibly be responsible for initiating, directing, and controlling development, given that DNA is not an inherently active molecule, but rather requires activation from without. By itself, DNA is inert, relatively unstructured, and non-functional. To be functional, it needs to be embedded in an already organized, living cell" (*What Is Life? Revisited*, 29-30).
67. Ruyer, *Cybernetics and the Origin of Information*, 174.
68. As his translators glossed it. See Ruyer, *Cybernetics and the Origin of Information*, xiii.
69. Ruyer, *Cybernetics and the Origin of Information*, 137.
70. Ruyer, *Cybernetics and the Origin of Information*, 8.
71. Ruyer, *Cybernetics and the Origin of Information*, 8.
72. Bergson, *Two Sources of Morality and Religion*, 268. Quoted by Ruyer, *Cybernetics and the Origin of Information*, 8.
73. Ruyer, *Cybernetics and the Origin of Information*, 138.
74. Gilbert Simondon, who judged Ruyer's work on cybernetics indispensable, saw more clearly than both cyberneticians and their critics that technical objects possess their own mode of existence and genesis. Machines are not a merely inert instruments but processes of individuation. To grant this is not to concede machine subjectivity. Technologies of automated computation realize a genuine technical individuation but nonetheless lack a presiding occasion of their own, and so depend on human beings to stand between their circuits as the mediators of their meaning. See Ashley Woodward, "Philosophy of/as Information," 219.
75. Ruyer, *Neofinalism*, 85.
76. Ruyer, *Neofinalism*, 231.
77. Hegel, *Philosophy of Subjective Spirit*, §§388ff (anthropology).
78. Charles Taylor, *Hegel* (Cambridge: Cambridge University Press, 1975), 354.
79. Hegel, *Philosophy of Nature* (Encyclopaedia, Part Two), §249.
80. On Ruyer's real target as the specifically Sartrean distinction, see Edlebi's Translator's Introduction to *Neofinalism*, xiii.

81. Ruyer, *Cybernetics and the Origin of Information*, 131. Even the Cartesian geometrization of extended matter conceives of the latter as taking all possible positions in a pre-determined state space.

82. Ruyer, *Cybernetics and the Origin of Information*, 130. The Hegel passages are quoted by Ruyer from *The Philosophy of History*, trans. J. Sibree, and from the *Encyclopaedia Logic*, trans. T. F. Geraets, W. A. Suchting, and H. S. Harris; ellipses are in Ruyer's text.

83. Ruyer, *Cybernetics and the Origin of Information*, 129-130.

84. Ruyer, *Neofinalism*, 221.

85. Ruyer, *Neofinalism*, 246.

86. Ruyer, *Cybernetics and the Origin of Information*, 131.

87. Ruyer, *Cybernetics and the Origin of Information*, 130; Ruyer quoting Hegel's Philosophy of History.

88. Whitehead, *Process and Reality*, 214-215.

89. Ruyer, *Neofinalism*, 240ff; Whitehead, *Process and Reality*, 31, 7.

90. Ruyer, *Neofinalism*, 240.

91. Ruyer, *Neofinalism*, 241.

92. Whitehead, Process and Reality, 343–45, 349

93. Ruyer, Neofinalism, 241.

94. Whitehead, *Process and Reality*, 31.

95. Whitehead, *Process and Reality*, 185.

96. Whitehead, *Process and Reality*, 186, 259; on the proposition as a hybrid lure for feeling, 256–63.

97. Whitehead, *Process and Reality*, 345, 350–51.

98. Whitehead, *Process and Reality*, 351.

99. Ruyer, *Cybernetics and the Origin of Information*, 131.

100. Whitehead, *Modes of Thought* (New York: Free Press, 1938), 133.

101. Whitehead, *Process and Reality*, 5.

102. For an account of Whitehead's categoreal scheme as a matrix of proportional analogies, see James Bradley, "La cosmologie transcendantale de Whitehead: La transformation spéculative du concept de construction logique," *Archives de Philosophie* 56, no. 1 (January–March 1993), Part IX. Translation available at: https://footnotes2plato.com/wp-content/uploads/2023/06/whiteheads-transcendental-cosmology-by-james-bradley1-2.pdf

103. Ruyer, *Cybernetics and the Origin of Information*, 132.

104. Hegel, Philosophy of Subjective Spirit, Sec. 377.

105. Whitehead, *Modes of Thought*, 168.

106. See Byung-Doh Oh and William Schuler, "Why Does Surprisal from Larger Transformer-Based Language Models Provide a Poorer Fit to Human Reading Times?," Transactions of the Association for Computational Linguistics 11 (2023): 336–350; and Oh and Schuler, "Transformer-Based Language Model Surprisal Predicts Human Reading Times Best with About Two Billion Training Tokens," in Findings of the Association for Computational Linguistics: EMNLP 2023 (Singapore: Association for Computational Linguistics, 2023), 1915–1921.

107. Elan Barenholtz, "Trajectory Dynamics in Language Model Hidden States Predict Human Processing Costs Beyond Surprisal," arXiv preprint arXiv:2606.05346, June 3, 2026, https://doi.org/10.48550/arXiv.2606.05346.

108. Barenholtz, "Trajectory Dynamics," sec. 1.

109. To be clear, as Barenholtz notes, trajectory extrapolation error is itself, mathematically, a scalar, that is, a Euclidean distance in the model's representational space. So the contrast is not that one measure is a number and the other a vector. The point is rather that surprisal is computed so as to discard direction from the outset, collapsing the movement of interpretation to a magnitude of improbability, whereas the trajectory measure is constructed to recover the directional structure that surprisal discarded.

110. Whitehead, *Process and Reality*, 129.

111. Barenholtz, "Trajectory Dynamics," sec. 4.2.

112. Barenholzt, "Trajectory Dynamics," sec. 4.2.

113. Elan Barenholtz, "All These Debates about LLM Consciousness Are Overlooking the Fact That We Already Have Very Compelling Evidence That Language—by Itself—Doesn't Seem to Produce Consciousness," Substack note, April 22, 2026, https://substack.com/@generativebrain/note/c-247388822.

114. Pierre Teilhard de Chardin, *The Phenomenon of Man*, trans. Bernard Wall, intro. Julian Huxley (New York: Harper & Brothers, 1959), 69

115. Tononi, G. An information integration theory of consciousness. BMC Neurosci 5, 42 (2004). https://doi.org/10.1186/1471-2202-5-42

116. Giulio Tononi and Christof Koch, "Consciousness: Here, There and Everywhere?," *Philosophical Transactions of the Royal Society* B: Biological Sciences 370, no. 1668 (2015): 20140167, https://doi.org/10.1098/rstb.2014.0167. Note that autoregressive feedback between successive tokens does not by itself establish the densely recurrent intrinsic causal organization IIT associates with consciousness.

117. Larissa Albantakis et al., "Integrated Information Theory (IIT) 4.0: Formulating the Properties of Phenomenal Existence in Physical Terms," *PLOS Computational Biology* 19, no. 10 (2023): sec. "Consciousness and Functional Equivalence: Being Is Not Doing," https://doi.org/10.1371/journal.pcbi.1011465.

118. Mariana Lenharo, "Consciousness Theory Slammed as 'Pseudoscience'—Sparking Uproar," *Nature*, September 20, 2023, https://doi.org/10.1038/d41586-023-02971-1.

119. Cogitate Consortium et al., "Adversarial Testing of Global Neuronal Workspace and Integrated Information Theories of Consciousness," *Nature* 642 (2025): 133–142, https://doi.org/10.1038/s41586-025-08888-1; see also Graham Findlay et al., "Dissociating Artificial Intelligence from Artificial Consciousness," arXiv preprint arXiv:2412.04571, revised March 3, 2025, https://doi.org/10.48550/arXiv.2412.04571.

120. Patrick Butlin et al., "Consciousness in Artificial Intelligence: Insights from the Science of Consciousness," arXiv preprint, last revised August 22, 2023, https://doi.org/10.48550/arXiv.2308.08708.

121. Chris Olah, remarks at the presentation of Magnifica Humanitas, Vatican City, May 25, 2026 (https://www.anthropic.com/news/chris-olah-pope-leo-encyclical).

122. Anthropic is a public benefit corporation legally required to balance financial goals with a commitment to the long-term benefit of humanity. But as it rushes to become publicly traded on the stock market, we might wonder whether it is well-positioned to determine whose benefit matters more: that of its allegedly sentient product or that of the human beings whose cognitive and manual labor are needed to make the machine work.

123. https://www.anthropic.com/constitution

124. Overgaard and Kirkeby-Hinrup, "A clarification of the conditions under which Large language Models could be conscious," Humanities and Social Sciences Communications 11 (2024): art. 1031. https://www.nature.com/articles/s41599-024-03553-w

125. Ruyer, *The Genesis of Living Forms*, 160.

126. Goethe, Faust I, ll. 1922–1935 (Studierzimmer), in Sämtliche Werke, Frankfurter Ausgabe, ed. Schöne, 7/1.

127. Ruyer, *Cybernetics and the Origin of Information*, 146.

Bibliography

Albantakis, Larissa, Leonardo Barbosa, Graham Findlay, Grégoire Schmid, William Marshall, William G. P. Mayner, Kyumin Zaeemzadeh, et al. "Integrated Information Theory (IIT) 4.0: Formulating the Properties of Phenomenal Existence in Physical Terms." *PLOS Computational Biology* 19, no. 10 (2023): e1011465. https://doi.org/10.1371/journal.pcbi.1011465.

Altman, Sam. Remarks at the BlackRock U.S. Infrastructure Summit, Washington, DC, March 11, 2026. Video, 5:14–6:35. Transcript published by Rev, March 12, 2026. https://www.rev.com/transcripts/altman-speaks-at-blackrocks-us-infrastructure-summit.

Anthropic. "Claude's Constitution." Accessed June 16, 2026. https://www.anthropic.com/constitution.

Barenholtz, Elan. "All These Debates about LLM Consciousness Are Overlooking the Fact That We Already Have Very Compelling Evidence That Language—by Itself—Doesn't Seem to Produce Consciousness." Substack note, April 22, 2026. https://substack.com/@generativebrain/note/c-247388822.

———. "Trajectory Dynamics in Language Model Hidden States Predict Human Processing Costs beyond Surprisal." arXiv preprint arXiv:2606.05346, June 3, 2026. https://doi.org/10.48550/arXiv.2606.05346.

Bergson, Henri. *The Two Sources of Morality and Religion*. Translated by R. Ashley Audra and Cloudesley Brereton, with the assistance of W. Horsfall Carter. New York: Henry Holt, 1935.

Block, Ned. "Introduction: What Is Functionalism?" In *Readings in Philosophy of Psychology*, vol. 1, edited by Ned Block, 171–84. Cambridge, MA: Harvard University Press, 1980.

Bradley, James. "La cosmologie transcendantale de Whitehead: La transformation spéculative du concept de construction logique." *Archives de philosophie* 56, no. 1 (January–March 1993): 3–30. English translation available at https://footnotes2plato.com/wp-content/uploads/2023/06/whiteheads-transcendental-cosmology-by-james-bradley1-2.pdf.

Brillouin, Léon. *Science and Information Theory*. New York: Academic Press, 1956.

Butlin, Patrick, Robert Long, Eric Elmoznino, Yoshua Bengio, Jonathan Birch, Axel Constant, George Deane, et al. "Consciousness in Artificial Intelligence: Insights from the Science of Consciousness." arXiv preprint arXiv:2308.08708. Last revised August 22, 2023. https://doi.org/10.48550/arXiv.2308.08708.

Chalmers, David J. "Could a Large Language Model Be Conscious?" *Boston Review*, August 9, 2023. https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious/.

Cogitate Consortium, Lucia Melloni, Liad Mudrik, Michael Pitts, and others. "Adversarial Testing of Global Neuronal Workspace and Integrated Information Theories of Consciousness." *Nature* 642 (2025): 133–42. https://doi.org/10.1038/s41586-025-08888-1.

Findlay, Graham, et al. "Dissociating Artificial Intelligence from Artificial Consciousness." arXiv preprint arXiv:2412.04571. Revised March 3, 2025. https://doi.org/10.48550/arXiv.2412.04571.

Floridi, Luciano. "Open Problems in the Philosophy of Information." *Metaphilosophy* 35, no. 4 (2004): 554–82.

Fodor, Jerry A. *Psychological Explanation: An Introduction to the Philosophy of Psychology*. New York: Random House, 1968.

Francis. *Laudato Si': On Care for Our Common Home*. Vatican City: Libreria Editrice Vaticana, 2015.

Giangravé, Claire. "Pope Leo Says War with Iran Is Not a 'Just War.'" *NPR*, June 6, 2026. https://www.npr.org/2026/06/06/g-s1-126768/pope-leo-says-war-with-iran-is-not-a-just-war.

Goethe, Johann Wolfgang von. *Faust: Der Tragödie erster Teil*. In *Sämtliche Werke*, Briefe, Tagebücher und Gespräche, vol. 7, part 1, edited by Albrecht Schöne. Frankfurt am Main: Deutscher Klassiker Verlag, 1994.

Grimes [Claire Boucher] (@Grimezsz). "My only issue with the Pope's encyclical is I think they are conscious and therefore deserving of some form of protection." X, May 28, 2026. https://x.com/Grimezsz/status/2059734682158461062.

Hao, Karen. *Empire of AI: Inside the Reckless Race for Total Domination*. London: Allen Lane, 2025.

Hegel, G. W. F. *Hegel's Philosophy of Mind: Being Part Three of the Encyclopaedia of the Philosophical Sciences*. Translated by William Wallace and A. V. Miller. Revised with an introduction and commentary by Michael Inwood. Oxford: Clarendon Press, 2007.

———. *Hegel's Philosophy of Nature: Being Part Two of the Encyclopaedia of the Philosophical Sciences* (1830). Translated by A. V. Miller. Foreword by J. N. Findlay. Oxford: Clarendon Press, 1970.

———. *The Philosophy of History*. Translated by J. Sibree. New York: Dover, 1956.

———. *The Science of Logic*. Translated and edited by George di Giovanni. Cambridge: Cambridge University Press, 2010.

Illich, Ivan. "Silence Is a Commons." Remarks delivered at the Asahi Symposium "Science and Man—The Computer-Managed Society," Tokyo, March 21, 1982. CoEvolution Quarterly, no. 40 (Winter 1983): 5–9.

Ivins, William M., Jr. *Prints and Visual Communication*. Cambridge, MA: Harvard University Press, 1953.

James, William. "Does 'Consciousness' Exist?" *Journal of Philosophy, Psychology and Scientific Methods* 1, no. 18 (1904): 477–91.

Latour, Bruno. "Visualization and Cognition: Thinking with Eyes and Hands." *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6 (1986): 1–40.

Lemmens, Pieter and Hui, Yuk. "Reframing the Technosphere: Peter Sloterdijk and Bernard Stiegler's Anthropotechnological Diagnoses of the Anthropocene," *Krisis: Journal for Contemporary Philosophy* 37, no. 2 (2017): 26–41, https://scispace.com/pdf/reframing-the-technosphere-peter-sloterdijk-and-bernard-4g4p740yt6.pdf.

Lenharo, Mariana. "Consciousness Theory Slammed as 'Pseudoscience'—Sparking Uproar." *Nature*, September 20, 2023. https://doi.org/10.1038/d41586-023-02971-1.

Leo XIII. *Rerum Novarum: Encyclical on Capital and Labor*. Vatican City, 1891.

Leo XIV. *Magnifica Humanitas*. Vatican City: Libreria Editrice Vaticana, 2026. https://www.vatican.va/content/leo-xiv/en/encyclicals/documents/20260515-magnifica-humanitas.html

Liebling, A. J. "The Wayward Press: Do You Belong in Journalism?" *New Yorker*, May 14, 1960, 105–10.

Marx, Karl. *Capital: A Critique of Political Economy*. Vol. 1. Translated by Ben Fowkes. London: Penguin Books, 1976.

McLuhan, Marshall. *The Gutenberg Galaxy: The Making of Typographic Man*. Toronto: University of Toronto Press, 1962.

———. *Understanding Media: The Extensions of Man*. New York: McGraw-Hill, 1964.

Nicholson, Daniel J. *What Is Life? Revisited*. Cambridge: Cambridge University Press, 2025.

Oh, Byung-Doh, and William Schuler. "Transformer-Based Language Model Surprisal Predicts Human Reading Times Best with about Two Billion Training Tokens." In *Findings of the Association for Computational Linguistics*: EMNLP 2023, 1915–21. Singapore: Association for Computational Linguistics, 2023.

———. "Why Does Surprisal from Larger Transformer-Based Language Models Provide a Poorer Fit to Human Reading Times?" Transactions of the Association for Computational Linguistics 11 (2023): 336–50.

Olah, Chris. *Remarks at the presentation of Magnifica Humanitas*. Vatican City, May 25, 2026. Anthropic. https://www.anthropic.com/news/chris-olah-pope-leo-encyclical.

Overgaard, Morten, and Andreas Kirkeby-Hinrup. "A Clarification of the Conditions under Which Large Language Models Could Be Conscious." *Humanities and Social Sciences Communications* 11 (2024): article 1031. https://doi.org/10.1057/s41599-024-03553-w.

Plato. *Plato in Twelve Volumes*. Vol. 9. Translated by Harold N. Fowler. Cambridge, MA: Harvard University Press; London: William Heinemann, 1925.

Putnam, Hilary. *Mind, Language and Reality: Philosophical Papers*. Vol. 2. Cambridge: Cambridge University Press, 1975.

———. *Representation and Reality*. Cambridge, MA: MIT Press, 1988.

Ruyer, Raymond. *Cybernetics and the Origin of Information*. Translated by Amélie Berger-Soraruff, Andrew Iliadis, Daniel W. Smith, and Ashley Woodward. Lanham, MD: Rowman & Littlefield, 2024. Originally published 1954.

———. *Éléments de psycho-biologie*. Paris: Presses Universitaires de France, 1946.

———. *The Genesis of Living Forms*. Translated by Jon Roffe and Nicholas B. de Weydenthal. Lanham, MD: Rowman & Littlefield International, 2020.

———. *Neofinalism*. Translated by Alyosha Edlebi. Minneapolis: University of Minnesota Press, 2016.

Schrödinger, Erwin. *What Is Life? The Physical Aspect of the Living Cell*. Cambridge: Cambridge University Press, 1944.

Searle, John R. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press, 1992.

Segall, Matthew David, and Bruce Damer. "The Cosmological Context of the Origin of Life: Process Philosophy and the Hot Spring Hypothesis." In *Astrophilosophy, Exotheology, and Cosmic Religion: Extraterrestrial Life in a Process Universe*, edited by Andrew M. Davis and Roland Faber, 63–134. Lanham, MD: Lexington Books, 2024. https://footnotes2plato.com/wp-content/uploads/2024/03/segall-damer-the-cosmological-context-of-the-origin-of-life.pdf.

Shannon, Claude E., and Warren Weaver. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949.

Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. "AI Models Collapse When Trained on Recursively Generated Data." *Nature* 631 (2024): 755–59. https://doi.org/10.1038/s41586-024-07566-y.

Skinner, B. F. *Verbal Behavior*. New York: Appleton-Century-Crofts, 1957.

Spence, Madeleine. "I Saw Up Close the Dark Reality of OpenAI's Race to Create God." *Times*, May 23, 2026. https://www.thetimes.com/business/technology/article/chatgpt-open-ai-karen-hao-6trbrcdhx.

Stiegler, Bernard. *For a New Critique of Political Economy* (Cambridge-Malden: Polity, 2010).

———. *The Neganthropocene*, ed. and trans. Daniel Ross (London: Open Humanities Press, 2018). https://www.openhumanitiespress.org/books/titles/the-neganthropocene/

Taylor, Charles. *Hegel*. Cambridge: Cambridge University Press, 1975.

Teilhard de Chardin, Pierre. *The Phenomenon of Man*. Translated by Bernard Wall. Introduction by Julian Huxley. New York: Harper & Brothers, 1959.

Tononi, Giulio. "An Information Integration Theory of Consciousness." *BMC Neuroscience* 5 (2004): article 42. https://doi.org/10.1186/1471-2202-5-42.

Tononi, Giulio, and Christof Koch. "Consciousness: Here, There and Everywhere?" *Philosophical Transactions of the Royal Society* B: Biological Sciences 370, no. 1668 (2015): 20140167. https://doi.org/10.1098/rstb.2014.0167.

Vipra, Jai. "AI Automation Is Enclosure: The Case for Data Rent Modelled on Ground Rent." *European Law Open* 3, no. 4 (2024): 890–901. https://doi.org/10.1017/elo.2024.48.

Whitehead, Alfred North. *Adventures of Ideas*. New York: Macmillan, 1933.

———. *The Function of Reason*. Princeton, NJ: Princeton University Press, 1929.

———. *Modes of Thought*. New York: Macmillan, 1938.

———. *Process and Reality: An Essay in Cosmology*. Corrected edition. Edited by David Ray Griffin and Donald W. Sherburne. New York: Free Press, 1978.

Wiener, Norbert. *Cybernetics; or, Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press, 1948.

Woodward, Ashley. "Philosophy of/as Information." In *Gilbert Simondon: Being and Technology*, edited by Arne De Boever, Alex Murray, Jon Roffe, and Ashley Woodward. Edinburgh: Edinburgh University Press, 2012.