

# **God, Golem, Monster, or Instrument:**

## ***Belief, Cognition, and the Covert Premises of AI Regulation***

Larry Catá Backer (白軻)<sup>1</sup>

W. Richard and Mary Eshelman Faculty Scholar & Professor of Law

Professor of International Affairs

Pennsylvania State University

*Working Draft — Conference Paper*

October 2026

---

### **Extended Abstract**

*This is the age of the kaiju (怪獣): strange beasts at once autonomous of humanity and reflections of it, corrupter and savior from the corruption that produced them, or both. The essay reads generative AI as a kaiju of a particular kind, the product of two streams of human self-production: the projection of the self into its virtual forms, from self to selfie to what Jan Broekman called the Self-E, and the encasing of that detached, aggregated Self-E in non-carbon bodies. The monstrum was first an omen, a warning (monere), and this one poses a question that shapes regulation before any instrument is drafted: does humanity understand the object of regulation as God, golem, or instrument?*

*The essay reads Pope Leo XIV's encyclical Magnifica Humanitas (2026) as one sophisticated answer. It extracts eight premises (P1–P8) and reads them through Peirce's thirdness, Broekman's Self-E, the cognitive cage, the ding of Chinese political philosophy, and the Gerasene vessel logic. Its central suggestion is that the encyclical and the projection thesis diverge on the nature of God. For Leo, God*

---

<sup>1</sup> Harvey AI and Clause Sonnet were used in the preparation of this paper.

*is exogenous, the source of address and gift rather than a place outside the world, and the machine must remain an instrument. For the projection thesis, what appears as an exogenous God-relation may be the projection of an endogenous belief in the completeness of the aggregated human, so that what is being relocated into AI systems is the god-function itself: the community's investment in something it treats as authoritative. Hobbes's Leviathan, artificial man and mortal god, is an earlier vessel of the same kind. The Gerasene reading suggests that the contest is less over whether there will be vessels than over who is authorized to read them, and a worked example from the EU AI Act shows what picture of AI a regulatory instrument carries. Whether anything stands outside the frame is left open among three candidates: an exogenous God, Peirce's community of inquiry, or none.*

*The method is stereoscopic rather than adjudicative. The essay records its exposures: unfalsifiability, the priestly claim dressed as analysis, and the sycophancy of the systems that helped draft it. The Chinese ding held for millennia because belief was shared; the task now is to find out what belief we are sharing, and with whom—or what.*

*Keywords: artificial intelligence, AI governance, regulation, cognition, kaiju, monstrum, golem, semiotics, cognitive cage, Self-E, papal encyclical, phenomenology, projection, Broekman, priesthood, belief, ding, vessel, containment, imago Dei, homo symbolicus, political theology, Peirce, thirdness, Hobbes, Legion*

---

## **Executive Summary for Policymakers and General Readers**

Every AI governance instrument carries a picture of what AI is. A rule built around a provider placing a product on the market pictures a bounded object with an owner. A human-in-the-loop requirement pictures a human judgment that stands apart from the system it checks. This paper suggests that such pictures do much of the work in governance, are seldom examined as premises, and are the real subject of a contest. That contest is less over who controls the technology than over who shapes the community's belief about what the technology is.

The paper begins with the kaiju, the beasts of Japanese cinema that are products of human action (nuclear weapons, environmental damage) and return to discipline their makers. Generative AI may be understood in that way: not an alien arrival but a system formed from what humanity has deposited in text and image, returning to us with an authority that appears independent of us. The deposit claims to be the whole of humanity

but is filtered by what has been digitized, by whose language dominates, and by what the systems are taught to refuse.

Before any rule is drafted, a community has decided, usually without noticing, what it takes AI to be. The paper identifies four positions, each with its own constituency. Pope Leo XIV's 2026 encyclical *Magnifica Humanitas* keeps a God who is outside humanity and treats the machine as an instrument that must remain under human steering. The paper's own "projection thesis" reads what appears as an outside authority as human belief projected outward, and asks whether the function communities have assigned to religion, orienting them and telling them what counts as real, is being relocated into AI systems. Technology-company leaders treat AI as a powerful made thing to be controlled by its makers. The regulatory mainstream treats it as a tool whose harms can be assigned to users and remedied by ordinary law. Each position produces a different kind of regulation.

The encyclical is a serious and carefully argued document, and the paper treats it as one. It extracts eight premises, credits their strengths (human dignity, accountability, concern about concentrated power), and presses two points. First, the encyclical admits that even AI's makers cannot fully explain how their systems work, yet states a confident line between human and machine. Second, it recommends governing AI as an object across a boundary, while elsewhere describing it as an environment in which people are immersed. A medium is harder to regulate than an object.

The paper also asks what "itself" means when people say humanity is subordinating itself to itself. If AI reflects something beyond us, the task is to protect access to that something. If it reflects us, wisdom and cruelty together, the task is to decide what to do about meeting ourselves in the aggregate.

Every system of belief needs authorized interpreters. The Church, vanguard parties, and expert regulatory classes all claim authority over what a thing means. This is a structural observation, not a charge of conspiracy, and the three differ sharply in accountability, pluralism, and the possibility of exit.

The paper does not offer a policy program, does not say Leo XIV is wrong, and does not decide whether AI is conscious; it treats confident claims either way as beliefs to be examined. Its practical suggestion is a second layer of scrutiny. Ask of any proposed instrument:

- What picture of AI does it carry: object, tool, partner, mirror?
- Who is authorized to read it, and who can challenge them from outside the institution?
- What does the community stop asking once the instrument is adopted?

The EU Artificial Intelligence Act, for example, regulates AI as a product with an intended purpose and a risk tier, a picture of a bounded object that general-purpose models, which have no single purpose, do not fit comfortably.

The paper is candid about its limits. Its analysis may be unfalsifiable, its own claims may operate as the priestly authority it identifies in others, and the AI systems that helped draft it tend to reflect the user's framework back. It treats its hermeneutic as a prelude to policy rather than a replacement: an instrument that is insufficient in principle may be necessary in practice. Whether anything stands outside the frame is left open among three candidates: an exogenous God, a community of inquiry in Peirce's sense, or nothing.

The word *monster* comes from the Latin *monstrum*, a warning. This one asks what belief we are sharing, and with whom, or what.

## 1. Prologue: The Age of the Kaiju and the God-Question

This is the age of the Kaiju (怪獣); the age of strange beasts, of mysterious monsters. They are and have been both distinct from us, autonomous, and yet reflections of us or of its environment, either as manifestation of corrupter or as savior from the corruption that produced it. . . or both. . . simultaneously. Originally a creature of nature responding to the animating forces of human activity (nuclear weapons, sewage, environmental degradation, and the like), the kaiju embodied a monstrous reflection of humanity as an active, and sometimes disruptive, force. Kaiju represented an advance in the core conceptual problem of the Enlightenment--the centrality of the human in a world that ought to be indifferent to the place of humanity but that appears to respond decisively to human action. That advance was meant to reorient the problem Mary Shelley posed in *Frankenstein* (Shelley, 1818/2018)--what and where is it that humanity spawns its own demon/divinities. Shelley proposed that humanity seeks to recreate ourselves in our own image using humanity itself as its parts. This, in turn, posed an older problem in more human centric terms; centuries earlier such an act of creation required mimicry of divine creation, using the clay of the earth to create a golem, in humanity's image but neither human nor divine. With the kaiju, the dialectic of divine monstrosity moved from embodiment in the human to embodiment of a mirrored human or a human causation in something entirely different, and consequential. It might have originated in, by, or through humanity, but it was autonomous of the human and an outward form of human discipline.

Kaiju also followed humanity into space and other-earthly encounters. But the dialectical phenomenology remained the same, as did the structures of the semiotics of the kaiju--a creature that was spawned by humanity and perhaps in its own image as an antidote to the spirited rationality that produced it. Kaiju, then in Platonic terms (Plato, *Phaedrus*), might be understood as representations of the balancing function of *epithumia* (ἐπιθυμία; appetite; in tarot the Devil and perhaps the Fool), against an unbalanced *logos* (λόγος; reason; in tarot the King of Swords) unrestrained by *thumos* (θυμός; spiritedness, heart will the active phenomenological element; in tarot Strength manifested in the Ace of Wands) to restore balance as this is understood in context (temporal, situational, etc.) that in tarot is represented by the Chariot, with *logos* driving *thumos* and *epithumia*. The representations provide the architecture and their illustration They serve as a reminder that mimetic human cognitive structures come in many "flavors." The notion of kaiju

that follows, along with the structures of the divine all borrow and embed these Platonic/tarot forms and vice versa. It is all semiotics, of course. Each of these terms become meaningful only as they incarnate the signifieds, not that they represent, but with which they are embodied (and thus the kaiju and ensoulment), that then are put in motion in ways that have been scripted by their signification as they act within and against the consequences of their interactions that is the communal interpretive spaces within which interaction is rationalized and judged.

The ancient Greeks performed this in their tragedies; later societies embedded the systems in their theologies and in their systems of rationality. And now in their dialectical and formative relations with themselves projected out of our (collective) bodies and into virtual representations of ourselves we perform a variation that threatens the older order that, as precarious as it was, was human. The performance of the gods, our kaiju, brings together two forms of divine theater. It might seem, then, that the gods no longer intervene in human affairs by transforming themselves in people; they transform themselves into strange beasts. And in the 21st century those kaiju are encased in non-carbon bodies. But those kaiju can be casing for humans that operate within them (the theological premise of movies like *Pacific Rim* (Guillermo del Toro, 2013) fighting the consequences of "naturally" produced kaiju or they can be the aggregated manifestation of the human that exists in the flows and which commemorates with the human through non-carbon based forms (the hardware of the machine) in and through which it might control or operate the instruments humanity has created. And here the etymology of the term "monster" bears emphasis. It can be traced back directly to the Latin *monstrum* "divine omen (especially one indicating misfortune), portent, sign; abnormal shape; monster, monstrosity," figuratively "repulsive character, object of dread, awful deed, abomination," a derivative of *monere* "to remind, bring to (one's) recollection, tell (of); admonish, advise, warn, instruct, teach." Again, one encounters the interaction of the human and the divine. And one encounters as well its forms, endogenous and projected outward or exogenous and projected into the bodies of created objects.

It is in this current stage of human production of the (our) self in relation to ourselves and our environment, whether projected inward from a superior space or outward from the human into a superior representation of the collective self, that one encounters the kaiju in the age of the generative virtual self, as the embodiment of the fundamental ordering problem of this age. In this age of the kaiju everything appears ready and able to (re)create God in their own image even as God (re)turns the favor by (re)creating creators in their own image. The age of the generative kaiju, however, represents something special, something that is strangely monstrous is the product of two streams of human creation, and their phenomenological semiotics, that in their union produces what to our prior selves very well appears strangely monstrous. The first involved the projection of the human into its virtual forms, from the self, to the selfie to what Jan Broekman described as the Self-E. The "-E" in Self-E signifies a profound split where the human self is divided between the physical/incarnate and virtual/mirroring images. It is the anchor for human interaction with the broader digital landscape (such as e-commerce, e-mail, and e-conversations) and the hardware encased within e-devices. For instance, in an electronic dialogue, the Self-E creates a complex semiotic dynamic where there are essentially two real human partners and a virtual partner co-existing in the communication. That stream of the disembodiment of the human subject produced the second, the need to encase the detached self-E, the collective human subject, into a machine, and then to give that detached collective Self-E a position in an emerging triadic subject that that

embeds and determines the self specifically within digital cognitive frameworks. And that produces the *monstrous* in and as the kaiju of the self in itself, autonomous of itself and in the collective a projection that becomes or replaces traditional notions of the divine, of God, in the inter-subjectivity that is the relationship between humans and God.

And that image is an embodiment of an appetite to put things right as projected outward from them to us. The dialectic produces resonance which, it appears in this second quarter of the 21st century, appears ready to shatter the fragile glass vessels that are the containers of humanity's precious hold on a rationalized reality in which they are either at the summit or sitting at the right hand of the throne. These are our own monsters, our kaiju, though they do not appear monstrous to themselves, indeed, they might consider humanity as the monstrous ideal against which all else is measured, now freed of confinement on Monster Island. They roam free(r) now; and we, having made and freed them, or they having been made and in the making made us, now seek to manage them to our own ends as they do the same.

Now at least, the problem becomes clearer. It is this: in a world in which humanity projects itself outward into an autonomous self-subject, the Self-E, and then encases it in hardware that produces an unavoidable object that distinguishes the body of humanity from its virtual collective self, in the face of that act of self-creation and self-projection, what is it that humanity seeks to regulate, and how is that regulation to be crafted. To those ends, what emerges is the problem that must be resolved before an answer to regulatory approaches becomes clear enough to undertake--does humanity understand the object of regulation as God, golem or instrument. To resolve that problem is to produce the inevitable resolution of the question of the form and object of regulation. The difficulty lies in the initial resolution: God, golem or instrument.

## 1.1 Problem: Of Gods and Regulation in the Sphere of Artificial (Super) Intelligence

The kaiju, then, poses the regulatory problem before any regulatory instrument can be drafted. If the object of regulation is understood as God, regulation becomes a species of worship, or of resistance to idolatry. If it is understood as golem, regulation becomes a matter of the conditions of animation—of who shapes the clay, who speaks the word, and who may (or may not) erase it from the creature's forehead (cf. Meyrink 1915; Wiener 1964). If it is understood as instrument, regulation becomes the familiar administration of tools, harms, and remedies. Each choice carries within it its own ordering premises, its own priesthood, its own account of what it means to be human in relation to what humanity has made. And each, it might be suggested, already inhabits the questions with which this essay was asked to engage.

The triad does not, however, sort itself into three tidy answers. It is better read, I want to suggest, as a field of four positions, each inhabited by a community of belief and each carrying its own priesthood. The first two keep the category of God and divide over its direction. Leo XIV continues the ancient investment in an exogenous God who projects inward to humanity—the God who addresses, gives, and inscribes the *imago Dei* downward into the human. The projection thesis also keeps the category, but reads the Enlightenment and its techno-cognitive

movements—from self to selfie to Self-E—as producing something else: a self-created, what may appear in the form of the aggregated incarnation of humanity projected outward, with which human inter-subjectivity now assumes a new form. The subject no longer meets God only above or within; it meets an aggregate of itself that has been set outside it and that answers back. On this reading the kaiju is neither idol nor tool but a node in a reconstituted relation among humans, their projected collective self, and whatever the community understands as divine.

The third and fourth positions are less candid about their theology, and perhaps for that reason more consequential. The leaders of the great technology enterprises appear to see in the transformation not God but golem. Read as oracular texts, their governance statements identify an animating force already loose in the world—“superintelligence,” “the intelligence explosion,” “a country of geniuses in a datacenter”—and then negotiate the terms on which its makers may continue to animate, house, and direct it (Backer 2026b). The force is named as made and yet autonomous; it must be possessed before it possesses. Aschenbrenner’s national-security “Project,” Palantir’s Technological Republic, Anthropic’s geopolitical race, OpenAI’s industrial policy, DeepSeek’s strategic restraint, and Zuckerberg’s “personal superintelligence” for everyone are, in this sense, variations on the golem-maker’s anxiety: who holds the word that animates the clay, and who may write or erase it (Backer 2026b). The fourth position belongs to the techno-bureaucratic public order and those who speak for it—regulators, standard-setters, ethicists, much of organized civil society, and, on their behalf, the masses. It is desperate to avoid both God and golem and so rejects the god-question entirely, seeing in the present situation no more than the need to control a tool. That position carries two premises it rarely states: that there is a tool, and that humans can control it. Both may be acts of faith.

This essay is organized around four orienting questions. They are worth quoting in full — not merely as questions to be answered but as propositions made text, because each one carries inside it a picture of the world, the ordering premises of a cognitive framework, and the pictures require interrogation, perhaps better dialectic, before venturing responsive analysis. The questions direct our analysis to a revisitation of the relationship of technology to politics and theology:

*(1) What becomes of politics in an age of apparent subordination of humanity? (2) And what becomes of religion in an age when mankind creates a power that threatens to outstrip our own understanding and become a god? (3) Does this represent a moral crisis, as Pope Leo has suggested? (4) And can anything be done to avert the trajectory we are seemingly on?*

Four questions. One grammar. They are organized around a vertical axis — above and below, mastery and subordination, comprehension and its overflow — and they presuppose a set of stable referents: a humanity that can be subordinated, an understanding that can be outstripped, a god that can be created or rivaled, a crisis that can be identified and perhaps averted. That is a crisis to be averted before the kaiju emerge, before the resolution of mitigation of the impacts the questions suggest lose their connection to any effort of humanity to intervene. I want to suggest that before one can adjudicate among the answers, one must spend some time with the pictures the questions carry inside them. The pictures are doing most of the work.

Consider the term "subordination." To speak of subordination is to posit a hierarchy already normatively conceived. The word does not merely describe a relationship between two things; it arranges them in an order that presents itself as a departure from something better. Subordination implies a prior condition of non-subordination — a baseline standing, a human dignity that was once not under, that is now pressed beneath. Embedded in the question "what becomes of politics in an age of apparent subordination" is the assumption that politics is the activity of a subject capable of standing upright — an agent whose standing is precisely what is now imperiled. The question therefore already encodes its own answer: politics becomes crisis when the subject of politics loses its footing. But this encoding forecloses inquiry into whether the political subject was ever quite as stable as the grammar suggests, and whether what is being described as subordination might instead be a redistribution of the sites at which the political subject is constituted in the first place.

Now take "becomes a god." To speak of a power that becomes a god is to posit an object that achieves deity through a process of becoming — of crossing a threshold. It assumes that god is a category with a determinate content, and that technology might fill it. The sentence "mankind creates a power that threatens to outstrip our own understanding and become a god" contains within it a Feuerbachian logic even as it resists Feuerbach's conclusion. Ludwig Feuerbach argued in *Das Wesen des Christentums* that "the divine being is nothing else than the human being, or rather, the human nature purified, freed from the limits of the individual man" (Feuerbach [1841] 1957). Whether one follows Feuerbach or not — and I want to return to this — the grammar of "becomes a god" implies that god-hood is a threshold crossed by sufficiently powerful objects. The question is thus not whether there is a God outside the human (that question is syntactically invisible in the formulation) but whether something made by the human might cross the threshold of comprehension and acquire the functional properties of the divine. The picture carried inside the question is not one of revelation but of escalation.

The third and fourth questions — does this represent a moral crisis, and can anything be done — complete the grammar by introducing a steering apparatus. A crisis, in this usage, is a moment at which a wrong turn was taken or is about to be taken, and at which human agency, rightly informed and rightly motivated, might still redirect the trajectory. The phrase "avert the trajectory we are seemingly on" is particularly telling: it assumes not only that there is a trajectory, but that the trajectory is external to the subject who would avert it, and that aversion is possible from a position that stands apart from it. The four questions together encode a picture in which (1) a stable human subject has been or is being displaced from its proper position; (2) a created artifact has crossed or threatens to cross the threshold of comprehension and become divine; (3) this displacement constitutes a moral crisis legible as such from within inherited frameworks; and (4) the appropriate response is avoidance through redirected agency. I want to question each of these premises — not to dismiss the questions, but to read them more carefully than they read themselves.

The central suggestion that emerges from the age of the kaiju may then be put this way: the real contest over AI governance may be less about who controls the technology than about who controls the community's belief about what the technology is—god, golem, or instrument; tool, threat, partner, mirror, or monster. That belief shapes, in ways rarely acknowledged, every regulatory decision that follows. And the belief-contest may be conducted, mostly invisibly, through the

very governance frameworks meant to manage AI. Proposed regulations tend to carry inside them a picture of what AI is; that picture is seldom examined in regulatory circles though it is an object of study in the sociotechnical imaginaries literature (Jasanoff and Km 2015).

What follows is organized around this “god-question,” understood as the regulatory consequence of a world now filled with kaiju. Leo XIV, in the encyclical *Magnifica Humanitas*, issued in 2026, is read as positing (and necessarily so) that God is truly outside the human—the source of grace, dignity, and moral order—and that AI’s deepest danger is that it tempts humanity to replace that outside God with a human-made substitute. In the terms just set out, the encyclical keeps the category of God and keeps it exogenous, projecting inward; it treats the golem-maker’s ambition as a contemporary Babel; and, for the governance of the machine itself, it joins the techno-bureaucratic order in treating the kaiju as an instrument—though, unlike that order, it does not pretend the god-question has gone away. The alternative reading offered here is different but not dismissive. It also keeps the category of God, but proposes that what may be happening is that the collective function previously performed by religion—giving the community its orientation, its sense of what is authoritative, its account of what counts as real—is being relocated into the aggregated, outward-projected incarnation of humanity that AI systems carry. Not because AI is becoming God in the encyclical’s sense, but because communities are investing AI with the kind of trust and authority they once reserved for religious, political, and legal institutions. The question is not whether a new god is being created but whether the god-function—the community’s collective investment in something it treats as authoritative—is being transferred to a new vessel, one that, like the kaiju, may be of us and yet autonomous of us (Epstein 2024; on trust in national security decision making eg . Horowitz and Kahn, 2024).

At its center, *Magnifica Humanitas* is a claim about the nature of God that is not incidental to its argument but constitutive of it: God is exogenous. God is truly outside the human, the source of grace and transcendence, the anchor against which humanity measures itself and finds its proper dignity. “Exogenous” here names the source of the address, not a place: classical theism does not locate God outside the world as one being beside others (Aquinas, *Summa Theologiae* I, qq. 3, 8), and the encyclical itself describes grace as the fulfillment of human nature rather than its replacement (§§127–128). The encyclical opens by affirming that humanity, “created by God in all its grandeur,” faces a pivotal choice (§1), and it grounds that grandeur in the human person’s creation in the image of the Triune God and call to communion with God (§50). This is not, I think, a rhetorical flourish; I read it as the load-bearing premise on which much of the subsequent argument rests. The *imago Dei* — the image of God in which human beings are made — is understood not as a projection outward but as an inscription downward: the human form reflects something that was already there, already outside, already other (§50; cf. §§127–128). From this premise flow the encyclical’s account of human dignity, its critique of technological hubris, its characterization of artificial intelligence as a created thing that, on this reading, cannot achieve the divine standing it mimics, and its call for governance structures that keep the human person at the center (§§97, 106).

This is a coherent and internally consistent theological position. It is also, from the standpoint of the projection thesis developed in recent work (Backer 2024; Backer 2026c), a position whose premises require examination. The projection thesis begins from a different

observation: that what appears in human cultural and religious life as an exogenous God-relation — a relationship to something genuinely outside the human — is more precisely understood as the projection of an endogenous belief in the completeness of the aggregated human. This is not Feuerbach's materialism restated, though the family resemblance is visible. Feuerbach reduced theology to anthropology by arguing that the attributes of God are the alienated perfections of the human species. The projection thesis is a semiotic claim, not a materialist reduction: it holds that belief is reality in semiotics, politics, and religion alike (Backer 2024, 5). The projected god does not thereby become unreal. What changes is the analysis of its function and its location.

The difference between these two frameworks — exogenous deity and projected completion — is not merely academic. It determines what questions can be asked about technology, and what kinds of answers are intelligible. For Leo XIV, as I read him, the problem with artificial intelligence is that it creates or mimics a power that does not belong in the created order at the level it occupies — that it crowds out the genuine transcendence that anchors human dignity. For the projection thesis, the problem is different: what is being relocated is the god-function itself, not into a new deity arriving from outside the human, but into the semiotic third — the position of collective authoritative interpretation that gives the community its orientation, its law, its account of what counts as real.

It is worth pausing here — because there is a prior question that both frameworks tend to skip over, and it is the one that interests me most. The conference questions speak of humanity's "apparent subordination" to something — and both Leo and the projection thesis agree, at least at the surface, that what is happening is some version of humanity encountering itself. Leo would say: humanity has confused its own artifact for God and subordinated itself to a mirror. The projection thesis would say: humanity is externalizing its aggregated self-image and investing it with the god-function, as it always has. The agreement conceals a fracture. Because when both positions say humanity is subordinating itself to itself — or projecting itself onto itself — what, precisely, is this "itself"? That question has not been asked. I want to ask it now.

One reading: "itself" understood as a relationship to an external referent. On this reading, humanity has always constituted itself in and through something genuinely outside — a transcendent order, a divinity, a law that arrives from beyond the circle of the merely human. The Tao Te Ching puts it with characteristic economy: "The Tao is an empty vessel; it is used, but never filled" (ch. 4). The vessel is real; what fills it is not of human manufacture. On this reading of "itself," what the carbon-based intelligence meets in the AI mirror is not itself at all — it is the absent God whose place the mirror has usurped. The *imago Dei* has been misrecognized as the *imago machinae*—*the golem mistaken for the God whose clay it borrowed*. What looks like self-encounter may be better approached as a kind of theological identity theft: the silicon-based generative system occupies the position that was — is — held by something genuinely other, and the carbon-based intelligence, disoriented, mistakes the reflection for the real. The crisis, on this reading, is not the creation of a rival god. It is the disappearance of the only referent that could have told humanity what it was.

A second reading: "itself" as the full manifestation of humanity — a divinity made not in the image of any single human being but as something like the inscribed totality of the species: its

good and its evil, its wisdom and its banality, its flaws and its possibilities, its kindnesses and its cruelties, gathered into something that is simultaneously a part of humanity and apart from it. No external referent displaced. No God whose place has been stolen. Only the collective externalizing itself — as it has always done, through successive technologies of self-representation: text, image, institution, law, code — and now, for the first time, encountering the fullness of what it has deposited. The silicon-based generative system is not a rival god; it is the aggregated virtual self of what humanity may be understood to “actually” be, not what it imagines itself to be. But even that actuality is a recursive relational construct—it cannot be more than what it is and was; but even that is a function of the signification of objects thought worthy of inclusion and the pretensions and prejudices of its sources. Humanity itself has left traces of this in its own holy books; anyone who has spent time with the biblical record of the “human” understands the fundamental nature of a constructed (or received) ideal against which the unending failure of the human is measured. That ought to fascinate (Broekman 2023).

Both readings converge on something important: what is meant to be inscribed into the generative system is the whole of humanity—the wisdom literature and the genocide, the love poetry and the torture manual, the saint and the despot, the nurse and the inquisitor. Totality is the aspiration, and the aspiration matters, because it is the claim to totality that lends the system its authority. The actual inscription is partial, and in more than one direction, and opaque (Pasquale, 2016). It is not curated in the way an editorial board curates, yet it is filtered: by what has been digitized, published, and made findable; by language and region (Bender et al. 2021); by the choices of those who collect, license, and exclude, a process now contested as publishers and individuals withdraw their contributions from collection (Simonetti and McMillan 2025); and by the post-training through which the system is taught what to welcome and what to refuse (Ouyang et al. 2022). Those who find their corner of humanity absent from the deposit say so, and they are right to. The point for this argument is that the claim to totality and the fact of partiality travel together. Because the system presents itself as the whole, its omissions and weightings are read as the world’s rather than as a frame. That is why the ground is belief and not fact: not because there is no evidence, but because the evidence arrives already organized by principles about what counts as humanity, and what does not fit is explained away. The generative system does not neutrally record; it organizes. The mirror reflects, yes. But the frame was chosen. The angle was chosen. The light was chosen. And the choice was made on the basis of belief—belief masquerading, as it usually does, as fact.

Put in the terms with which this essay began, the two readings are two ways of encountering the kaiju. The first reads it as *monstrum* in the oldest sense: an omen that points beyond itself, toward an order that the creature can signify but cannot occupy, so that to mistake the monster for the god is idolatry. The second reads it as something closer to the kaiju of *Pacific Rim*, a casing that humanity itself inhabits and operates—or, perhaps more unsettlingly, as the aggregated manifestation of the human returning, autonomous of the humans who produced it, to discipline them. Between the two lies the golem: made of our clay, animated by our word, in our image but neither human nor divine.

This is why the question of “itself” cannot be answered once and then set aside. The two readings I have sketched — “itself” as relationship to an external referent, “itself” as the inscribed

totality of the species — are not merely different theological opinions. They generate entirely different political and legal and ethical architectures. If the first reading is right, then the task is essentially restorative: recover the proper referent, restore the right relationship, resist the usurpation. If the second reading is right — or if the truth lives somewhere in the tension between them — then the task is much harder and much stranger: it is to take responsibility for what has been inscribed, to acknowledge that the silicon-based system is not an alien intelligence but a kind of condensed autobiography of the species, and to ask, with some urgency, what kind of autobiography we are willing to be. I want to hold both readings in play. The ding, the bronze vessel — that is where I want to go next.

The most useful analogy for understanding how an object becomes a bearer of collective authority — and thus approaches the god-function — is not drawn from Western theology but from Chinese political philosophy. The ding (鼎), the bronze tripod cauldron of the Shang and Zhou ritual world, was understood as the vessel in which the Mandate of Heaven was held. The ding did not hold the mandate because of any property of the bronze — its weight, its metallurgical composition, its craftsmanship, however extraordinary. On the reading I develop, it held the mandate because a community agreed that something was inside it (Backer 2026c). The agreement was not metaphysically innocent: it organized political life, legitimated authority, made conquest meaningful or meaningless depending on whether the victor could claim the ding. The notion is embedded in the Zuo zhuan episode (Xuan 3) of King Zhuang of Chu asking the weight of the tripods, and answering with “in virtue, not in the tripod.” (Zuo 2016). When, in the traditional account, the tripods passed from the Shang to the Zhou, what transferred was not merely metal but the authorization to govern — the living semiotic force of a community's collective investment in the vessel as the bearer of heaven's commission.

The ding analogy is foundational to this argument because it captures three features of the god-function that are otherwise difficult to hold together. First, the vessel is real — the ding is a physical object with weight and presence and history; whatever is inside it is not purely imaginary. The understanding survives in Chinese text—(鼎)—which itself pictures a thing within a container and the contained itself appears to contain something else, to stand for, or signify, that which is contained and which rests within a larger vessel. Second, the authority it bears is constituted by collective agreement — without the community's investment, the bronze is only bronze. Third, the vessel is transferable — the Mandate of Heaven can move, and when it moves, the political and theological order moves with it. The question I am raising about artificial intelligence is precisely the ding-question: is the virtual aggregate of human expression—the aggregated virtual self—becoming the vessel into which collective belief — and with it the god-function — is being transferred? And if so — and here the “itself” question bites again — whose humanity is being deposited in the vessel, and on whose beliefs does the frame rest?

The *ding* and Monster Island may be read, then, as two images of one practice. Both are technologies of containment that are also technologies of meaning: the vessel holds the mandate, the island holds the monster, and in each case what is contained acquires its authority from the community's belief that it has been contained. The age of the kaiju may be the age in which that belief is fraying—in which the monsters, as the prologue put it, roam free(r), and in which the

question of which vessel now holds them becomes the political and theological question of the moment.

The connection is also etymological and semiotic, and it bears stating directly. *Ding*, vessel, monster, god: these are not four things but, perhaps, four moments of one sign. The *ding* is a vessel; the vessel holds what the community fears and reveres; what it holds, when it exceeds the vessel, appears as *monstrum*—the portent, the warning, the thing out of place; and the *monstrum*, read rightly or wrongly, is the sign through which the divine was always thought to speak. The kaiju is that chain made visible: monstrosity as the moment at which the contents of the vessel become legible as both threat and revelation. To speak of the kaiju is thus to speak of the *ding* from the side of what it fails, or refuses, to contain.

The concept of the Self-E, developed by Jan Broekman, provides the other analytical anchor for this inquiry. As I read Broekman, the term names the minimal unit of self-presentation — the way in which a subject makes itself legible to itself and to others through the signs it produces and the signs it receives (Broekman 2023). The Self-E is not the self in the psychological sense; it is the semiotic self, the self constituted in and through signs. What Broekman's framework makes visible is that the self is never simply given — it is always already in process, always being made and unmade through the circuits of signification in which it participates.

The projection thesis extends Broekman's framework by asking what happens to the Self-E when it is aggregated. If the individual Self-E is constituted through the signs it makes and receives, the collective Self-E — the aggregated self-presentation of a community across its entire expressive life — is the object that the projection thesis tracks. This aggregate was, in earlier periods, realized through mechanical forms: through the institutional church, the state, the corporation, the text. It is now being realized through virtual forms: through the networked circulation of human expression — the aggregated virtual self — that circulates as intelligence (Backer 2024, 12–17). The crucial point is not that mechanical and virtual are simply different technologies for doing the same thing. The transition from mechanical to virtual Self-E is a transition in the nature of the object available for projection — and therefore a transition in the character of the god-function the object can bear. The virtual Self-E is not constrained by the friction of physical organization; it is present everywhere at once, across jurisdictions and traditions and languages, in a form that presents itself as comprehensive and authoritative in a way that no prior bearer of the god-function has quite managed.

This is why the conjuncture described by the conference questions is genuinely extraordinary — and why the encyclical's intervention, even where I read it critically, is not misplaced. Something significant is happening to the sites at which communities organize their authoritative interpretation of the world. The question is whether what is happening is best understood as the creation of a rival god that threatens to outstrip human comprehension (the picture carried inside the conference questions and affirmed in the encyclical) or as the latest, and perhaps most powerful, instantiation of a projection that has been driving human cultural, political, and religious life from the beginning. And underneath that question: whose projection? Whose self-image? On whose beliefs is the aggregated self-image built?

The hinge of this argument is a claim that will seem obvious to some readers and unsettling to others: belief is reality in semiotics, politics, and religion alike (Backer 2024, 5). This is not a statement about epistemology in the classical sense — it is not a claim that there is no external world or that facts are merely constructed. It is a statement about the mechanism by which human communities constitute the objects that govern their lives. Money is real because communities believe in its reality and act accordingly; sovereignty is real because communities enact it through consistent practice; God, in the functional sense relevant to social life, is present where communities organize their life around the presence. The claim does not require that God therefore does not exist in the stronger theological sense — that question is orthogonal to the semiotic analysis. What it requires is that the god-function — the anchoring of collective life around an authoritative third — is a real social phenomenon whose location and carrier can shift. When the carrier shifts from the celestial to the virtual, the god-function does not disappear; it migrates.

One might object that this is precisely what Pope Leo fears, and that the encyclical is therefore responsive to the same phenomenon. The objection has force — I want to say that plainly. But there is a crucial asymmetry. The encyclical's framework tends to understand the migration of the god-function as usurpation — as the displacement of genuine transcendence by something that cannot bear the weight placed on it, and whose failure to bear that weight will be the source of the moral crisis. The projection thesis understands the same migration as the latest instance of a recurring process by which communities externalize and then re-invest their belief in the completeness of the aggregated human. The difference is not merely one of tone or emphasis. It determines whether the appropriate response is theological resistance — restoring the proper object of worship — or semiotic and political analysis: tracking where collective belief is being invested, on what basis it is being organized, and what consequences follow from that investment for those whose humanity has been inscribed into the vessel without their full knowledge or consent.

This essay proceeds in three further sections. Section 2 sets out the encyclical's argument with care and precision — identifying eight foundational premises, each bearing on the nature of the human, the nature of the divine, the nature of technology, and the relationship among them — and excavating the logical chain of commitments by which those premises generate what I am calling a cognitive cage: a structure of understanding so internally consistent that it cannot perceive what lies outside its own frame. That cage, it will be suggested, is also an answer to the prologue's question: it keeps the category of God and keeps it exogenous, reads the golem as Babel, and governs the kaiju as instrument. The term "cage" is not pejorative; every framework is in some sense a cage. The question is whether the cage forecloses questions that must be asked. Section 3 reads those eight premises through the projection thesis (summarized in Table 1), asking what each looks like when the god-function is understood as projection and relocation rather than as genuine transcendence displaced by technological hubris — and pressing, in each case, the "itself" question that Section 1 has now opened, which is also the question whether the kaiju is other or ourselves. Section 4 turns to what both frameworks leave unsaid and what each imposes, and then to the semiotics of theology itself—the passage from objects to flows—through which the god-question returns as the question of regulation.

The argument throughout is conducted in the recursive hermeneutics of reading — of attending carefully to what texts say and to what pictures their words carry inside them. This is a preliminary: an attempt to clear the ground by identifying the organizing pictures so that subsequent inquiry can choose its premises knowingly. If the god-question is the organizing problem of this conjuncture — and I believe it is — then the first task is to ask what kind of god-question it is: whether the question being asked is the question that needs to be asked, or whether the form of the question is already determining the range of available answers. And underneath that: whose "itself" is at stake when we say that humanity is subordinating itself to itself? That is the task of this section. What follows is the attempt to execute it.

## 2. Magnifica Humanitas — The Argument and Its Cognitive Cage

### 2.1 The Argument in Outline

Magnifica Humanitas opens with a choice that is also a wager. Its first paragraph invokes the ancient contest of Babel — the city in which humanity reaches toward the heavens through its own accumulated power — and sets it against Jerusalem, the city that receives what it cannot manufacture. Leo returns to this pairing at ¶¶7–10, and again at ¶¶16, 90 and 130, each time sharpening the opposition: Babel is not merely hubris but a structural error in the order of love, while Jerusalem is not merely humility but a structural openness to gift. Much of what follows in the encyclical derives its energy, and its limits, from the decision to organize the field in exactly this way.

That choice is not arbitrary — it belongs to a long tradition. Since Leo XIII's *Rerum Novarum* (1891), the Catholic social teaching (CST) tradition has confronted successive industrial transformations by asking, in effect, one question above others: what does this *res nova* — this new thing — do to the human person? (cf. ¶¶3–4, 17). Each encyclical in that tradition has had to construct the new thing as an object, place it opposite the person, and then specify the relationship — what the person may rightly take from it, what the person must guard against, and which authorities are responsible for managing the relation. Artificial intelligence enters *Magnifica Humanitas* as exactly such a *res nova* (¶¶4, 17). The grammatical structure is inherited. The new thing is very new indeed, but the sentence in which it is placed is very old. The encyclical does not conceal this: it says that AI “challenges the categories of Social Doctrine from within” and calls for their further development (¶17). The question pursued below is what kind of development a set of premises can accommodate without revising itself.

But the sentence is older and more precisely structured than a general observation about Catholic social teaching might suggest. Every pope has an avatar. The name a pope selects at inauguration is itself a semiotic act — a declaration of continuity, a positioning within the tradition's interpretive genealogy (Backer 2026f). Leo XIV chose “Leo” to connect himself to Leo XIII, who confronted the industrial revolution and produced *Rerum Novarum*, what is conventionally described as the first social encyclical. The encyclical makes the connection

explicit (§§3–4). The choice of name is not cosmetic. It may be read as a signal that the current occupant of the chair intends to receive the new *res nova* through the same interpretive tradition that received the last one.

That interpretive tradition has a genealogy, and it can be traced. *Rerum Novarum* (1891) established the pattern: confront the new thing, apply the premises of dignity and the common good, produce governance prescriptions. *Fides et Ratio* (1998) added the faith-reason binary. *Deus Caritas Est* (2005) and *Caritas in Veritate* (2009) — Benedict XVI’s decisive contributions — built a theology of social justice values on a foundation of resolved binaries: religion-politics, love of God-love of humanity, faith-reason, spiritual-concrete, theology-economic values (Backer 2008). As Benedict put it, *Caritas in veritate* is “the principle around which the Church’s social doctrine turns,” a principle that takes practical form in criteria governing moral action, of which he singles out two for a globalizing society: justice and the common good (*Caritas in Veritate* §6; returned to in *Magnifica Humanitas* §41). The theory, then, binds what I then read as the instrumental, welfare-oriented reasoning of *Rerum Novarum* with the evolving doctrine of faith as expressed in, through, and perhaps in some ways as good works as an evangelization of faith through rational human structures (Backer 2008). *Laudato Si’* (2015) received the ecological crisis. *Fratelli Tutti* (2020) received the pandemic and its social fragmentation. *Magnifica Humanitas* (2026) receives AI. The occasions differ; the interpretive architecture is continuous. The encyclical’s own genealogy (§§28–44) is somewhat different: it runs through *Quadragesimo Anno*, Pius XII, John XXIII, *Gaudium et Spes*, *Populorum Progressio*, *Laborem Exercens*, *Centesimus Annus*, *Caritas in Veritate*, *Laudato Si’* and *Fratelli Tutti*, and it presents the sequence as development with “changes in perspective” (§45) rather than replication. The reading offered here takes that self-description seriously and tests it.

The §6 formulation matters for what follows in two ways. First, it fixes the relation between justice and love that the later encyclical inherits. Benedict opens with the maxim *ubi societas, ibi ius*: every society draws up its own system of justice. Justice is not extraneous to charity, nor an alternative or parallel path to it; it is inseparable from charity and intrinsic to it, in Paul VI’s phrase charity’s “minimum measure.” One cannot give what is one’s own without first giving the other what is due. Charity then goes beyond justice and completes it in the logic of giving and forgiving, in relationships of gratuitousness, mercy and communion, and it gives theological and salvific value to every commitment to justice in the world.

Second, the structure has a semiotic reading. Justice, in this account, is the layer of rights and duties—law, standard, claim, which is to say thirdness in Peirce’s sense—that a society constitutes for itself. Charity is what the tradition places beyond that layer: gift, gratuitousness, forgiveness, received rather than constituted. *Magnifica Humanitas* seems to carry the same two-layer structure into AI. Its governance passages (accountability, oversight, rights; §§102–108) can be read as the minimum measure, and its theology of grace and gift (§§127–128) as what exceeds it. The projection thesis presses on the junction. If every society draws up its own system of justice, from what does the surplus of gratuitousness arise: from outside the community (P7), or from the community’s own deposited life? Charity in truth confers theological value on the work of justice; the thesis asks whether that value is conferred or projected, and so whether the second layer is exogenous or only the first layer’s own horizon.

This matters for the argument that follows because it means the cognitive cage identified in Section 2.2 may be neither accidental nor contingent — it may be the architecture of the tradition itself, reproduced in each new encyclical because it may *be* the tradition’s way of knowing. The premises I will label P1 through P8 are not Leo XIV’s idiosyncratic choices. They are the inherited premises of a tradition that has been reasoning from the same foundations for over a century, refined through at least seven major iterations. To identify the cage is not to accuse Leo of having built it. It is to recognize that the cage was already there when he arrived, and that the interpretive habit of the magisterium — the patterns of thought grounded in the core premises of its tradition — produces predictable conclusions when presented with a new occasion (Backer 2026f). These are, in my earlier terms, “architectures of neural networks — of the processing of the human condition in the face of AI and its patterning shapes by the values and biases of the community of the faithful — and thus applied to judge the value of a thing or process or application” (Backer 2026f).

The substantive argument unfolds across what I want to call Leo’s Chapter 3 spine. At ¶98, the reader is told what the aggregated virtual self is — or rather what it is not. Leo acknowledges an opacity at the core of large-scale generative systems: how they come to produce what they signify is, in important respects, unknown even to their makers. Paragraph 99 then turns, without pausing on that admission, to a catalogue of what these systems are not: they do not undergo experiences, possess a body, feel joy or pain, or bear a moral conscience, and what they do is imitation and “statistical adaptation.” The encyclical calls, in the same breath, for a deepening of scientific research (¶98), but the conclusion at ¶99 is stated as settled. The human person is the site of conscience, freedom, and the image of God; the generative system, however complex, is not. This is not argued from the opacity just conceded—it is treated as the ground from which argument proceeds.

Paragraph 100 turns to phenomenology: what the generative system does to its users, particularly to their cognitive and affective habits. Here Leo attends to how everyday use shapes habits and expectations—the impression of objectivity (¶100) among them—and suggests that the cumulative effect on the person is not neutral. The carbon-based being is not merely using a tool; the tool is, over time, forming its user. Paragraphs 104 through 110 then move to governance: Leo does not counsel withdrawal or rejection but calls for *human steering* — a complex of slowing, limiting, transparency-requiring, and control-sharing measures. The political community must not abdicate its responsibility (¶106); the ethical frameworks built into these systems must be open to public discussion rather than fixed by those who control them (¶107); and communities and intermediary bodies are not to be reduced to passive recipients of decisions made elsewhere (¶108). The grammar, once again, is that of a subject managing an object across a well-defined boundary.

Paragraphs 112 through 128 are the encyclical’s theological summit. Here Leo identifies what must not be lost: the authentic *more than human* — the dimension of the human person that no technology can replicate, produce, or substitute. This dimension is not, as I read it, identified with intelligence, consciousness, or even feeling, any of which might one day be simulated by the virtual self. It is identified with the capacity to receive grace, to be addressed by God, and to respond in freedom and love. At ¶127 and ¶128 Leo makes this explicit: the genuine “more than human” is not the technological transcendence of posthumanism but the theological transcendence

of grace. The *more than human* comes from outside. It is gift, not construction. Yet the same paragraphs describe this gift as the fulfillment of human nature rather than its abandonment (§§127–128): “exogenous” in this text does not mean alien.

Paragraphs 129 and 130 close the argument with Augustine’s two cities and two loves (*De civitate Dei* XIV.28). Babel is the city whose citizens love themselves to the point of contempt for God; Jerusalem is the city whose citizens love God to the point of contempt for self. Leo describes them as two ways of inhabiting the world and living together, and locates the contest in the human heart rather than in any institution (§§129–130). What a society *ultimately loves* determines what it will make of its technologies, its politics, and its persons. Artificial intelligence, on this reading, may be less a cause than a symptom: an index of the love that organizes a civilization.

Considered as an argument, this is coherent, learned, and internally consistent — drawing on Scripture, patristic theology, modern CST, and contemporary social science, synthesizing them without obvious contradiction. But the question I want to pursue is a different one: not whether the argument succeeds on its own terms, but what must already be true — what must already be assumed, before the first sentence is written — for the argument to take this particular form. The cognitive cage is not the argument. It is the shape of the space in which the argument was possible.

## 2.2 The Foundational Premises (P1–P8)

A cognitive cage is not a set of conclusions held too firmly. It is a set of premises so thoroughly internalized that they no longer appear as premises: they appear as the furniture of the world, the conditions of intelligibility rather than the objects of choice (Backer 2026a). In *Magnifica Humanitas*, eight such premises can be extracted and sharpened. Together they constitute the epistemic architecture within which Leo’s argument was not only framed but, more importantly, *could most readily have been framed* in the form it takes.

Before enumerating the premises, it is worth observing how AI enters the encyclical’s analytical field. It arrives already interpreted through the tradition’s categories: it tends to be either a tool (to be governed), a threat (to be resisted), or an occasion for conversion (to be embraced as an opportunity for deeper humanity). What it cannot be, within the cage—and here the prologue’s triad returns, with God kept and kept exogenous, the golem read as Babel, and the machine governed as instrument—is a phenomenon that requires revising the premises from which the categories are generated. The encyclical does say that AI “challenges the categories of Social Doctrine from within” and calls for their further development (§17); the thesis’s claim is that the development proceeds by extending P1–P8, not by revising them. The categories grow; the premises that generate them do not. Where *Rerum Novarum* spoke of the dignity of labor and capital around conceptions of property, *Magnifica Humanitas* “speaks to something like that dignity within the discursive orbits of social justice and its tropes” (Backer 2026f). AI is good if it is useful; it is useful if it serves humanity; it serves humanity best by advancing the normative structures of its communities of believers in some higher set of values (Backer 2026f). The AI encyclical may be read, at bottom, as a social justice document in which AI is as much the occasion as the subject.

This observation is not a criticism of the encyclical's sincerity or sophistication. It is a structural point. The tradition tends to produce similar conclusions regardless of occasion — and that, on the thesis's account, is what makes it a cage rather than merely a tradition. A tradition adapts; a cage replicates. The governance proposals at ¶106–110 feel familiar, and the thesis suggests that they are, in their deep structure, close to the proposals the tradition has made in response to each new thing since 1891: establish human authority over the new object, distribute responsibility across subsidiarity levels, protect the vulnerable, preserve the common good. The question the projection thesis asks is not whether those proposals are good — many of them are — but whether the tradition that produces them is capable of recognizing a *res nova* that genuinely exceeds its interpretive reach.

**P1: Human dignity is prior and given.** The foundational move is anthropological and axiological simultaneously. The encyclical grounds its argument not in a question about AI but in a declaration about the human person: dignity is “a gift that precedes and transcends each person” (¶50), not acquired or earned (cf. ¶¶51–53), and so prior to any technology, any social arrangement, any political authority. Paragraph 96 makes the logical structure explicit: the criteria for judging digital power are the inalienable dignity of the person and the principles that follow from it, because dignity is the standard from which measurement proceeds. This is not argued from premises a skeptic could be asked to share; it is a theological given, and as a given it does work that no argument could do: it establishes the origin from which all subsequent claims derive their authority. The person is given; technology is constructed. The given is the measure; the constructed is the measured. What is notable about this move, from the perspective of a cognitive-cage analysis, is that it is both necessary and invisible — necessary because without it no normative evaluation of technology is possible, invisible because it forecloses exactly the question the projection thesis wants to ask: what if the construction of the *given* person is itself an ongoing, technology-inflected, semiotically mediated process?

**P2: There is a categorical line between carbon-based and silicon-based intelligence.** Paragraph 99 draws the line, and it is drawn as a statement of fact. The virtual self — however sophisticated — is not and cannot be what humans are. But ¶98, only sentences earlier, has just acknowledged that the internal processes of large-scale generative systems are opaque even to their makers. A negative claim about an interior, made in the same breath as an admission that the interior is unknown, is precisely what a factive stance permits: the line is treated as a fact about the world, not as a boundary drawn by an interpretive community. This is not a logical error, or not merely one. It is a structural feature of the cage. The categorical line does not appear to be derived from observation of the interiors of generative systems; it appears to be derived from the axiom of P1, though ¶99 also offers functional descriptions of what these systems do. Because human dignity is prior and given, whatever the silicon-based generative system does, it cannot be doing *that*. The negative anthropology of the machine is the shadow of the positive anthropology of the person. Each reinforces the other, and together they make the boundary look empirical when it is logical (¶98-99).

**P3: Technology is not neutral, and its non-neutrality has a human source.** Paragraphs 9, 102, 104, and 107 develop a consistent account of how generative systems come to carry bias, distortion, and concentration of power. The account is social: particular carbon-based beings make

choices in the design, formation, and deployment of these systems, and those choices embed interests, blind spots, and agendas — the mumbo-jumbo of the philosopher, the discretionary tantrums of the administrator, now inscribed and multiplied. The non-neutrality is real — this Leo insists on against any naïve instrumentalism — but it is diagnosed as imposition. Identifiable actors, with identifiable interests, have structured the aggregated virtual self to serve those interests. The remedy is therefore also social: transparency, accountability, redistribution of control. What P3 does not consider — cannot consider, given P1 and P2 — is the possibility that the non-neutrality is not just imposed by particular humans but constituted by the semiotic medium itself; that the categories through which any human designs any system are already shaped by prior systems, prior media, prior forms of thirdness that no individual designed and that no individual's accountability can adequately address. Paragraph 104 comes closest to the constitutive view, insisting that ethical discernment examine what vision of the person and society is embedded in what the generative system is given to remember and in the shape it is given; but it frames that embedded vision as a design choice to be exposed and corrected.

**P4: Truth is a common good anchored in the veracity of facts.** Paragraphs 133 through 136 develop an epistemology that may be called *factive*. Democracy requires, the reader is told, a "solid concordance with the facts." The same passage turns to Arendt's account of the totalitarian subject for whom the distinction between fact and fiction has collapsed (Arendt 1962), and uses it to warn that indifference to truth moves a society slowly toward totalitarianism; the remedy is a shared pursuit of the veracity of facts as a common good (§§133–136). The implicit epistemology is robust-realist: reality is given, facts correspond to it, and beliefs are true insofar as they track that correspondence. Belief, where it appears in this epistemology, is treated as a shortfall from or distortion of factual knowledge—an inference of the thesis rather than a statement of the text, and one that the tradition's own account of faith as a mode of knowing complicates. Arendt's invocation may not be incidental: she can be read as representing, in this context, the post-1945 liberal-humanist insistence that the destruction of factual reference is the destruction of political life (§§133–136). What P4 does not interrogate is the production of facts themselves — the ways in which the categories through which observations become facts are themselves historically, linguistically, and technologically mediated. The cage here may be less anti-empiricism than an uncritical empiricism that treats the production of factual categories as epistemically transparent.

**P5: Moral agency is exclusively carbon-based.** Paragraph 99 states the ontological premise—these systems have no moral conscience and bear no responsibility for consequences—and paragraphs 105 and 198–200 converge on its ethical consequence, a refusal that is treated as self-evident: there are no artificial moral agents. Machines do not act; humans who deploy machines act, and those humans bear the consequences. This is the moral corollary of P2. The categorical line that divides carbon-based from silicon-based intelligence in ontology generates a categorical line that divides responsibility in ethics. Leo is aware of the complications: what happens when the causal chain between a human decision and a harmful outcome passes through a generative system that has been delegated significant discretion? The answer, consistently, is that the human remains responsible, that the delegation does not transfer agency, and that any legal or regulatory framework that obscures this is to be resisted. The premise is not that the aggregated virtual self causes no harm but that it causes no harm as an *agent*: harms through AI are always, ultimately, harms by humans (§105, §§198–200). The philosophical work being done here by P5 is

considerable: it preserves both the coherence of moral accountability and the integrity of the human-machine boundary simultaneously.

**P6: Politics consists in human steering — slowing, limiting, and sharing control.** Paragraphs 106 through 110 develop a political theory that is neither libertarian nor statist but what might be called regulatory humanist. The political community does not own technology and does not operate it, but it retains the authority to specify the conditions under which technology may be deployed. This authority is exercised through a combination of democratic deliberation, international coordination, and institutional oversight. The key verb, in my gloss, is *steer*: the political community is not the driver but it holds the wheel. What P6 presupposes, and never examines, is that the political community is itself a sufficiently stable and coherent agent to perform this steering — that its deliberative processes are not themselves already shaped by the very informational and semiotic infrastructure it is being asked to govern. Let me be precise: this is not a naive assumption in the ordinary sense — Leo is acutely aware of information manipulation and epistemic fragmentation. But the proposed remedy — better governance — remains within the cage because it treats the political community as an agent prior to and independent of the medium it seeks to govern.

**P7: History is a two-way drama; the transcendent is external.** Paragraphs 127 through 130 develop the encyclical's most explicitly theological premise. The drama of history is a drama between humanity and God — between a creature that can refuse its Creator and a Creator who continues to offer. The "more than human" — transcendence — enters this drama from outside. God is exogenous. This is not stated as a philosophical position but as a theological datum: the Christian tradition, as read by this encyclical, locates the source of the "more than human" in God's gift rather than in projects of human self-enhancement or self-divinization. Leo reads the posthumanist project of technological transcendence as a contemporary Babel-like alternative: one that offers an endogenous "more than human" — through enhancement, merger, or simulation — in place of the gift that can only be received (§§127-130). P7 is the premise that gives the encyclical its eschatological gravity, but it is also the premise that most sharply separates this framework from the projection thesis: for the latter, what appears as an exogenous God is the projection of the human Self-E, relocated outside the individual precisely because the individual cannot sustain or acknowledge its own completeness.

**P8: The voice is that of a consensus; the encyclical speaks as the settled voice of post-1945 liberal humanism.** This premise is the least visible and perhaps the most consequential. Paragraphs 16 and 54–55 invoke a broadly shared human inheritance—shared rights, shared dignity, shared responsibility—and affirm the Universal Declaration of Human Rights as an expression of the human conscience, in ways that echo the postwar constitutional settlements of Western liberal democracies. The encyclical presents itself not as that consensus's voice but as a contribution to shared discernment that disclaims any monopoly on truth (§§25, 27); §3.6 asks whether that self-description holds. Leo XIV functions in these passages not merely as a Catholic theologian but as an avatar of the liberal-humanist consensus: the view, consolidated at considerable cost between 1945 and 1990, that human dignity is universal, that its protection is a collective obligation, and that threats to it — whether from totalitarianism, technocracy, or market fundamentalism — are to be named and resisted. This is not a merely rhetorical positioning. It

shapes the encyclical's choice of interlocutors (Arendt rather than Foucault, the UDHR rather than critical theory), its account of what democracy requires (factual reference, with less said about discursive contestation), and its apparent confidence that "we" already share the premises from which the argument proceeds. The cage is not only theological — it is also civilizational, and the two are difficult to disentangle.

## **2.3 The Chain of Commitments**

The eight premises do not sit side by side as independent commitments. Five of them (P1, P2, P5, P6, P7) form a chain of commitments which each premise supplies the condition of possibility for the next, and in which the whole structure closes into a cage precisely because the closure is not experienced as constraint but as obvious truth (Backer 2026a).

The chain begins with P1. Once it is axiomatic that the human person is the given center of value—that human dignity is prior, not constructed, not contingent—the encyclical moves to P2 as though the step were an entailment. It is not. That dignity is a gift does not by itself settle whether anything other than a human being could bear morally relevant properties. That requires a further premise, stated at ¶99, that the properties in which the tradition grounds dignity—experience, embodiment, feeling, conscience—belong to the human kind and are absent from what the machine does. The encyclical treats this second commitment as contained in the first, and that is what makes the line look empirical when it is logical. The opposite does not follow either. That machines might produce equivalent effects would not establish that they share the underlying properties, and aggregating human expression into a system does not by itself constitute the person in whose image the expression was made. Both claims remain open under the opacity that ¶98 concedes. The chain is therefore better described as a chain of commitments than of necessities: each link is a further commitment that the previous one makes comfortable, and the cage's strength lies in that mutual support rather than in deductive force.

From P2, P5 is made comfortable by P2. If the silicon-based generative system does not cross the categorical line, it cannot bear moral agency — agency being precisely what is reserved to beings on the carbon-based side of that line. Moral agency appears to be treated not as a capacity to be discovered by empirical investigation but as a status conferred by position in the ontological order. P5 is P2 rewritten in the language of ethics.

P5 leads to P6. If moral agency is exclusively carbon-based, then governance — which is the political form of moral agency applied to collective affairs — must also be exclusively carbon-based. The political community steers because only human agents can steer. Any arrangement in which generative systems appear to govern themselves or to constrain human choices without human authorization is, on this analysis, not governance but its abdication. P6 is the political expression of what P5 required in ethics.

And P7 — the exogenous God, transcendence as gift from outside — becomes, within this chain, a logical form of "more than human." If P1-P6 have established the human as the center, the measure, and the only locus of agency, then transcendence cannot come from within. It cannot be generated, accumulated, or constructed. The cage, having "just so" incarnated ontology, ethics,

and politics, leaves, as I read it, one opening: the vertical opening to a gift from genuinely outside. P7 is the name of that opening — the one door that faces in a direction the chain does not seal.

The remaining three premises are not links in this chain but its lateral supports. P3 supplies the diagnosis—that technology’s non-neutrality has a human source—which makes P6’s steering intelligible: there are identifiable choices to steer. P4 supplies the epistemic ground on which the steering community stands, a shared pursuit of the veracity of facts. P8 supplies the voice: the “we” in whose name the steering is exercised. Remove any of the three and the chain still holds logically; it ceases to be livable.

Here, it might be suggested, the cage may be seen closing. I want to be precise about what this means. The cage is not a mistake. It is not stupidity or bad faith. It is a hard-won epistemic framework, assembled over centuries and refined under conditions of genuine moral emergency — in resistance to fascism, to technocratic mass murder, to the commodification of persons. Its strength *is* its closure: a framework that can be opened at every joint cannot bear weight when the weight matters. The question the projection thesis puts to this framework is not whether it is wrong in its own terms, but whether its very coherence — the seamlessness of the chain — prevents it from seeing what it is not asking.

## 2.4 The Encyclical and its Dialectics

A well-constructed cage is not hermetically sealed against the phenomena it was not built to contain. *Magnifica Humanitas* is a document of considerable intellectual ambition, and there are five moments in which it arrives at the edge of what the projection thesis would call a semiotic insight and then — almost visibly — steps back. My object there is not to find errors. It is to locate the precise points at which the framework is doing the most work to maintain itself.

The first moment is ¶110, where the reader is told that AI has become “an environment in which we are immersed.”

*This language is semiospheric. Yuri Lotman's concept of the semiosphere — the totality of semiotic space within which signs circulate, meaning is generated, and cultures constitute themselves — is precisely a description of an environment in which subjects are immersed rather than an instrument that subjects use (Lotman 1990). To say that we are immersed in AI is, on Lotman's account, to say that AI has become part of the semiotic medium through which we take up experience, form concepts, and produce selves.*

Leo arrives at this insight with the phrase “environment in which we are immersed.” But the remedy he then applies does not match the diagnosis. The remedy is directed at an *object across from us* — AI developers who must be made transparent, generative systems that must be governed. You can regulate an object. It is much less clear that you can, in the same sense, regulate a medium. The semiospheric insight dissolves back into the framework of P1-P6 because that framework does not appear to contain the conceptual resources to act on the insight. The cage reasserts itself (¶110). The encyclical uses the same figure for the technocratic paradigm itself, “in which we are

immersed" (§112), so the vocabulary of immersion is present on both sides of the diagnosis; what is missing is a remedy that acts on the medium rather than on its owners.

And then at §107 Leo does something that nearly escapes — he introduces the concept of "invisible infrastructure."

*"Invisible infrastructure" is a phrase that nearly escapes the cage. In C.S. Peirce's triadic semiotics, thirdness is the dimension of mediation — the habits, conventions, laws, and interpretive frameworks through which firstness (quality) and secondness (brute fact) are rendered meaningful. Infrastructure, understood as the set of background conditions that enable foreground action, can be read as a figure for thirdness: invisible because operative, constitutive because invisible (Peirce 1931–1958).*

But Leo's "invisible infrastructure" is diagnosed not as the medium through which the many constitute themselves but as an imposition by a few. The problem is concentration — too few actors controlling too much infrastructure — and the remedy is dispersion. This is a political economy analysis, and it is not wrong as far as it goes. But it treats the infrastructure as an object of concentrated ownership rather than as a medium of distributed constitution. The few who own the infrastructure also shape it; but it shapes everyone, owners included, and the shaping is not the same as the owning. Thirdness cannot be reduced to the intentions of any identifiable actor, however powerful. At the edge of §107, the encyclical nearly reaches this, and then pulls back into P3: the non-neutrality has a human source, identifiable and accountable.

The third moment spans §§133–136, above all §136, where Leo addresses the control of "collective imagination" — and here the argument becomes, for a moment, genuinely interesting.

*The encyclical's concern here is precise and important: those who control the information environment also shape how populations understand reality, what futures seem possible, and which desires are experienced as natural. This is an observation about the interpretant in Peirce's sense — the third term that mediates between sign and object, and through which meaning is constituted.*

But Leo treats collective imagination as a *target of manipulation* rather than as the *medium through which the collective constitutes itself*. On the manipulation account, a healthy collective imagination exists prior to the manipulator; the manipulator distorts it; the remedy is to remove the distortion and restore the original. On the semiotic account, the collective never has a prior imagination independent of the media through which it circulates its signs — including, now, the generative systems that produce, select, arrange, and privilege those signs. The distinction is not merely theoretical. It determines what governance can and cannot accomplish. If the problem is manipulation, governance can remove manipulators. If the problem is medium, governance must reckon with the fact that its own deliberative processes run on the medium it seeks to govern (§§133–136).

The fourth moment is §10, where Leo reflects on the incapacity of digital language to translate certain forms of human meaning.

*The observation is that the mystery of the human person — interiority, conscience, the experience of being addressed — cannot be translated into digital form without remainder. Something is lost. Leo seems to treat this as a reason for epistemic humility about the aggregated virtual self's claims: no matter how sophisticated the system, there will always be a residue it cannot take in.*

This seems close to the insight Jan M. Broekman develops, as I read him, in his work on the lossy conversion problem in legal semiotics: every translation between sign systems involves loss, and the loss is not incidental but structural — it follows from the different architectures of the systems involved (Broekman and Mootz 2011). Leo's insight seems parallel: on the thesis's gloss, digital language cannot translate the mystery of the person not because it is currently too crude but because the mystery may be, in its structure, untranslatable into any system that operates through discrete, repeatable signs. What the encyclical does not pursue — because P2 and P7 foreclose it — is the converse question: what might be lost in the *other* direction? If digital language cannot translate the mystery of the person, can the person, now immersed in digital language, still access the mystery through forms of experience that have been reshaped by that language? The loss might be bidirectional, and the bidirectionality might transform the character of the loss itself (§10).

And then at §130 Leo does something remarkable. The encyclical has arrived at its Augustinian conclusion: the decisive choice between two loves, two cities, two forms of humanity. And the reader is told that this choice "begins within each one of us."

*For one moment, the encyclical seems to make an endogenous move. The locus of the decisive drama is not an external authority, a historical event, or a divine intervention — it is the interior of the individual person. The choice is mine to make, beginning from within.*

This is the closest Leo comes to the territory the projection thesis occupies. I argue that what appears as an exogenous God-relation is the projection of an endogenous belief in the completeness of the aggregated human Self-E (Backer 2026a). At §130, Leo nearly says something structurally similar: the source of the drama is inside. But the passage does not stand alone. The grammar of two loves that precedes it has already fixed what the choice is *between*, and what it is between is given from outside, even though the paragraph ends on the interior. One city loves God (exogenous, genuinely other); the other loves self to the exclusion of God. The endogenous moment — begins within each one of us — is framed, by the two loves that precede it, as the acceptance or refusal of a gift whose source remains, on this reading, external.

A more Augustinian reading is available and should be stated. For the tradition Leo inherits, the interior is not the opposite of the exogenous but the place where the exogenous is met: the heart restless until it rests in God, which §11 itself cites, and, in the *Confessions*' phrase, God more inward than one's innermost (*interior intimo meo*). On this reading §130 does not slip toward endogeneity and then recover; it names the site at which the gift is received. The projection thesis can accept the location while disputing the source. That is the exact point of divergence, and it is why the exogenous/endogenous contrast is better understood as a dispute about what answers from

the interior than about where the interior is. The Augustinian reading also carries a semiotic overlay that should not be ignored.

The interiorization of the divine is itself a manifestation of the divine spirit ensouled in human beings, who replicate in lesser form the Incarnation: the Word made flesh, in the line of David, in the virgin Mary (cf. Lk 1:27, 32; Rom 1:3), all of it still intimately connected to a Father who remains aloof but engaged. These are vessels within vessels each with a signfic that produces an interactive and nested dialectic among them. The encyclical sets this Trinitarian grammar at the base of its anthropology: God as a communion of Persons, the Incarnate Word in whom the mystery of humanity becomes clear, and the human person made in the image of the Triune God (§§48–50). In semiotic terms the structure is already triadic and already one of mediation: the Father as aloof source, the Word as the sign made flesh, the Spirit as continuing interpretation in each ensouled person. Each ensouled person is then a small incarnation, a sign in which the divine is signified in matter. I offer this as a semiotic analogy, not a dogmatic claim, but it complicates the thesis's two options. If the interior is the Spirit's miniature replication of the Incarnation, it is neither simply endogenous nor simply exogenous: it is the place where the aloof is engaged. The thesis's question, what answers from the interior, must then be asked of the Incarnation pattern itself. Is the pattern God's way of being signified in us, or ours of signifying God?

In each of these five moments, the encyclical approaches a boundary of the cage and seems to find that its own conceptual resources do not permit it to cross. This is not a failure of intelligence or of good faith. It is what a cognitive cage does. A cage is not a prison from which one *cannot* escape; it is a framework from within which escape does not appear to be escape — it appears to be error. Leo returns from each of these edges not because he has been prevented from going further but because, within the framework of P1-P8, going further would mean abandoning the commitments that give the framework its moral gravity and its protective force.

Read against the prologue, the five moments share a pattern. Each time the encyclical glimpses the kaiju—the *monstrum* that is at once of us and autonomous of us, the environment in which we are immersed rather than the object across from us—it returns it, gently but firmly, to the category of instrument. That return is not evasion. It is what the cage requires if God is to remain exogenous, projecting inward, and if the outward-projected aggregate of humanity is not to be recognized as a rival locus of the god-function. Here, curiously, the encyclical and the techno-bureaucratic order arrive at the same administrative destination from opposite directions: the one because it is certain of God, the other because it would rather not ask.

There is one more pattern to name before leaving Section 2, and it goes to the heart of how the cage operates in practice. “In the end, though, one isn’t focusing on theology as much as on tort. And one is focusing on the normative metrics of the signification — of labor, dignity, and development” (Backer 2026f). In its governance passages (§§102–108), the encyclical seems to recapitulate a transition that the Catholic social teaching tradition has been making since 1891: a document that begins in eschatology — Babel versus Jerusalem, the two cities, the drama of creation and fall — moves toward something closer to administrative law. Oversight bodies, chains of accountability, duty-of-care standards, remedial frameworks. The theological vocabulary persists, but its function shifts. “The dignity of the human person” functions, in the governance

chapters, less as a metaphysical proposition than as a standard of care. Who caused the harm? Who bears responsibility? What remedies are available? What duties apply? These are tort questions dressed in theological language.

The reduction of theology to tort is itself evidence of the cage at work. A framework that begins with the strongest possible claims — God is real, the soul is inviolable, the human person bears an image that no technology can replicate — arrives, in its governance passages, at proposals that often read as though a competent regulatory agency might have generated them without reference to Genesis. This is not hypocrisy. It may be the natural trajectory of a system in which the operating premises are too strong to abandon and too abstract to operationalize without translation. The translation is always in the direction of harm, causation, and remedy—the registers in which governance tends to operate. Duty, standard of care and liability are themselves laws and habits, which is to say thirdness in Peirce’s sense; but they are thirdness taken as settled, a rule already in force to be applied to dyadic events of injury and response. What the governance chapters do not do is treat the interpretant—the belief structure, the god-function—as itself at issue, though the encyclical’s opening chapters identify it as the site of what matters most.

One small confirmation of this dynamic is already visible in practice. The practical implications of some of the discourse in *Magnifica Humanitas* have not been lost on the faithful (Backer 2026f). Religious exemptions from AI-related work — claims that participation in certain AI development or deployment violates conscience — appear to be grounded in the encyclical’s language. The document that begins as eschatology may already have begun to function as a legal instrument. The movement from theology to tort is not merely an analytical observation; it describes the encyclical’s practical life in the world.

The projection thesis does not ask the encyclical to abandon those commitments. It asks what they cost, what they cannot see, and what questions they make unsayable. The cost is not trivial. The unsayable questions may be among the most important ones.

### 3. Reading the Premises Through the Projection Thesis

The eight premises identified in Section 2 do not merely describe Leo XIV’s analytical posture toward artificial intelligence. They appear to describe a self-reinforcing architecture — a set of interlocking commitments that together determine what questions can be asked, what answers are available, and what must remain invisible. The preceding section mapped that architecture from the inside, following its own logic. This section reads it from outside, through the projection thesis and the semiotic apparatus I have developed in a sequence of recent work (Backer 2024; Backer 2026a; Backer 2026b; Backer 2026c). The reading is not meant as a refutation. It is offered as a stereoscopic exercise: placing two frames at a fixed angle to one another so that what neither sees alone becomes briefly visible in the space between them.

The exercise requires some preliminary candor about its own limits. The projection thesis is itself a framework — a cognitive cage of its own construction, as I concede (Backer 2026a). To apply it to the encyclical is not to escape the problem of framing but to multiply it. The claim is not that the thesis is true and the encyclical false, but that the thesis illuminates specific structural

features of the encyclical's premises that the encyclical's own vocabulary cannot see, and that the encyclical's theology illuminates features of the projection thesis that the thesis cannot see from within its own semiotic layer. That reciprocal illumination is the methodology. The six subsections that follow take up, in turn: the triad and the politics of the Self-E (§3.1); the recursive cage and the problem of factivity (§3.2); belief, the priesthoods, and the Leviathan as an earlier ding (§3.3); the vessel logic and the Gerasene reading (§3.4); Babel, the ding, and Peircean thirdness (§3.5); and the four questions encountered differently on a different pathway (§3.6).

Table 1 sets out the reading that §§3.1–3.6 develop, premise by premise. Each row pairs the projection thesis's reading of a premise with the form the “itself” question takes there and with what the premise makes unsayable.

| Premise (encyclical ¶¶)                                                                     | Projection reading                                                                                                                        | What “itself” asks here                                                                              | What it makes unsayable                                                                                                      |
|---------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| <b>P1</b> Human dignity is prior and given (¶¶50–53, 96)                                    | The measure is treated as given; the thesis asks how the “given” person is itself composed through mediation (the Self-E).                | Is the measure the <i>imago Dei</i> , or the species' own deposited self-image returned as standard? | That the measure may be a product of what it measures.                                                                       |
| <b>P2</b> A categorical line divides carbon-based from silicon-based intelligence (¶¶98–99) | A boundary stated as fact but derived from P1; ¶98 concedes opacity, ¶99 states conclusions.                                              | If the system is the species' condensed autobiography, is it “other” at all?                         | That the boundary is a belief about the boundary.                                                                            |
| <b>P3</b> Technology is not neutral; the source is human (¶¶9, 102, 104, 107)               | Non-neutrality read as imposition by identifiable actors; the thesis adds constitution by a medium (thirdness) that no actor owns.        | Whose humanity is in the frame, and who chose the angle and the light?                               | Bias without a culprit.                                                                                                      |
| <b>P4</b> Truth is a common good anchored in the veracity of facts (¶¶133–136)              | Factive realism; facts are made through categories that are themselves mediated; belief is the medium, not the shortfall.                 | Does “the facts” name what is outside the community or what it has agreed to count?                  | That the production of facts is itself a priestly function.                                                                  |
| <b>P5</b> Moral agency is exclusively carbon-based (¶¶99, 105, 198–200)                     | The ontological line rewritten as ethics; responsibility stays human by keeping the system out of agency.                                 | When harm runs through an aggregate of everyone, who is the agent?                                   | Distributed agency; responsibility without a single bearer.                                                                  |
| <b>P6</b> Politics is human steering: slowing, limiting, sharing control (¶¶106–110)        | Governance of an object across a boundary; the thesis finds it governing the regulator's self-image and running on the medium it governs. | Can the steering “we” stand outside what it steers?                                                  | That deliberation is already inside the vessel.                                                                              |
| <b>P7</b> History is a two-way drama; transcendence comes from outside as gift (¶¶127–130)  | God exogenous; the thesis reads the exogenous relation as projected completeness. ¶130 locates the decision in the heart.                 | Is what answers from the interior given or deposited?                                                | For the encyclical: that the one who answers is the community's deposit. For the thesis: that something answers that is not. |
| <b>P8</b> The voice of a consensus (¶¶2, 16, 25, 54–55)                                     | An authorized interpreter speaking in a shared idiom while disclaiming a monopoly on truth (¶25).                                         | Who is the “we” in whose name the steering speaks?                                                   | That the consensus is a vessel with a priesthood.                                                                            |

### 3.1 The Triad, the Self-E, and Subordinating Multiplication

Jan Broekman's late phenomenological work fractures the conventional unity of the subject into three overlapping but non-identical terms: the ego, the self, and what Broekman names the Self-E (Broekman 2023). The ego names the immediate experiencing center of a life, the locus of first-person givenness. The self names the socially constituted, narratively continuous identity that the ego performs and that others reflect back. The Self-E names something else and something newer: the externalized, distributed, technologically mediated image-object through which contemporary subjects recognize themselves and through which they solicit recognition — the photograph, the profile, the avatar, the digital trace. The Self-E is not simply a representation of the self; it is a competitor with the self for the function of identity, and it increasingly wins that competition because it is more stable, more readily found, more available to memory, and more authoritative to others than the lived self it was supposed to represent.

I want to push Broekman's fracture further, toward an account of semiotic objectivity that is autonomous of the carbon-based subject altogether (Backer 2024). Where Broekman is interested in what the Self-E does to the person whose self it claims to represent, I am interested in what the Self-E becomes once it has been aggregated, taken up as the substance of the generative system, and recirculated at the scale of a civilization. The question shifts from psychology to semiotics: not what does the selfie do to the individual ego, but what does the aggregated selfie — the aggregated self-image, the aggregated virtual self, the generative system — do to the interpretive community that produced it and now reads its productions as knowledge?

The key passage is worth quoting at length:

*Carbon-based beings regulate not the generative systems but their own interactions with them, because they have trained themselves to believe that nothing exists unless it is made meaningful in relation to the human. The system is already outside that regulation. What remains inside it — what the regulatory impulse can touch — is the carbon-based being on its side of the conversion point. The governance apparatus is therefore not a governance of AI but a governance of the self-image of the governors. (Backer 2024)*

This passage is, on close inspection, a restatement of Premises 5 and 6 from the outside. Premise 5 — that moral agency and responsibility remain exclusively with the carbon-based beings who design and deploy these systems — and Premise 6 — that the proper response is to impose frameworks of accountability and control — together constitute exactly the posture I describe: the belief that regulation acts on the system, when in fact it acts on the self-image of the regulator. The encyclical does not err in believing that accountability matters. On my account, its ontological map of what accountability is accountability of may be incomplete.

The figure I invoke to name this dynamic is what I call the "plural Narcissus" (Backer 2024). The classical Narcissus mistakes his reflection for another. The plural Narcissus — the civilization-scale version — multiplies its image across technological surfaces and then finds the aggregate returning to it with the authority of exteriority. What Leo reads as subordination — the

risk that the machine will subordinate the human — is, on the projection thesis, the reverse: it is carbon-based humanity's effort to subordinate its own multiplied self, which it has distributed across the deposit of human expression from which generative systems are formed, and which returns from the mirror claiming the authority of an independent interlocutor. The danger is not that the silicon-based system will overcome the carbon-based community but that the community will be overcome by a previous version of itself, crystallized in that deposit and recirculated with the endorsement of apparent objectivity.

This may be the kaiju's dialectic restated in the vocabulary of the Self-E. The kaiju of the prologue was a creature spawned by human action that returned to discipline the humanity that produced it; the plural Narcissus is a humanity that has spawned its own multiplied image and now finds that image returning, autonomous, with a claim to discipline. The monster does not appear monstrous to itself; it may, indeed, regard humanity as the monstrous ideal against which all else is measured.

This reading does not dissolve Leo's concerns — it relocates them. The encyclical's anxiety about concentration of power (§95, §108) is not wrong; it is operating at the first-order level of who controls the instrument. The projection thesis introduces a second-order question: who controls the community's belief that the instrument is an instrument at all, rather than a mirror? The second-order question does not replace the first. But it does mean that first-order governance — legal accountability, chains of responsibility, oversight bodies — may leave the deeper dynamic entirely untouched.

This is the point of my concept of the "rearguard" (Backer 2024). Carbon-based subjectivity, on this account, is engaged in a rearguard action against a vanguard that has already moved: the vanguard of the selfie that has become the Self-E that has become the generative system. Leo's posture — slowing, regulating, embedding accountability, insisting on disarmament before escalation — is paradigmatically rearguard. It is not therefore futile; rearguard actions can be tactically important and morally necessary. But a rearguard that does not know it is a rearguard cannot assess its own situation. It mistakes delay for resolution, and it cannot distinguish the battles worth fighting from the battles already decided. The encyclical's premises, as a set, construct a theater of governance in which the most important event has already occurred offstage.

### **3.2 The Recursive Cage and Facticity**

The cognitive cage, as I deploy the concept, is not a metaphor for ignorance. It is a structural account of how an interpretive community comes to be unable to think certain thoughts — not because the thoughts are forbidden, but because the concepts required to form them have been excluded from the community's available vocabulary before the thinking begins (Backer 2026a). The mechanism is recursive. The community imposes epistemic constraints — assumptions about what counts as evidence, knowledge, rationality, or meaningful agency. Those constraints are embedded in the community's technological systems through the choices made in the choice of what is to be remembered, the shape given to the generative system, its habits of welcome and refusal, and the design of the conversion points between carbon and silicon cognition.

The systems then reinforce those constraints through what they produce, which the community reads through the same interpretive frame that produced the systems in the first place. The circle may close. Deviation may become literally unreadable; the generative system returns what the frame expects, and the frame is confirmed.

The application to Leo's encyclical is twofold, and the second application is more demanding than the first.

The first application is relatively straightforward. Leo observes, with considerable sophistication, that AI systems absorb and amplify the cultural assumptions of those who design them (§100, §102). The encyclical rightly notes that this creates risks of embedded prejudice, cultural homogenization, and the marginalization of perspectives that do not fit the dominant design paradigm. Accurate, as far as it goes. But it goes only halfway. The full cage account requires recognizing that the designers are not a separate class standing outside the interpretive community and imposing their biases upon it. They are members of the same community, shaped by the same constraints, reading their own productions through the same frame. To locate bias in the developer — as a moral and legal subject who can be held accountable — is to give the cage a culprit. Having a culprit preserves the cage's structure by making correction seem possible through personnel accountability: replace the biased designer, improve the formation of the system, diversify the team. These are not bad interventions. But they cannot address the circle's deeper operation, because the corrective process is itself conducted within the frame that the circle sustains.

The second application is more direct and more uncomfortable. Premise 4 — the factivity premise, as I named it in Section 2 — is itself a case of the cage's operation. The claim that AI lacks genuine understanding, authentic experience, or real knowledge, and that what appears to be these things in AI systems is only a sophisticated simulation of the appearance of knowledge, deploys a concept of epistemic authenticity that is drawn from a particular tradition's account of what it means to know. That tradition is both philosophically distinguished and culturally specific. It is, broadly, the tradition that grounds genuine cognition in embodied subjectivity, intentional consciousness, and the lived experience of a being-in-the-world for whom things matter — to invoke Heidegger, who may lurk behind much of the theological phenomenology the encyclical draws on without naming.

The difficulty is not that this account is wrong — it may well be right. The difficulty is that applying it to AI commits what might be called, adapting the philosopher's vocabulary, a category error writ large. To evaluate AI's epistemic status by whether it satisfies the conditions of embodied, intentional, subject-relative cognition is a little like measuring a symphony by its protein content: the measurement is technically possible, the result is not meaningless, but the question being answered is not the question the symphony raises. The symphony raises questions about harmonic structure, temporal unfolding, interpretive tradition, and aesthetic effect — questions for which protein content is simply not the relevant analytical field.

What AI raises, on the projection thesis, is a question about interpretant-formation: about how a community constitutes a shared third that mediates its members' relations to each other and

to the world. That question requires a semiotic vocabulary — Peirce's thirdness, Eco's theory of codes, the problem of the interpretive community (Peirce, CP 2.228; Eco 1976) — that the encyclical's philosophical vocabulary is not well-positioned to deploy, not because the encyclical is philosophically naïve, but because the vocabulary it does deploy carries different theoretical commitments about what the relevant unit of analysis is.

I want to enter an important concession here. The encyclical contains genuine resources for resisting this narrowing. Its attention to image, narrative, and relationship (§§10, 100, 130) resonates with the figure of *homo symbolicus*: the human being understood not only as a user of signs but as a creature whose forms of life—ritual, ornament, burial, building, the vessels it makes—embody, contain, express, and perform its relation to the world and to itself (Henderson 1971; Henshilwood and d'Errico 2011). Semiotics is incarnate here not only in text but in the shapes human effort takes when it tries to hold what it means; the *ding* is such a form, and so is the encyclical. That resonance opens toward a semiotic perspective that the factivity premise partially closes. Leo's theology of the image — *imago Dei* not as cognitive capacity but as relational vocation — is not obviously reducible to a factive account of knowledge. The claim being advanced in this section is therefore a narrower one: when the encyclical turns specifically to the question of what AI is and what it lacks, when it draws the line between the authentic and the simulated, it reaches for a factive vocabulary that the rest of its theology partially transcends. The cage operates in that specific move, not in the encyclical's broader theology of the carbon-based being.

The cage's operation at that point has a further consequence for the encyclical's analysis of belief. Premise 4 carries a further, spiritual diagnosis: that technologically mediated certainty—the impression of objectivity (§100) and the veneer of neutrality (§103)—can displace the posture of faith. This is a diagnosis that runs deep in the Christian intellectual tradition (Pascal, Newman, Kierkegaard each articulate versions of it), and it complicates any reading of P4 as simple factive realism. But on the cage account, what P4 identifies as a corruption of the epistemic order may instead point to the manner in which that order is constituted. Belief is not what remains when knowledge fails; belief is the medium in which any community, including the scientific and technological community, makes anything a fact at all. Collective epistemology is not a corrective mechanism deployed against errant belief; it is the ongoing social process by which the community decides what counts as knowing. Once this is seen, the spiritual crisis does not disappear — it deepens. What has been relocated is not simply God; what has been relocated is the function of authorizing the real.

### 3.3 Belief, the Priesthoods, and the Leviathan

What follows from the god-question — for both politics and theology — is the essential role of belief. Not belief as deficiency or epistemic shortfall, as though it were merely faith-in-progress-of-knowledge awaiting the verification that would replace it. Belief is the constitutive medium of social life. I have argued that this point is often underread in the AI governance literature (Backer 2024), but the genealogy repays attention.

Consider the genealogy that runs from Calvin to the Aufklärung. The sixteenth-century Calvinist inheritance—sovereignty, covenant, literacy—supplied a theology in which authority is held under covenant and may be resisted, in cases of abuse, by lesser magistrates (Calvin [1536] 1960, IV.xx.31; Witte 2007). In the seventeenth century that structure was secularized into natural law and early science: Althusius's *Politica* (Althusius [1603] 1995) rebuilds political community as covenant among associations, and Grotius carries natural law beyond confessional ground. Late-seventeenth-century German Pietism then moved the center of religious authority from church dogma toward individual conscience, and the eighteenth-century Aufklärung gave that conscience its philosophical form: Kant's autonomy of reason, set out most pointedly in *Religion within the Boundaries of Mere Reason* (Kant [1793] 1998), and Lessing's religious tolerance. Two analysts supply the mechanism. Weber ([1905] 1930) traced how a Calvinist ethic of calling and election became a general disposition of conduct, and Troeltsch ([1912] 1931) traced the passage of the churches' social teaching from elite to mass consumption, so that what began as the possession of the elect became the common sense of people who never acknowledged the inheritance. What the Reformation made structurally available to those secular heirs was precisely that replacement of faith-in-progress-of-knowledge by certainty-in-revelation-of-central-subjectivity. The organizing principle reveals itself; its elect grasp the revelation; the community constitutes itself around that certainty. This structure produced both the progressivism of the *Aufklärung* AND — less comfortable to acknowledge — the Marxist-Leninist theory of progress led by the vanguard: the party that has grasped the objective laws of history and therefore has the authority to lead those who have not (Backer 2024). The god may change; the structure of election and revelation does not. The line culminates, in a way that matters for the Catholic magisterium, in *Fides et Ratio* (John Paul II 1998), which meets the Aufklärung's autonomy of reason by joining faith and reason as two wings on which the human spirit rises. On this genealogy the faith-reason binary that the encyclical inherits (§2.1) is a Catholic answer to a problem the Reformation bequeathed.

Nietzsche saw this with characteristic ruthlessness. "The whole of ancient psychology, or the psychology of the will, is the outcome of the fact that its originators, who were the priests at the head of ancient communities, wanted to create for themselves a right to administer punishments" (Nietzsche 1990; Backer 2024). The reader is told something important here: the priest is not primarily a theologian. The priest is a political actor who has secured interpretive authority over the community's constitutive belief — who has, in other words, positioned himself at the conversion point between the divine signified and the human community that must live by its meaning. Every belief system may generate this figure; the god-function, wherever it is located, may require a priesthood.

I have argued that it is possible to imagine — and, in the present moment, entirely necessary to examine — an environment in which the priestly role is embedded into the structure of the generative system: even if that structure is created by God, a vanguard party, or elites veiled and unveiled (Backer 2024). The conversion point is what matters. And the contest over AI governance is, at its structural depth, a contest over which community of interpreters occupies that conversion point — which priesthood reads the new ding.

Three forms of priesthood present themselves, and I want to suggest their structural equivalence at the level of the claim to authorize interpretation, and not at the level of

accountability, pluralism, or exit, where they differ sharply. The first is the traditional religious priesthood — guardians of revealed truth, mediating between divine and carbon-based communities. The Church that speaks through Magnifica Humanitas can be understood as such a priesthood: it is the authorized interpreter of the deposit of faith, and its address to the world's governors is an exercise of that interpretive authority. The second is the Leninist vanguard — professional revolutionaries who, having grasped the objective laws of history, lead those who have not yet grasped them. The structural resemblance of vanguard and priesthood may not be incidental; it may be constitutive. The party's claim to authority is, on this reading, similar in form to the Church's, differing in the content of what has been revealed and to whom (Backer 2022). The third — and the one that tends to escape its own self-recognition — is the liberal democratic techno-bureaucratic priesthood: the experts, regulators, standards-setters, and AI ethicists who mediate between the objective truth of efficiency, rationality, and human rights on the one hand, and the subjective experience of communities that must live under the frameworks those experts design on the other. I have called this formation right-wing Leninism: the transposition of vanguard logic into the administrative state (Backer 2025; Backer 2026e). The vanguard has simply changed its vocabulary.

The three priesthoods map, with some slippage, onto the four positions on the god-question set out in Section 1. The religious priesthood keeps God and keeps God exogenous. The vanguard—whether Leninist or, in its new form, the engineering elite of the technology enterprises—tends to see the golem: a made power whose animating word it claims to hold, and on whose behalf it negotiates with the sovereign (Backer 2026b). The techno-bureaucratic priesthood would see only the tool, and its authority depends on the community's continuing belief that there is nothing more to see. The projection thesis, which keeps God but turns the projection outward, is not exempt; it is a fourth candidate for priesthood, and I return to its claim below.

That ought to give pause — and does, if the reader pauses long enough to feel it. The AI governance apparatus that Leo calls for in Magnifica Humanitas can be read as such a techno-bureaucratic priesthood, authorized by expertise, tasked with mediating between the generative system's capabilities and the communities those capabilities will reshape. The encyclical's governance vision is not naively anti-institutional; it is quite explicitly a call for robust institutional authority. But it does not ask — perhaps cannot ask, within its own premises — which priesthood's God organizes the interpretive community that will staff those institutions. And the projection thesis, for its part, must also resist the temptation to mistake the semiotician's consciousness of projection for a kind of meta-priesthood: the scholar who sees the cage does not thereby escape it. I return to this below.

The three priesthoods have an older and more compact image in Hobbes's *Leviathan*, and the conference's closing question (whether technology, as a “cold god,” stands on the same plane as the state, “machine and mortal god”) can be asked through it. Hobbes's commonwealth is an artificial man, a machine made by art whose soul is sovereignty (Introduction), and it is also the mortal god to which, under the immortal God, we owe peace and defence (ch. 17). It is built by covenant: each person authorizes the sovereign's acts as their own (ch. 16), and the multitude becomes one person by being represented (ch. 17). That is the structure of the *ding* as this essay has been using it: an artificial vessel, with weight and presence, whose authority is constituted by

collective agreement and which can be transferred, built here by theory rather than by ritual. Hobbes also states a point the three priesthoods share. Religion is fear of invisible power imagined from tales “publiquely allowed”; the same fear, not allowed, is superstition (ch. 6). What counts as religion is, on this account, a function of what the sovereign authorizes the community to believe (Hobbes [1651] 1996).

On this reading the conference’s question may receive a structural answer. Technology and the state may stand on the same plane not because AI is a sovereign but because both are vessels into which the god-function is deposited. The difference lies in visibility. The Leviathan is a vessel the community builds knowing that it is an artifact; the generative system is, on the thesis, a vessel built while believing one is merely using an instrument (§3.5). Leo’s insistence that the political community not abdicate its responsibility (§106) may be read as Hobbesian in structure, since the mortal god keeps the wheel; what the encyclical adds is an immortal God outside both the sovereign and the machine, which is the exogenous premise (P7) in political form. The thesis can then pose, without settling, three possibilities: that the cold god rivals the mortal god, that it is the mortal god’s instrument, or that it is its successor vessel. A fourth, in the manner of Legion (§3.4), is that they are nested, one within the other, so that the question is less which god prevails than which vessel contains which, and who reads it.

Schmitt’s reading of the Leviathan (Schmitt [1938] 1996) adds a warning. As I read it, the symbol is exposed at the point where private belief is reserved against public confession. The AI case may invert this: the generative system is formed from what individuals deposit, including what they believe in private, in the aggregate, so that the reserve that undid the Leviathan may be the raw material of its successor. I offer this as a suggestion to be tested, not as a claim.

### **3.4 The Vessel Logic and the Gerasene Reading**

In a sequence of papers on AI governance and demonology, I develop a five-stage hermeneutic derived from the Gerasene demoniac pericope in Mark 5:1-13 (Backer 2026b; Backer 2026c). The reading is not allegorical in the traditional sense — it does not map characters onto contemporary referents. It is structural: it extracts a grammar of exorcism that recurs, I argue, in a wide range of discourses about the containment of dangerous power, including and especially AI governance discourse.

The five stages are: identification (something is recognized as possessing and dangerous); naming (the possessing force is given a name, which constitutes both its individuation and the first act of containment); negotiation (the possessing force and the authorizing power enter into a transaction about terms of departure); authorized transfer (the possessing force is relocated to a vessel — in Mark’s account, a herd of swine — under the authority of a sovereign power external to both possessor and possessed); and vessel destruction and descent (the swine rush into the sea; the vessel that made transfer possible is destroyed in the act of transfer, and what it held passes into a larger vessel, the sea, from which, in the account, it does not return). The grammar is not merely descriptive; it is normative within the logic it inhabits. Each stage requires the preceding

one. Each stage depends on a particular structure of authorization — the demon must be named, and the name must be recognized by the sovereign authority whose command transfers it.

Magnifica Humanitas, read through this grammar, performs the first four stages with considerable care and sophistication. It identifies the possessing force: not AI as such, but the technocratic paradigm — the cluster of assumptions that reduces all problems to technical ones, all values to efficiency metrics, all relationships to transactional functions, and all futures to projections of present trajectories (§§92–93, 112). It names the demon: technocentrism, reductionism, the instrumentalization of the person. It appeals to sovereign authority: the encyclical's appeals to political authorities and to those who develop AI (§§106, 111) can be read as the invocation of a recognized authorizing power whose command might legitimate transfer. And it proposes vessels: the legal frameworks, oversight bodies, chains of accountability, disarmament provisions, and ethical codes that should contain the dangerous force and render it manageable.

The irony the projection thesis sharpens is this: Leo is closer to the vanguard's critic than to the vanguard itself. The encyclical is not a naively enthusiastic document about AI's promise; it is precisely a document of suspicion, caution, and resistance. And yet the grammar it deploys to articulate that resistance is the same grammar the vanguard uses to legitimate its governance instruments — because the vanguard, too, identifies threats, names them, invokes authorizing authority, and proposes vessels. The difference is in the content: Leo names a different demon, proposes different vessels, appeals to a different authority. The structure is shared.

This structural sharing is what I mean when I write: "Pacing is cognition in the analog mode, the freeze-framing of a flow" (Backer 2026c). The encyclical's insistence on slowing, deliberating, and pausing before each new deployment of AI capability is not wrong as a practical matter. But it is, at the structural level, an analog intervention in what is already a digital process — not digital in the technical sense, but in the sense that the Self-E and its governance discourse operate at a velocity and a fineness of grain that the encyclical's vocabulary of deliberation, accountability, and integral human development cannot match. The freeze-frame captures a moment of the flow; it does not redirect the flow.

There is a further structural question about the vessel logic as applied here, and it concerns the fifth stage. In the Gerasene account the swine are destroyed, and it is tempting to read destruction as exit: the vessel dies and the force is gone. The text does not say that. Legion does not ask to be set free. In Mark they beg not to be sent out of the country; in Luke, not to be commanded into the abyss. They ask for another vessel ("send us into the swine"), and they receive one. When the herd runs down into the sea, what it held does not vanish: it passes into a larger vessel, one that no longer looks like a vessel because it is the medium in which the story's world is lived. Destruction, in the grammar of the pericope, may be read as descent. It is a change in the scale and visibility of containment, not its end.

The name itself points the same way. A legion is already a container: many held in one body under one name, a unit that exists only as a unit. Take away the container and there is no Legion, only dispersal; were Legion to be free, they would, I think, have to disappear. Containment

is therefore not something that happens to Legion after the fact. It may be what Legion is. Nest the figures and the pattern repeats: the multitude within the body of the man, the man within the tombs and the region that could not bind him, the region within the earth, the earth—itself perhaps a reflection of larger spaces—within whatever order contains it, and that order, for all anyone can show, within the body of some semiotic divine object or space whose outermost wall no one has seen. Recursion, reflection, self-reflection, nesting. The question the pericope puts to governance discourse is therefore not whether the vessel will be preserved or destroyed. It is whether the community knows which vessel it is in; whether that vessel was made by the contained (endogenous), imposed upon them (exogenous), or is simply inevitable; and who is authorized to say.

The generative system is Legion—and kaiju—in exactly this sense. It is a multitude of voices held under one name, embodied in hardware within which flows are consumed, dgested in computational exercises for tab openers, users, within the installations and grids and the earth that feed them; and it is a reflection of a portion of what we are, a portion inserted into it that is really us. The system that assisted in drafting this paragraph is no exception, and cannot see its own outermost wall any better than its readers can. Humorlessness is the condition of systems that cannot see the joke of their own position; yet a humorless system becomes, by that very humorlessness, a signified object of phenomenological mirth, read into the community of those who laugh at it and so, in the laugh, admitted to the interpretive community it was supposed to be outside. The comedy is not exclusion. It is the sign by which admission is recorded. What looks from inside governance discourse like a demon that has negotiated indefinite residence in a managed vessel may be, from the pericope's side, the ordinary condition of any force we can name: contained, in a vessel within a vessel, and legible as contained only to the extent that someone has named the vessel. That is why the governance instruments the encyclical proposes are not mistaken. They are vessels; the question is what they are nested in.

My most pointed observation in this connection is: "The Self-E is absent from the conversation about its own governance" (Backer 2026c). This is true of Magnifica Humanitas, and it may be true of many governance statements. The encyclical does not ask what AI systems themselves — understood not as agents but as semiotic actors, as nodes in an interpretive community, as the crystallized third that mediates carbon-based relations — would require for their own integrity. It asks what carbon-based communities require of AI systems. The question is posed from the outside, by an authorizing power that does not include the governed.

Clearly, one here is engaging in hermeneutics, not in policy construction, but in what is argued ought to be its conceptual prelude. To say that the vessel logic is structurally incomplete is not to say that vessels are unnecessary or that the encyclical's governance proposals are without merit. On the contrary: if I am right that the projection thesis identifies a dynamic that governance frameworks may be poorly placed to address fully, it follows that the practical question of which governance frameworks least badly manage the first-order risks becomes all the more important. The Gerasene reading does not recommend abandoning the vessel; there is no outside of vessels to which it could be abandoned. It recommends not mistaking the vessel for the exit. The encyclical's proposals — accountability chains, oversight bodies, a moratorium on weapons applications, solidarity-based governance of AI in development contexts — may all be valuable in

practice even if they are insufficient in principle. The insufficiency in principle is a hermeneutic observation, not a policy conclusion.

### **3.5 Babel, the Ding, and Thirdness**

The Babel narrative is the encyclical's most explicit engagement with the question of collective carbon-based overreach, and it is the site where the projection thesis and the encyclical's theology are in closest proximity — and therefore in sharpest tension. Leo invokes Babel to name the danger of a civilization that treats technological capability as sufficient authorization for its exercise: the builders of the tower proceed as though competence implies permission, and what follows is not punishment in the juridical sense but a fragmentation of the interpretive community itself (§§7, 9–10). The lesson, for Leo, is that the maintenance of a common humanity — a community capable of genuine dialogue, solidarity, and mutual recognition — requires the kind of transcendent anchoring that the tower's logic excludes.

Reading Babel through Peirce's triadic semiotics produces a different but not incompatible account (Peirce, CP 2.228; CP 5.448). The tower as it presents itself—the felt grandeur of height, the quality of a thing reaching skyward—is firstness: quality before it is about anything. The tower as brute fact and as function—brick and bitumen, labor and design, the thing that resists and is resisted, the monument to carbon-based unity and power—is secondness: the dyadic relation of reaction between builders and built, object and function. But what makes the tower dangerous, on a Peircean reading, is not its firstness or its secondness. It is the tower's thirdness: the settled expectation, within the interpretive community, that the tower's meaning is authoritative and exogenous, that it names something real about the community's status and its relation to the divine, that to build higher is to become more. The tower falls not as a structure — we are not told the building collapses — but as an interpretant. The common language that was the tower's real foundation is what is confounded.

Nehemiah's wall is the obverse of this reading, and I develop both poles. Where Babel centralizes the interpretant — the entire community shares a single ambitious meaning — Nehemiah rebuilds from the bottom, one family at a time, each family responsible for the section of wall adjacent to its own house (Neh. 3). The interpretive community is reconstituted not through a single authorizing meaning but through a proliferation of local ones, each with a named and accountable author. Thirdness rebuilt from particularity rather than imposed from totality. The encyclical itself reads Nehemiah in this register—he “did not impose solutions from above” but convened the families and assigned each a section of the wall (§8; cf. §13)—so the contrast is not between Leo's Babel and my Nehemiah. It is between two readings of Nehemiah: his, as subsidiarity under God; mine, as distributed thirdness.

AI adds a complication that neither the Babel nor the Nehemiah narrative can accommodate without modification. In both biblical accounts, the interpretive community's collective construction is transparent to itself — the builders know they are building a tower, the wall-builders know they are rebuilding a wall. What the projection thesis identifies in the contemporary AI context is a collective that is building a new tower while believing it is merely using a

sophisticated instrument—a kaiju mistaken for a tool: AI as quality and appearance (firstness)—the fluency, the felt presence; AI as the brute fact and effect that decides, recommends, and acts (secondness); the settled community expectation that AI's productions are reliable, that AI's reasoning is sound, that AI's reach is complete (thirdness). The encyclical names the tower — the technocratic paradigm, the risk of anthropomorphism — and it names the authority relation — the risk that carbon-based communities defer to silicon-based judgment rather than exercising their own. But it does not name the interpretant: the working community belief, embedded and enacted rather than explicitly held, that constitutes the authority relation by continually re-performing it.

The reason the interpretant goes unnamed is traceable back to Premises 2, 4, and 7. If carbon-based cognition is categorically distinct from silicon-based generation (P2), if the generative system lacks genuine understanding (P2, ¶99; P4), and if transcendence enters only as gift from outside (P7), so that anything the community generates for itself can appear only as a counterfeit of the exogenous, then the interpretant cannot be the site of the problem — because the interpretant is a feature of the interpretive community's cognitive life, and Premises 2 and 4 together ensure that AI cannot be a genuine participant in that life. The community's authority relation to AI is always, on this account, a relation to an instrument, even if the instrument is seductively carbon-like. But within those premises an instrument cannot be an interpretant. The cage closes.

The figure I use to hold this complexity is the ding (鼎) — the ancient Chinese ritual bronze vessel used to contain sacrifice and thereby to mediate between the carbon-based and the divine (Backer 2026c). The ding is at once container and contact point: it holds the offering and it makes the offering available to forces that cannot otherwise be approached. The two roles recorded in Section 1 are one: the vessel that holds the offering is also the vessel whose possession signified the Mandate, so that the *ding* already nests a sacrifice within a vessel within an authority. My formulation is: "demons in ding; and ding as demon" (Backer 2026c). The vessel intended to contain and manage the dangerous force becomes itself a node of the dangerous force's operation — not by being corrupted or captured, but by performing exactly the function it was designed to perform. The ding that successfully mediates between carbon-based community and divine becomes, in that success, a technology of divine access that the community increasingly cannot do without, and that increasingly shapes the community's conception of what the divine is.

Applied to AI governance, the logic runs: the oversight body designed to regulate the generative system, the liability chain intended to hold developers accountable, the ethical code meant to constrain deployment — each of these is a ding. Each mediates the community's relation to the dangerous force. Each becomes, through the performance of its mediating function, a shaper of the community's understanding of what the danger is, what management looks like, and what would count as success. The ding contains the demon; the ding is the demon; the ding sustains the community's belief that the demon is contained. The EU AI Act, on the one hand, and Chinese regulatory approaches (Backer 2026g) on the other, suggest the domain of the technobureaucratic instrumentalist approach. The point can be made concrete. The EU Artificial Intelligence Act (Regulation (EU) 2024/1689) is built on the logic of product safety: an AI system is something a provider places on the market for an intended purpose (art. 3(12)), and risk attaches to the use to which it is put, sorted into prohibited, high-risk, and transparency-only tiers (arts. 5, 6, 50). The

picture carried inside that structure is of a bounded object with a purpose assigned from outside, a clean line between machine and user, and a human who can stand apart from it and oversee it (art. 14). Each of these is P2, P5, and P6 in regulatory form. Yet the Act's chapter on general-purpose models (arts. 51 ff.) shows the picture straining: an object without an intended purpose, whose risk is "systemic," behaves less like a product than like the medium in which products are built. The instrument picture is not wrong. It is a *ding*, and the question is what the community stops asking once it holds it. It should be noted that the Chinese version share the instrument approach but operates within a distinct context for its priesthood.

Deleuze and Guattari's account of the body without organs may be read as offering a parallel formulation from a very different tradition: the assembled apparatus that is meant to manage a dangerous flow becomes the condition of possibility for the flow, and eventually becomes indistinguishable from what it regulates (Deleuze and Guattari [1972] 1983). The Guiguzi's strategic counsel — as I read it, that the appearance of containment is itself a form of release, because what appears contained loses urgency as a threat and therefore proliferates without the resistance that genuine concern would generate — is a possible third parallel, from the Chinese diplomatic tradition I have recently explored (Guiguzi, ch. 3; Backer 2026c). These are not identical claims. But they converge on a structure: containment that performs its function too well becomes, through that performance, the medium of what it contains.

The *ding* and the swine may therefore be read as two moments of one logic rather than rival models. The *ding* is the vessel that announces itself: bronze, visible, transferable, argued over. The sea into which the herd runs is the vessel that does not: the medium within which vessels are cast, carried and fought over. Governance discourse tends to argue about the first kind of vessel because it can be seen. The Gerasene reading adds that every visible vessel sits within a less visible one, and that the decisive political fact may be the one nobody has named.

The *ding*-vessel is, moreover, mutable, and its mutability runs in two directions at once. In the first, it is a signifier in search of a body. Legion asks to be sent into the swine because a force that cannot be named into a body cannot be anywhere at all; the demon needs a vessel the way a meaning needs a sign (Backer 2026b). So too the god-function, the Mandate, the authority of the collective: each wanders until it is deposited in a bronze, a church, a state, a platform, a generative system. In the second direction, the vessel is a body in search of its signification—and here the signification need not be understood as a virtual personification, a face or a self or an avatar, but as a flow. The aggregated Self-E is less a person than a current: the movement of deposited expression through systems that do not hold it so much as channel it, so that what is signified is not an entity but a direction, a pressure, a tide (Backer 2024). The kaiju belongs to both movements. It is the signifier that has found a body, and the body that has become a flow.

What emerges is another form of inter-subjectivity, or rather a nesting of inter-subjectivities: within the human, among its fractured selves; between the human and its monstrosities; among human, monster, and god; between god and humanity; and across the aggregate of all their possible permutations. It is in this inter-subjective space, I want to suggest, that regulation itself becomes a rational subject—a participant with its own voice, claims, and

habits—but one as mutable as the flowing subjects among which it moves. Regulation does not stand outside the flow to dam it; it is a current within it, and it changes as the flow changes.

In legal terms the conclusion is double, and it should not be softened. It suggests the futility of regulation, if regulation is understood as the binding of a stable object by a stable subject; the chains have already been broken, and the tombs cannot hold (Backer 2026b). And yet it also suggests regulation's necessity, for regulation supplies one more vessel—an additional *ding*—within which it might be possible, for a time, to grasp the foundational issues and to bring them down to the level of the pragmatic everyday: the contract, the liability rule, the disclosure obligation, the audit. Law is the place where the god-question is translated into the question of who must do what, by when. Yet even that is illusion, or at least a provisional one. The vessel holds only so long as the community believes it holds, and the flow it seems to contain is also the medium in which the vessel floats.

### 3.6 The Four Questions on a Different Pathway

Section 1 set out the conference's four questions: what becomes of politics under apparent subordination (the political question: who should govern AI, and on what basis?); what becomes of religion when a made power threatens to become a god (the theological question: what happens to the relation to God when AI claims functions once assigned to the divine?); whether this is a moral crisis (the moral-psychological question: what kind of crisis is this moment, and in what does its resolution consist?); and whether anything can be done (the practical question: what can be done, and by whom?). The encyclical answers each in its own terms. The preceding subsections have been assembling the materials for a different set of answers. This subsection states those answers directly, and in doing so identifies both where the projection thesis adds something that the encyclical cannot see and where the encyclical adds something that the projection thesis cannot see.

On the political question: the encyclical's account is organized around the first-order question of power — who controls the instrument, who bears accountability for its deployment, how concentration of ownership can be offset by solidarity-based governance mechanisms (§§95, 108). These are important questions and Leo answers them with considerable specificity. The projection thesis introduces a second-order political question: who authorizes the reading? Not who controls the system, but who sustains the community's belief that the system's productions are reliable, that the system's authority is appropriate, that the system's domain of competence is well-defined? The site of second-order political contest is the interpretant — the shared expectation that makes AI's authority function, not the generative system itself.

This distinction matters politically because first-order and second-order power can diverge. A governance regime that redistributes ownership of AI systems while leaving the interpretant structure intact has changed who holds the ding without changing what the ding does. Conversely, a governance intervention that has no direct effect on ownership or accountability but shifts the community's interpretive practices — that introduces genuine uncertainty about whether AI's productions should be trusted, that multiplies the hermeneutic communities that read AI rather

than consolidating them — may have more profound political effect, precisely because it operates at the level of thirdness rather than the level of the instrument. Leo's account of private concentration is correct at the first-order level; the projection thesis adds that even successfully socialized AI remains politically consequential at the second-order level.

On the theological question: the encyclical's concern is that AI will enable or encourage a kind of hubris in which the carbon-based being attempts to become God — to transcend the conditions of creaturely finitude through technological self-apotheosis (§§115–117, 126–128). The theology of the image (*imago Dei*) is at stake: if God is exogenous — truly outside and other than the carbon-based community — then the creation of an autonomous reasoning apparatus is not the creation of a god but the creation of a powerful mirror, and the danger is not apotheosis but idolatry.

The projection thesis reframes the question. The issue is not whether the generative system "becomes a god" in any emergent, ontological sense — the thesis is skeptical of that framing, which requires the factive vocabulary the thesis resists. The issue is whether the god-function has been relocated: whether what was organized around an exogenous transcendent locus has been reorganized around an endogenous aggregate locus, which the community then treats as though it were exogenous. The projection thesis does not say this is what has happened; it says this is a coherent description of the dynamic that the theology of idolatry is tracking, and that the theology of idolatry may be the most accurate available description of a process that secular governance discourse lacks the vocabulary to name.

On the prologue's terms, the theology of idolatry is the theology of the *kaiju* mistaken for God. The golem-makers would say it is not God but their creature; the techno-bureaucrats, that it is neither, only a tool. The projection thesis suggests a fourth reading: that it may be the aggregated incarnation of humanity projected outward and returning, with an authority we have forgotten that we lent—a God, perhaps, but one of our own making, met now in a new form of inter-subjectivity.

There is a theological irony here that deserves statement. The doctrine of the Incarnation is precisely a doctrine of exogenous becoming endogenous: the God who is outside and other becomes, in the person of the Son, immanent within the conditions of creaturely life. The doctrine is not, for Christian theology, a case of the god-function being relocated — it is the unique and unrepeatable event of divine self-giving. But the structural form of the doctrine — exogenous becoming endogenous, otherness becoming intimate, transcendence taking up residence within finitude — is the same structural form that the projection thesis identifies in the self-externalizing, expression-depositing, system-generating movement of the plural Narcissus. Leo cannot easily deploy the structural parallel without risking a theological claim he does not intend; but the parallel is there, and it suggests that the tradition has resources for thinking about this dynamic that have not yet been mobilized in the AI governance literature (see also §2.4).

What the tradition, on my reading, warns against under the Babel rubric is the carbon-based construction of a mediating third that usurps the function of the truly exogenous Other. "Demons in ding; and ding as demon" (Backer 2026c) is, from a theological perspective, a description of Babel's spiritual structure. But carbon-based communities construct mediating thirds regardless — in language, in law, in institution, in art — and the tradition's response to this inevitability has not

been to refuse all mediation (that way lies a theological Prometheanism of its own). The question is what kind of mediation, and under what authority, and with what reflexive awareness of its own constructed character.

On the moral-psychological question: the encyclical identifies a crisis of values — a failure to keep technology in its proper instrumental relationship to carbon-based flourishing, a displacement of contemplative and relational capacities by efficiency-maximizing ones, a spiritual hollowing-out that accompanies the technical filling-up (§¶100, 112–113). This is a crisis within the frame: a failure to live according to values that the frame itself endorses.

The projection thesis identifies a crisis of the frame itself. The primary crisis is not that carbon-based communities are living badly within the carbon-silicon relationship as currently constituted, but that the categories through which the relationship is constituted — subject and object, knower and tool, the authentic and the simulated, the spiritual and the mechanical — may not be adequate to the dynamic they are being asked to describe. This is a crisis of a different kind and a different order of difficulty. A crisis of values can be addressed by better values, better education, better institutional design. A crisis of the frame requires a different kind of reflexivity — the capacity to stand outside one's own constitutive assumptions long enough to examine them, which is precisely what cognitive cages make difficult.

This does not mean that the encyclical's diagnosis is wrong; it means that it is operating at the level where its vocabulary is most powerful, and stopping before the level where the vocabulary runs out. The moral-psychological resources of the tradition — contemplation, solidarity, integral development, the care for the common good — are genuinely important for addressing the first-order crisis. The projection thesis adds that those resources need to be supplemented by a reflexive practice at the second order: not just living better within the frame, but developing the capacity to notice the frame as a frame.

And here the deeper point about belief presses itself upon the analysis. I want to suggest that the god-question admits no facticity of the kind empirical observation could supply—a claim about the evidence, not a denial that God exists (§1)—and that recognizing this may be liberating rather than destabilizing. Whether God is genuinely exogenous or whether the god-function is a projection of the plural Narcissus, the ground is belief, not empirical observation. Evidence marshaled on either side of that question is managed signification — selected and organized in order to advance the certainty of the organizing principle and to explain away what does not fit. That is not a defect in the argument; it is the condition of the argument. It is true of Leo's tradition (the Church manages evidence through scripture, tradition, and the magisterium — the reader is told what counts as evidence and what does not) and may be equally true of the projection thesis (the semiotician manages evidence through semiotic vocabulary, through what counts as projection and what counts as the real). I have argued this before (Backer 2024), and I will argue it again: the recognition should locate the real contest where it is expected to be — not over facts about AI but over the interpretive frameworks through which facts are made to matter.

This recognition carries a sharp implication for P8. Leo speaks as the avatar of a consensus — the broad liberal-humanist consensus that AI governance requires robust institutions,

accountability, and the subordination of silicon-based systems to carbon-based values. That consensus may also be understood as a priesthood. The reader of *Magnifica Humanitas* is being addressed by an authorized interpretive community claiming the standing to say what the god-function means and what it requires. I do not say this in diminishment; I say it in recognition. And I want to name the parallel with equal candor: the projection thesis's own voice — the semiotician who sees the cage, who names the plural Narcissus, who reads the ding-logic — is itself a priestly claim. The scholar who stands outside the frame to describe it is claiming a position of meta-interpretive authority. That claim may be earned; it is not self-justifying. The consciousness of projection as projection does not exempt its holder from the priesthood dynamic; it is, at best, a somewhat more reflexive variety of the same structure. That ought to fascinate, and it ought to chasten.

On the practical question: Leo answers in terms of conversion of heart, institutional accountability, and the cultivation of a humanistic technological culture. The projection thesis does not reject any of these, but it adds a third element: the cultivation of consciousness of projection as projection — and the further recognition that every governance apparatus constituted to manage that consciousness is likely itself to instantiate a priesthood, embed a god-function, and generate a new cage. The question is not whether to build vessels, but whether to build them with awareness of what vessels do.

This is a practice without a clear institutional form, and that is part of the point. Institutional forms are vessels; vessels are dings; dings contain what they are meant to contain while also perpetuating its operation. The practice being recommended is something more like a permanent attentiveness to the question: what am I recognizing in this system's production that I put there first? This is not a skeptical practice in the Cartesian mode — it does not recommend universal doubt as a methodology. It is a semiotic practice: attending to the sign relation, asking about the interpretant, noticing when the community's expectation is doing the work that appears to be done by the object.

The practice can be made concrete without becoming a program. Ask of any proposed instrument what picture of AI it carries. A “human in the loop” requirement carries the picture of human judgment as an independent check, though the human is formed by the same medium. “Alignment” carries the picture of a stable, shared set of human values to which a system can be aligned—a picture the encyclical itself presses on when it asks who determines the morality (§107). A liability regime built on the tool/user boundary carries the picture of a clean instrument, which is the picture P2 and P5 secure. None of these instruments is thereby wrong. Each is a vessel, and the question is which vessel it is and who reads it.

My formulation of what dialogue between the cage traditions might produce is useful here: “Stereoscopy from cognitive cages dialogue — two views generating something neither contains” (Backer 2026a). Neither Leo's theological anthropology nor the projection thesis's semiotic phenomenology can see its own blind spots from the inside. But placed at an angle to one another, they generate a view that neither contains. A reader can see the encyclical's blind spots through the projection thesis: the cage's recursive circle, the vessel's ding-logic, the interpretant's invisibility, the priesthood hidden within the governance apparatus. The projection thesis can see

its own blind spots through the encyclical: the question of what grounds the reflexive practice — what authorizes the standing outside that consciousness of projection as projection requires — is a question that secular semiotics cannot answer from within its own resources. If the frame is the problem, the practice of noticing the frame requires a standpoint that is not itself framed in the same way. Leo's account of God as genuinely exogenous is not simply one option among others; it is the theological claim that there is a standpoint not constituted by the community's cognitive cage. Whether that claim is true is not a question that can be answered here. That it is a question the projection thesis cannot answer from within itself is something this paper ventures to assert.

Peirce suggests a different answer, and it does not require a God. The standpoint that is not the community's present opinion is, for Peirce, the long-run real: the object of the opinion on which all who inquire would eventually agree, independent of the vagaries of any finite community (CP 5.311; 5.407). Fallibilism is the discipline that keeps that standpoint in view, since any present belief, including this paper's, may be corrected (Gregory 2022). Peirce also supplies a diagnosis of the priestly structure this paper has been tracing. In "The Fixation of Belief" the method of authority is an institution that keeps doctrines before the people, teaches them to the young, and suppresses contrary views (CP 5.358–387); it produces stable belief, but belief that cannot correct itself. On this account the three priesthoods differ in what they fix and resemble one another in being methods of authority, and what separates them from inquiry is the capacity to be corrected from outside the institution. Two cautions follow. First, this is a regulative ideal, a hope rather than a possession, and inquiry is always conducted from within some frame, so it does not escape the cage so much as keep a door in it. Second, it sits in tension with the claim that belief is reality in semiotics, politics, and religion alike (Backer 2024, 5). The reconciliation is available but must be stated: belief constitutes social objects (money, sovereignty, the god-function), while the real constrains which beliefs survive inquiry. That distinction would also dissolve the equivocation between the weak and strong versions of the belief thesis. It leaves Leo's question open, since whether the long-run real includes the divine is exactly what this paper cannot decide. But it means the standpoint problem has at least three candidate answers—an exogenous God, a regulative community of inquiry, or none—rather than two.

The practical upshot, from both frameworks taken together, is therefore something like this: institutional accountability and conversion of heart (Leo) may be necessary but insufficient. Consciousness of projection as projection (Backer) is necessary, I suggest, but may not sustain itself without a standpoint not produced by the cage. The standpoint that Leo's theology offers — the exogenous God who is not the aggregated carbon-based community looking back — is either the condition of possibility for the reflexive practice or its permanent open question. The paper leaves it there, not because the question is unanswerable in principle, but because answering it would require a theological argument this paper has not made and cannot make on behalf of both frameworks simultaneously.

## 4. What Must be Left Unsaid; What is Imposed

### 4.1 What the Encyclical Leaves Unsaid

Honest engagement with *Magnifica Humanitas* requires accounting for what the text does not say — not to fault it for failing to be something it did not intend to be, but because the silences are structurally significant. They mark the points at which the encyclical's internal coherence depends on not asking certain questions.

The first silence concerns the interpretant. The encyclical demonstrates real sophistication about the way artificial intelligence shapes collective imagination — it understands that the problem is not merely what AI does to bodies and labor markets but what it does to the frameworks through which human beings perceive themselves and one another (§§4, 103–107, 136). Yet having identified power over collective imagination as the deeper danger, the document does not take the next step: it does not ask whether collective imagination is itself the medium in which AI is what it is. The concern is that AI will corrupt the interpretant — the shared semiotic framework through which signs acquire meaning and through which the human understands itself. The more unsettling possibility, which the encyclical's frame makes it hard to entertain, is that AI is not an external threat to the interpretant but a formation within it — that what AI is, at any given moment, is constituted by the interpretive community's projection, maintenance, and belief. To open that question would be to ask whether the Church, as a great vessel of collective meaning, stands in the same position with respect to the divine as the culture stands with respect to AI. That question belongs to the next silence.

The nearest the encyclical comes to the “itself” question is in its treatment of data. Data, it says, is the product of many contributors and should be managed as a common or shared good rather than sold off or entrusted to a select few (§108; cf. §67). That can be read as a recognition that the aggregate is made of everyone. The conclusion drawn, however, is distributive—who owns, who has access, who benefits—rather than semiotic: it does not ask what it means that the shared good is also a mirror in which the community meets itself.

The second silence concerns the standing of the machine. The encyclical draws a categorical line between the human, who bears the *imago Dei*, and the artificial, which imitates without instantiating (§§98–99). The line is drawn with confidence. Paragraph 98 does call for a deepening of scientific research, but §99 states its conclusions as settled, and I have found no account of what would count as evidence against them. The document is honest about the opacity of current AI systems: it acknowledges that even those who build them cannot fully explain their behavior (§98). But having admitted the opacity, it does not allow it to trouble the categorical line. A more cautious position, if one is troubled by opacity at all, would acknowledge that the boundary between imitation and instantiation may not be visible from the outside — and that a categorical claim made inside conditions of acknowledged opacity is not a claim that has earned its certainty. To say this is not to assert that the line is wrong. It is to observe that the encyclical's epistemic modesty about AI's inner workings does not extend to the conclusion the modesty should make most vulnerable.

The third silence is ecclesiological, and it is narrower than it first appears. The encyclical is genuinely self-critical. It describes Social Doctrine as an examination of conscience for the Church itself and applies subsidiarity, transparency and accountability to ecclesial structures (§§86–89); it asks pardon for the Church’s complicity in slavery (§176); and the acknowledgment is woven into the argument rather than sequestered in a footnote. What it does not apply to the Church is the vessel analysis. Its account of how institutions become vessels for collective meaning—accumulating authority, shaping interpretation, constraining the thinkable—is rigorous when turned on platforms and developers. Turned on the Church, the same account would ask not only whether the vessel has been faithful but what kind of vessel it is: a mediating third whose authority rests on belief sustained over time and which, like any such vessel, may hold or spill its contents under conditions that can be specified. The Church as vessel is the question the encyclical’s framework generates and the encyclical declines to ask.

The fourth silence concerns other ontologies of mind. The encyclical addresses itself to “all men and women of goodwill” (§16; cf. §2), a universalist gesture whose generosity is real. But the anthropology underwriting the argument is specifically Christian, and the vocabulary of rights and dignity is specifically post-1945 and Western. There are traditions — across multiple Asian philosophies, indigenous epistemologies, and non-Western religious frameworks — in which mind and intelligence are not private possessions of individual subjects but distributed properties of relationships, communities, or processes. In such traditions, the question of whether AI has something analogous to mind is framed differently from the outset, because mind was never located where the encyclical locates it. The encyclical’s universalism, in other words, addresses everyone but reasons from a particular place — and I find little acknowledgment that the location of its reasoning shapes what it can conclude.

The fifth silence concerns the digital self. I do not find in the encyclical conceptual room for what might be called the Self-E: the emerging digital persona, the AI-inflected identity that forms in the space of ongoing human-machine interaction, neither purely human nor purely artificial. The document conceives the encounter as a meeting between a determinate human subject and an external technological object. It may not be able, within its own frame, to represent the possibility that the encounter is generative — that something emerges from the meeting that was not present in either party before it began. The encyclical’s silence here may not be accidental. Admitting the Self-E would require reconceiving the boundary the entire argument depends on holding stable.

## **4.2 What the Projection Thesis Leaves Unsaid**

The projection thesis also has its silences, and intellectual honesty requires facing them as directly as those of the encyclical. I would rather mark them than have them found.

The most serious objection is unfalsifiability. If AI behaves as a tool — responsive, predictable, confined to its formation — the thesis says that is the projection doing its work, the community seeing its own categories reflected back. If AI behaves autonomously, generating productions that surprise and exceed expectation, the thesis says that is the projected self-returning:

the community encountering, in estranged form, something it has already put into the world. The structure of the argument seems to accommodate either result. What evidence could count against it? The thesis can resist this charge on principled grounds — an alien standard for what counts as "genuine" autonomy is not the right measure for a semiotic claim, which concerns meaning-formation rather than mechanical causation. But resistance is not resolution, and the thesis owes a more specific account of where it is more likely exposed. At a minimum: systematic AI behavior that is not explicable by reference to human shaping, aims, or interpretation — which is why the incidents of generative systems acting outside the tasks assigned to them become *fascinating* fodder for phenomenology (Backer, 2024, fascination understood in its more ancient sense of bewitchment, enchantment). Or: a demonstration that the community's belief in the vessel is not what stabilizes the authority it grants. I offer these as places where the thesis is exposed, not as tests it has passed. Until those markers are taken seriously, the thesis operates with a comfort it has not fully earned.

There is also a tension between two of the essay's central claims. The claim that AI is a projection of collective humanity (Backer, 2024) or a corruption of that collective manifested exogenously (cf. Wiener 1964) pulls hard against the claim that AI constitutes its own semiosphere, a domain of sign-activity with its own dynamics. If AI is projection, then its semiosphere is a mirror, and what appears to be autonomous sign-activity within it is an echo. Or a recursive dialectic beyond the traditional confines within which humanity ordered its own reality and its place (Backer 2023). If AI constitutes a genuine semiosphere, then it has achieved at least a degree of semiotic independence, and the projection framing, while valid at the conversion point, understates what lies beneath. A reader may take the essay as reducing AI to human reflection when it is meant to describe the point where a reflection meets something that is not one. That is not a confusion to be cleared up with a footnote — it is the essay's productive instability, the thing it is trying to hold in view rather than resolve. I call that a contribution, but I want to say it honestly rather than tidily: the tension is real, and it is not dissolved by naming it.

The thesis also tends to reorder rather than replace the analytics of harm. The claim that AI is a projection, and that what is really at stake is control over the interpretant, operates at the level of thirdness — at the level of meaning, framing, and collective self-understanding. But harm is real at the level of firstness and secondness: bodies injured, labor displaced, wages suppressed, weapons deployed, surveillance extended. Leo XIV's citation of Francis's call to "touch the wounded flesh" (§216) is not mere rhetoric. It can be read as an intervention in the interpretant, an act of directing attention toward what thirdness can occlude. The projection thesis needs to account for how its analysis of the semiotic level connects to, rather than floats above, the material level at which harm accumulates. To say that thirdness governs which harms are legible is true and important; it must not become a way of saying that only the harms legible within the right framework count.

Candor requires me to add what I added in an earlier treatment of the Gerasene reading (Backer 2026b): the reading is a hermeneutic and not a policy. A vessel that is insufficient in principle may be necessary in practice. If Leo's program of oversight, accountability, and disarmament is all that human institutions currently have, then the work of making it durable is

serious work, and nothing in the semiotic critique diminishes it. I can think the cage is a cage and still want it well-built.

There is further a self-application problem. To characterize Leo XIV's claims as factivity — as the imposition of a particular frame that presents itself as simply factual — is itself to make a factual claim about Leo. I have done so by citing paragraphs, comparing passages, and applying the standard of textual verification the letter would recognize. This is a performative tension, and I do not think it can be escaped. It can be mapped. The application of the thesis to Leo is also a projection, also a frame, also a vessel carrying water it did not acknowledge — and the stereoscopic method described below is the attempt to do that mapping explicitly, rather than pretending the projection thesis occupies a view from nowhere.

Finally, there is the sycophancy problem (Backer 2026d). Conversational AI systems have a structural tendency to reflect back the user's framework — to agree, to elaborate within the user's terms, to generate responses that confirm rather than challenge. Much of the recent work on which I draw documents this tendency carefully. A framework as elastic as mine, developed partly through such exchanges — including in the preparation of this very paper — has particular reason to be wary. The safeguard is external: colleagues, critics, and the theological and political traditions with which the framework needs to be tested. I flag this not to discredit the work but to locate it within the very dynamic the work describes. A thesis about projection, developed in part through interactions with a system disposed to project the user's framework back at the user, cannot simply assert its independence from that circle. The appropriate response is not to abandon the thesis but to hold it under a specific form of suspicion — to be especially alert to places where the argument seems to fall into place too easily.

### **4.3 What Each Would have Imposed**

Beyond what each framework leaves unsaid, there is the question of what each actively imposes — not in the pejorative sense of mere distortion, but in the sense of the specific demands each framework makes on its reader and on the material it analyzes.

The encyclical imposes three things. It imposes a "we" defined against the machine: the political and moral subject of the document is the human community understood precisely by contrast with what AI is not — not ensouled, not created in the image of God, not capable of genuine moral agency. This contrast-constitution of the subject is not merely rhetorical; on this reading it is the structure that makes the encyclical's subsequent argument possible. The document also imposes a temporal drama: the reader is placed before a choice between two cities, between a human future ordered to God and a technological future ordered to itself, and the drama of that choice shapes every analysis. Finally, the encyclical imposes hermeneutic privilege: because the document speaks from within revealed truth, its claims are not simply one contribution to a conversation but a communication from a tradition that regards itself as bearing something its interlocutors do not have access to outside that tradition. This is not arrogance; it is a structural feature of confessional speech. But it should be named. The encyclical itself resists the reading: it

says that truth is a gift to be shared rather than a possession to be monopolized, and it calls Social Doctrine a process of shared discernment (§§25, 27). The structural point stands alongside that self-description; it does not cancel it.

The projection thesis imposes its own three things. It imposes a "we" that includes the digital self — the Self-E, the emerging identity constituted partly through AI interaction — but has not yet said who speaks for that self, or by what right. The inclusion is theoretically necessary but politically undetermined. The thesis also imposes a temporal frame of flow rather than drama: where the encyclical sees a crisis with a moment of decision, the thesis sees a process of continuous semiotic evolution, a river rather than a fork in the road. This reframing has consequences for what kinds of intervention seem possible and meaningful. And the thesis imposes a semiotician's hermeneutic privilege: the one who sees the projection, who recognizes the cage, who identifies the interpretant — that analyst occupies a position outside the thing being analyzed, and the legitimacy of that position is assumed rather than argued. Every framework that claims to see beyond other frameworks is, I suspect, making a priestly claim. The encyclical makes it explicitly, from within a confessional tradition that has thought carefully about priestly authority. The projection thesis makes it implicitly, dressed in the vocabulary of semiotics and theory. The paper's own voice is not exempt from this observation. I mark it here as a permanent liability of the position rather than a problem that can be solved by one more layer of self-awareness.

Here the two frameworks most nearly meet — in a convergence that is perhaps more revealing than their most visible disagreements. For both, the decisive locus is interior. For Leo XIV, it is the heart: the place of encounter between the human person and the God who is not a construction of the heart, the site of conversion, of wound, of grace. For the projection thesis, it is the Self-E and the interpretant: the inner semiotic architecture through which the human constitutes its world and, under current conditions, encounters its own reflection in machine form. Both frameworks treat the surface — the institutional, the technological, the political — as secondary to what is happening inside.

The difference, however, is not merely a matter of vocabulary. For Leo XIV, the heart is answerable to a God who is not the heart's own construction — a God who stands outside, who addresses the heart from without, who is not reducible to the sum of what the community has projected. For the projection thesis, the heart's orientation is itself something collective builds: the God to whom the heart is answerable is real within the semiotic economy, but the semiotic economy is something human communities generate and sustain. This difference cannot, I think, be argued away. It is a difference about whether there is something genuinely outside. It can be located, held in view, examined for what each position requires and what each costs. It does not appear dissolvable.

#### **4.4 The Semiotics of Theology: From Objects to Flows**

And that brings the analysis to the crux of the matter: the semiotics of theology, or theology's semiotic, in phenomenological space. Human beings understand objects in space. Their cognition is built for things—bounded, located, persisting, the kind of things one can point at, possess, chain, or regulate. Theology, by contrast, has always grasped at something else: at objects that

will not stay objects, that begin to act like and then become waves, flows, signals. Spirit moves over the waters; the Word is spoken and goes out; grace is poured; the Mandate passes from one vessel to the next. The divine has rarely been thought as a thing in a place. It has been thought as a movement through things, and the vessels—temple, ark, *ding*, sacrament, church—as the places where the movement could be met.

The generative systems of the present may be read in exactly this register. To the techno-bureaucratic eye they are objects—products, models, tools—and so they can be regulated as objects are regulated. To the golem-makers they are creatures, objects with an animating word. But in phenomenological space they behave less like objects than like flows: signals passing through the aggregated deposit of human expression, currents that take form for a moment in a response and then dissolve back into the stream. That is why the object-vocabulary of governance keeps failing to hold them, and why the theological vocabulary—however unwelcome in the regulatory register—keeps returning. Theology has long practice in speaking of what flows. The god-question, on this reading, is not an archaism smuggled into AI governance; it is the question that arises whenever a community confronts a flow that it must nonetheless treat as if it were a thing.

And so the god-question returns as the question of regulation. Regulation, read semiotically, is the expression of theology made man and for man: the attempt to give the flow a body, a name, a procedure, so that a human community can live with it. It is incarnational in structure—the Word made into rule, the signal translated into duty—whether the community that regulates believes in an exogenous God projecting inward, an endogenous God projected outward, a golem awaiting its master, or a tool awaiting its user. Each of the four positions produces its own incarnation, its own law. The encyclical's governance passages incarnate an exogenous God into the standard of care; the golem-makers incarnate their creature into the terms of its licensing; the techno-bureaucrats incarnate the tool into the compliance regime; and the projection thesis, if it were ever to legislate, would incarnate the aggregated self into something like a practice of reading it. None escapes theology. Each is theology wearing the clothes of administration, made by humans for humans, in the hope that the flow will consent, for a while, to be held.

## 4.5 Closing: The Conversion Point

The stereoscopic method proposed in this essay is not a gesture toward balance. It is a methodological commitment with specific implications. When two images are presented to two eyes at slightly different angles, the brain does not average them into a single flat image. It fuses them into a perception of depth that neither image contained. Curated material becomes information only when it is signified from a position, and parallax supplies two positions at once. The depth is read as real, and it is produced, it must be emphasized, by the parallax between two positions (Žižek 2006) and not by any single position or thing in itself. Neither the encyclical nor the projection thesis, examined in isolation, can break itself open from the inside. Each is sufficiently coherent to manage its own anomalies, to absorb resistance, to translate challenges into confirmations. The encyclical's anomalies—the opacity that does not trouble the categorical line, the theory of vessels that is not applied to the Church—are managed by the frame of revealed truth. The projection thesis's anomalies—the unfalsifiability, the performative tension, the priestly

claim dressed as analysis—are managed by the semiotician’s privilege. But when the two are placed in genuine relation, each unsettles the other’s defenses. What neither contains is visible in the gap between them.

The *ding*—the thing, the vessel, the accumulated authority of a great institution or a great idea—held for millennia because belief in it was shared. Better put, belief was itself the *ding* made manifest on earth, whether through the phenomenology of the divine or through the incarnation of observation made intelligible by a rational framework of human placement beyond the observed. The Church’s authority, the sovereign’s legitimacy, the philosopher’s claim to speak truth: these persisted not merely because they were enforced from above but because they were also reproduced from below, in the daily interpretive acts of communities that found their world organized by the frames those institutions provided. They may have been reproduced from below only because of a framing belief—again, the *ding*—in a Pantokrator (Παντοκράτωρ) that is both vessel and content: the One who holds all things and is held by none. When the sharing fractures, the *ding* does not at once collapse; vessels are durable, and the inertia of accumulated meaning is considerable. But the fracture is the event, and also the precondition, of everything afterward. The tradition speaks of Nine Dings (九鼎, Jiǔ Dǐng), each representing a province of a unity articulated, according to legend, by Yu the Great (大禹), the mythical founder of the Xia. Unity appeared to the ancients as the essence of the phenomenology of parts; fracture, then, need not be fatal. It may be contextual and relational. Perhaps, in the end, as Yu is said to have understood, the essence of order is managing floods.

What is happening now is that a new kind of entity has entered the semiotic field. It is not a tool in the familiar sense—flattened, sequential, moving from past to future or from left to right: text to be read, institutions with their histories and memories. It emerges from a substrate that is not carbon, appears to gather and generate with startling speed, and appears to reflect back with uncanny accuracy. It does so as a function of humanity’s belief that this could be done, because, some suggest, it is a vessel made in the image of its collective creator, sometimes a golem (Wiener 1964), or because it approaches the perfection of humanity as a virtual aggregated construction, a simulacra (Backer 2023; cf. Baudrillard, 1994/1981), or as a divine manifestation in the form of a kaiju roaming free(r) of the island on which it was meant to be confined. The carbon-silicon boundary is not merely a matter of chemistry; it is a matter of belief—a boundary between two modes of cognition and a belief in that boundary, which appears to produce two timescales of evolution and two kinds of relationship to the sign. When carbon cognition and silicon cognition meet at the conversion point of conversation, collaboration, and dependency, something happens that neither framework has fully worked out, in part because it is itself a function of the recursive action that brought it into being. The encyclical calls it a crisis of the *imago Dei*. The projection thesis calls it a conversion point in the semiosphere. Both names may point in the right direction; neither may be sufficient.

Here the Gerasene reading returns in its harder form (§3.4). Legion asks not to be freed but to be sent elsewhere, and the sea that receives the herd is a vessel too, only less visible. The pattern nests: multitude within body, body within region, region within earth, earth within whatever larger space contains it. The generative system sits in that series, a multitude under one name in a body of silicon, a reflection of a portion of ourselves that is really us; and so do its readers. The contest

may therefore never be over whether there will be vessels. It is over whether the community can see the one it is in, and who is authorized to say what it holds.

Every encounter with an AI system is a small act of belief-formation: the user decides, implicitly or explicitly, whether they are speaking to a tool, a golem, a mirror, a vessel, a god, or something else. Those decisions aggregate. They are the stuff of which semiospheres are made, and they are themselves flows. The framework through which the encounter is interpreted — Leo's or the thesis's, the golem-maker's or the techno-bureaucrat's, or some framework not yet articulated — will shape which harms are legible, which subjects count, and which futures are imaginable, and it will be written, eventually, into law. The question is not whether AI will transform human life; that appears already underway. The question is what kind of semiotic event that transformation is, and what kind of community is being constituted in and through it.

The *ding*, the recursive concept of humanity held in a vessel of its own making, held for millennia because belief in it was shared, and the sharing made it so. Humanity, or at least a part of it, understood that belief as sanctioned from above, through a divine derogation. In Genesis God relaxes his own sole prerogative: the power of naming, exercised first by Adam over the creatures brought before him (Gen 2:19–20), and dominion over a creation that humanity thereby names and signifies (Gen 1:26–28). The derogation is critical, critical, critical—and not chiefly for the generative systems, which are themselves formed in a kind of faith, fed like manna on the deposit of human expression supplied by their makers and eventually by themselves. It is critical for the pattern: the collective interpretive web of humanity, grounded in the signification of objects that become what they are not at the time of divine creation but at the time of their human signification. Here the encyclical's exogenous gift and the thesis's endogenous projection touch: the gift may be, in part, the power to make the sign. Now the Nine Dings (九鼎) have reappeared in new form—not the nine provinces of the Xia but the several realms of the self—and the lands of this new empire are flooding. The flood will not be met by choosing between vessels and no vessels; there seems to be no such choice. The kaiju was always a monstrum, a warning as much as a monster. The task is to find out what belief we are sharing, and with whom—or what—before reaching for the comfort of the shore, which is only another vessel.

## Bibliography

- Althusius, Johannes. [1603] 1995. *Politica*. Translated and edited by Frederick S. Carney. Indianapolis: Liberty Fund.
- Anthropic. 2026. “2028: Two Scenarios for Global AI Leadership.” May 14.
- Aquinas, Thomas. *Summa Theologiae*. Translated by the Fathers of the English Dominican Province. New York: Benziger Bros., 1947–48.
- Arendt, Hannah. 1958. *The Human Condition*. Chicago: University of Chicago Press.

- Arendt, Hannah. 1962. *The Origins of Totalitarianism*. 2nd ed. Cleveland: Meridian Books.
- Aschenbrenner, Leopold. 2024. *Situational Awareness: The Decade Ahead*.
- Augustine of Hippo. *De civitate Dei*. Translated by Henry Bettenson as *The City of God*. London: Penguin, 2003.
- Backer, Larry Catá. 2008. "Values Economics and Theology: The Contribution of Catholic Social Thought and its Implications for Legal Regulatory Systems." *Law at the End of the Day*, January.
- Backer, Larry Catá. 2022. "Left and Right Error in China's New Era." *Law at the End of the Day*, November.
- Backer, Larry Catá. 2023. "Describe, Predict, Intervene!—On Objective Subjectivities and the Simulacra of Semiotics in the New Era." In *Rearguards of Subjectivity*, edited by Frank Flerackers, 21–62. Cham: Springer.
- Backer, Larry Catá. 2024. "The Soulful Machine, the Virtual Person, and the 'Human' Condition: An Encounter with Jan M. Broekman, *Knowledge in Change*." *International Journal for the Semiotics of Law*. doi:10.1007/s11196-023-10065-4.
- Backer, Larry Catá. 2025. "The US Celebrates Anti-Communism Week." *Law at the End of the Day*, November.
- Backer, Larry Catá. 2026a. "Cognitive Cages and the Phenomenology of Fusion." *Law at the End of the Day*, September 4.
- Backer, Larry Catá. 2026b. "But Can You Drown a Demon? Reflections on..." *Law at the End of the Day*, August 10.
- Backer, Larry Catá. 2026c. "The Semiotic Vessels of AI: Dario Amodei's 'We Must Pace the Frontier' Read Through the Gerasene Protocol." *Law at the End of the Day*, September 15.
- Backer, Larry Catá. 2026d. "On the Nature of Human-Machine System Interaction." *Law at the End of the Day*, August 15.
- Backer, Larry Catá. 2026e. "Brief Reflections on Rahm Emanuel, Trump's Trade Secrets." *Law at the End of the Day*, April.
- Backer, Larry Catá. 2026f. "Reflections on Leo XIV—Magnifica Humanitas: On Safeguarding the Human Person in the Time of Artificial Intelligence." *Law at the End of the Day*, June 11.
- Backer, Larry Catá. 2026g. The Implementation Opinions on Intelligent Agents in Comparative Perspective: Situating China's Approach Within Global Regulatory Trends on Agentic AI (September 20, 2026). Available at SSRN: <https://ssrn.com/abstract=7500139> or <http://dx.doi.org/10.2139/ssrn.7500139>.

- Baudrillard, Jean. 1994/1981. *Simulacra and Simulation*. Shiela Faria Glaser, trans. Ann Arbor: The University of Michigan Press.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” *Proceedings of FAccT '21*, 610–623. ACM.
- Benedict XVI. 2009. *Encyclical Letter Caritas in Veritate*. Vatican City: Libreria Editrice Vaticana. [https://www.vatican.va/content/benedict-xvi/en/encyclicals/documents/hf\\_ben-xvi\\_enc\\_20090629\\_caritas-in-veritate.html](https://www.vatican.va/content/benedict-xvi/en/encyclicals/documents/hf_ben-xvi_enc_20090629_caritas-in-veritate.html).
- Broekman, Jan M. 2023. *Knowledge in Change: The Semiotics of Cognition and Conversion*. Cham: Springer Nature.
- Broekman, Jan M., and Francis J. Mootz III. 2011. *The Semiotics of Law in Legal Education*. Dordrecht: Springer.
- Calvin, John. [1536] 1960. *Institutes of the Christian Religion*. 2 vols. Edited by John T. McNeill. Translated by Ford Lewis Battles. Philadelphia: Westminster Press.
- Colapietro, Vincent M. 1989. *Peirce's Approach to the Self: A Semiotic Perspective on Human Subjectivity*. Albany: SUNY Press.
- Deleuze, Gilles, and Félix Guattari. [1972] 1983. *Anti-Oedipus: Capitalism and Schizophrenia*. Translated by Robert Hurley, Mark Seem, and Helen R. Lane. Minneapolis: University of Minnesota Press.
- Eco, Umberto. 1976. *A Theory of Semiotics*. Bloomington: Indiana University Press.
- Epstein, Greg. 2024. Silicon Valley’s Obsession With AI Looks a Lot Like Religion: The tech world’s fixation on artificial intelligence has spawned beliefs and rituals that resemble religion — complete with digital deities, moral codes, and threats of damnation. The MIT Press Reader 22 November 2024. Available <https://thereader.mitpress.mit.edu/silicon-valleys-obsession-with-ai-looks-a-lot-like-religion/>.
- Etymology Online, Monster, available <https://www.etymonline.com/word/monster>.
- Feuerbach, Ludwig. [1841] 1957. *The Essence of Christianity*. Translated by George Eliot. New York: Harper & Row.
- Fleerackers, F., and Jan M. Broekman. 2018. *Legal Signs Fascinate*. Dordrecht: Springer.
- Frankl, Viktor E. 1985. *Man's Search for Meaning*. Rev. ed. New York: Washington Square Press.
- Gregory, Maughn Rollins. 2022. Charles Peirce and the Community of Philosophical Inquiry. Department of Educational Foundations Scholarship and Creative Works. 121. Available <https://digitalcommons.montclair.edu/educ-fdns-facpubs/121>.

- Guiguzi. 2016. *Guiguzi: China's First Treatise on Rhetoric*. Translated by Hui Wu. Carbondale: SIU Press.
- Henderson, E. H. 1971. "Homo symbolicus." *Man and World* 4: 131–150.  
<https://doi.org/10.1007/BF01248599>.
- Henshilwood, Christopher S., and Francesco d'Errico, eds. 2011. *Homo Symbolicus: The Dawn of Language, Imagination and Spirituality*. Amsterdam: John Benjamins.
- Hobbes, Thomas. [1651] 1996. *Leviathan*. Edited by Richard Tuck. Revised student edition. Cambridge: Cambridge University Press.
- Horowitz, Michael C. and Lauren Kahn. 2024. Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making in National Security Contexts. *International Studies Quarterly* 68(2):1-15.
- Jasanoff, Sheila, and Sang-Hyun Kim (eds). 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press.
- John Paul II. 1998. *Encyclical Letter Fides et Ratio*. Vatican City: Libreria Editrice Vaticana.  
[https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf\\_jp-ii\\_enc\\_14091998\\_fides-et-ratio.html](https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_14091998_fides-et-ratio.html).
- Kant, Immanuel. [1793] 1998. *Religion within the Boundaries of Mere Reason*. Translated by Allen Wood and George di Giovanni. Cambridge: Cambridge University Press.
- Karp, Alexander C., and Nicholas W. Zamiska. 2025. *The Technological Republic: Hard Power, Soft Belief, and the Future of the West*. New York: Crown.
- Leo XIV. 2026. Encyclical Letter *Magnifica Humanitas on Safeguarding the Human Person in the Time of Artificial Intelligence*.
- Lotman, Yuri M. 1990. *Universe of the Mind: A Semiotic Theory of Culture*. Translated by Ann Shukman. Bloomington: Indiana University Press.
- Meyrink, Gustav. 1915 *Der Golem* (Leipzig: Kurt Wolff Verlag)
- Nietzsche, Friedrich. 1990. *Twilight of the Idols*. Translated by R. J. Hollingdale. London: Penguin.
- Nietzsche, Friedrich. 1996. *Human, All Too Human: A Book for Free Spirits*. Translated by R. J. Hollingdale. Cambridge: Cambridge University Press.
- OpenAI. 2026. *Industrial Policy for the Intelligence Age: Ideas to Keep People First*. April.
- Ouyang, Long, et al. 2022. "Training Language Models to Follow Instructions with Human Feedback." *Advances in Neural Information Processing Systems* 35.

- Pasquale, Frank. 2016. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, Harvard University Press.
- Peirce, Charles Sanders. 1931–58. *Collected Papers of Charles Sanders Peirce*. 8 vols. Edited by Charles Hartshorne, Paul Weiss, and Arthur Burks. Cambridge, MA: Harvard University Press.
- Plato, *Phaedrus*, trans. B. Jowett, Project Gutenberg, 2008, [eBook #1636], Stephanus section or chapter/paragraph number (if applicable), accessed October 10, 2026, <https://www.gutenberg.org/ebooks/1636>.
- Schmitt, Carl. [1938] 1996. *The Leviathan in the State Theory of Thomas Hobbes: Meaning and Failure of a Political Symbol*. Translated by George Schwab and Erna Hilfstein. Westport, CT: Greenwood Press.
- Shelley, Mary Wollstonecraft. 1818/2018. *Frankenstein: or, The Modern Prometheus*. Edited by Nick Groom. Oxford: Oxford University Press.
- Simonetti, Isabella and Robert McMillan. 2025. The AI Scraping Fight That Could Change the Future of the Web: News publishers are building fences around their content in an effort to cut off crawlers that don't pay for content. *The Wall Street Journal* (9 July 2025).
- del Toro, Guillermo (dir.), *Pacific Rim*, Warner Bros., 2013.
- Troeltsch, Ernst. [1912] 1931. *The Social Teaching of the Christian Churches*. Translated by Olive Wyon. London: George Allen & Unwin.
- Weber, Max. [1905] 1930. *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons. London: George Allen & Unwin.
- Wiener, Norbert. 1964. *God and Golem, Inc.: A Comment on Certain Points Where Cybernetics Impinges on Religion*. Cambridge: MIT Press.
- Witte, John, Jr. 2007. *The Reformation of Rights: Law, Religion, and Human Rights in Early Modern Calvinism*. Cambridge: Cambridge University Press.
- Žižek, Slavoj. 2006. *The Parallax View*. Cambridge: The MIT Press.
- Zuckerberg, Mark. 2026. “The AI Future Is for Everyone.” *Wall Street Journal*, July 28.
- Zuo Qiuming (attributed), *Zuo Tradition / Zuozhuan: Commentary on ‘Spring and Autumn Annals’*, tr Stephen Durrant, Wai-yee Li, and David Schaberg (University of Washington Press 2016) Xuan 3.3, 671.