Instruments of Virtue

A Thomistic Study in AI Systems Design

Rosaire Leon[1]

Disclosures: The author is employed as a senior model engineer at New Sapience.

Acknowledgements: Thanks to Emilia Stefankiewicz for her assistance in the editorial process, and to Patrick Misiewicz for his initial proposal to develop a Thomistic evaluative framework for AI systems by studying New Sapience's "sapiens" architecture.

[1] Correspondence to: Rosaire Leon, <rleon@newsapience.com>

Key Terms[2]

Statistics: the branch of mathematics concerned with collecting quantitative data, analyzing it, and using it to derive probable inferences through aggregation and correlation.

Stochastics: the study of systems whose behavior is governed by probability distributions; in the context of computing, a *stochastic* system is one whose outputs are not determined by fixed rules, but by extracting random samples from tabulated probabilities.

Determinism: in computing, this term characterizes a system in which identical inputs always yield identical outputs, with no random variation.

Generative AI: a stochastic AI system architecture in which the machine uses statistical patterns derived from digital libraries of human-authored textual, visual, or audio products to generate outputs resembling such products.

Large Language Model (LLM): a generative AI system that decomposes input text into sequences of word fragments ("tokens"), compares these sequences to statistical patterns derived from large quantities of human-authored text, and generates output text by computing which fragment most probably follows in sequence, given the patterns scanned, the input received, and the output fragments already generated.

Sapiens: a deterministic AI system architecture in which the machine executes formalizable aspects of human intelligence, organizes knowledge hierarchically across levels of experience and abstraction, and depends on human agents to supply explicit operating principles and practical judgment.

"Cognitive": a commonplace term in AI discourse, but Aquinas uses it and its relatives variously in the *Summa Theologiæ*. In contemporary usage, "cognitive" and "cognition" tend to connote an undifferentiated or incompletely differentiated conception of all human sapiential faculties (external senses, internal senses, *intellectus*, *ratio*, et cetera). We will avoid this vocabulary in favor of precise Thomistic terms, even in contexts that do not obviously evoke theology. "Cognitive" calls up confusing associations with modern "cognitive neuroscience," which is irrelevant here. Sapiens are not purported to replicate undifferentiated human "cognition," personality, or hypostaseity, but only to instrument particular rational operations.

---

[2] **Abbreviations for Aquinas's cited works:** *ST*: *Summa Theologiæ*. *DV*: *De Veritate*. *SCG*: *Summa Contra Gentiles*. *LC*: *Super librum De causis expositio.*

***Intellectus* (intellect):** the intellectual operation that grasps principles immediately, without discursive justification.[3]

***Ratio* (discursive reasoning):** the rational operation that moves from principles to conclusions, antecedents to consequents, through stepwise inference.[4]

***Anima rationalis* (the rational soul):** the complete rational faculty, which includes both *intellectus* and *ratio* as operations of the same intellectual power of the same *anima rationalis*.

---

[3] *ST* I–II, q.57, a.2: "Quod autem est per se notum, se habet ut principium; et percipitur statim ab intellectu. Et ideo habitus perficiens intellectum ad huiusmodi veri considerationem, vocatur intellectus, qui est habitus principiorum."

[4] *ST* I, q.79, a.8: "Ratiocinari autem est procedere de uno intellecto ad aliud, ad veritatem intelligibilem cognoscendam.…Patet ergo quod ratiocinari comparatur ad intelligere sicut moveri ad quiescere, vel acquirere ad habere."

Part I: Instrumentality, Chrēsis, and Complementarity

Contemporary discourse about AI routinely conflates distinct rational operations through uncritical repetition of euphemisms. It's primarily this use of euphemisms, and not the sophistication of the devices themselves, that makes it difficult for non-specialists to understand what generative AI systems actually do. The result is a lot of confusion about what we are dealing with in AI.

Since generative systems are being "procured" and "deployed" everywhere, such confusion significantly threatens human life and dignity. Different AI architectures have, and cultivate, fundamentally different relationships to truth and moral responsibility. If we don't understand what they're doing, then we don't understand what we're doing. Even that might be tolerable if the tasks proposed for AI were merely trivial, but that isn't what's happening: instead, so-called "agents" have fallen from the sky with alacrity, and they are now appearing in hospitals, universities, police precincts, immigration offices, policy-writing organs of the administrative state, courts of law, military command centers, and more besides. The moment presses: Do we know what we're doing?

Even if we don't, the Church has shown that she does, and not only by organizing this conference. In *Antiqua et Nova*, the Holy Office counseled that we distinguish human intelligence—which grasps truth as such through *intellectus* and reasons discursively through *ratio*—from AI systems that simulate these operations through statistical inference without performing either. Because moral causality belongs only to personal beings, decisions that touch human dignity must in all cases remain with human agents who alone can exercise conscience and prudence. The note warns that the opacity of current AI systems makes accountability obscure precisely where it is most needed.[5]

From St. Thomas Aquinas, we inherit the concept of instrumental causality: an instrument may participate in an act whose origin and end both lie beyond it, yet contribute to that act through its own proper power. When a sculptor cuts marble with a chisel, the work he applies and the form he intends are both extraneous to the chisel, and the chisel only cuts the marble according to its own material properties. Yet the chisel does contribute to the sculpture, indeed vitally so, thanks to the relationship it is *in* with both the sculptor and the marble. The principal agent works through an instrument that remains subordinate to him; even so, the instrument leaves its own signature on the finished work.

Through his recovery of the Greek concept of *chrēsis* ("use"), the Italian philosopher Giorgio Agamben offers a reading and development of Aquinas on instrumentality that bears fruit for

---

[5] Dicastery for the Doctrine of the Faith, Dicastery for Culture and Education, *Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence* (14 January 2025)

thinking about AI.[6] *Chrēsis* makes it possible to distinguish between two modes of instrumental relationship. What differs between them is not how the relationship is structured, but how that structure is *inhabited*. *Use* always affects both agent and instrument: neither emerges from use unmodified. The sculptor shapes the marble through his chisel, but at the same time, his work rebounds on him. By shaping the marble, he shapes himself: his grip improves, his technique develops, his instincts for matter, form, and his own powers deepen and mature *through* the ways instrument and medium both supply his art and resist it. Certainly, the chisel extends the sculptor's power and skill. But it also trains and refines the former into the latter. In turn, through use, the chisel acquires its particular material history and character: physical traces imparted to it by the sculptor's own habitual modes of application.

Joining Aquinas's and Agamben's concepts, we propose a tentative *principle of instrumental complementarity* as our place to begin: a machine that genuinely serves reason will be one that occasions reflective growth in the human principal who uses it. If a machine works authentically as an instrument through which the human intellect may extend its own proper power, then *use* of that machine should make the principal agent's knowledge transparent to itself by drawing it beyond what unaided reflection typically achieves.

LLMs simulate rational competence by producing outputs that resemble the products of human speculative and practical reason, but all such outputs are generated through a single sub-rational procedure: stochastic association.[7] What's more, because these machines, at least in their current iteration, utilize "attention" mechanisms that no human being can fully trace, they *cannot but* lend themselves to abuse by agents seeking an algorithmic alibi to evade legal, political, moral, emotional, and other kinds of liability.[8] These are inerradicable affordances of the machines themselves, whatever their architects' intentions. By design, LLMs and other generative AI systems function in ways that diffuse accountability, degrade expertise, distort and erode human relationships, and disintegrate the legitimacy of institutions on which a social organism depends.[9]

---

[6] Agamben, *The Use of Bodies* (2015); ibid., *Opus Dei: Toward an Archæology of Duty* (2013)

[7] Floridi et al., "What Kind of Reasoning (if any) is an LLM Actually Doing? On the Stochastic Nature and Abductive Appearance of Large Language Models (revised)" (15 December 2025)

[8] Zhang, "Attention Is Not What You Need: Grassmann Flows as an Attention-Free Alternative for Sequence Modeling" (22 December 2025)

[9] Hartzog and Silbey, "How AI Destroys Institutions" (forthcoming, draft accessed 13 January 2026)

With an LLM, the instrumental relationship is not merely concealed but progressively inverted.[10] Through his engagement with this machine, the human person becomes more or less unwittingly both instrument and patient of a statistical procedure without a determinate agent; and this happens precisely because LLMs do not participate in reason. Even when we intend to use such systems for genuine reasoning tasks, we find ourselves enlisted, knowingly or not, in a sub-rational simulation of reason. These machines excel only in *manipulating* information without in any sense *comprehending* it. The difference between knowledge, fiction, and lies, or more fundamentally that between *what is said*, *what is understood*, and *what is*, accordingly finds no purchase in them.

By contrast, a deterministic architecture that directly formalized structural features of human knowledge would yield an AI system that would remain fully subordinate to its human principal agents, transparent to human inspection, and dependent on human practical judgment. By instrumenting certain rational operations without simulating independent judgment, such a system would make human rational and moral agency absolutely and irreducibly explicit in every use of AI tools.

It happens that at least one such architecture has already been substantially developed. An American technology company named New Sapience has designed what it calls "sapiens": an AI system operating on a deterministic architecture that distinguishes *what is said* from *what is meant*, and organizes knowledge hierarchically across different levels of human experience and abstraction.

The reciprocal nature of interaction between human principal agents and sapiens expresses the design's core insight, and contrasts with the obscurity and asymmetry of the only relationship a human being can have with an LLM or other generative device. Insofar as human agents learn by using sapiens, while sapiens learn by being used, the relationship between agent and instrument here actively precipitates human introspection and the cultivation of reflective wisdom. Hence the design's name: both the architecture and its instances are called "sapiens" not for what these devices are puported to *be* or *do*, but for what they are intended to serve.

Yet, to discern whether sapiens actually achieve instrumental complementarity, we must know precisely what they do, how they exist, and what kind of relationship they have with human rational faculties. To that end, in what follows, we identify which rational operations they can perform and in what sense, in order to arrive where their limits expose what only a human being can do or supply.

---

[10] Knox, Bradford, et al., "Harmful Traits of AI Companions" (18 November 2025)

Threshold: Sapiens—The System Architecture

The Commonsense World Model (CWM) contains three domains: the linguistic (what is said), the ontological (what is meant), and the epistemological (what interprets).

Separating the linguistic domain from the ontological domain enables the system to map a natural-language dictionary to modeled ontology without conflating the two, while the epistemological domain situates the ontological objects in and among the "lamina," layers of knowledge. These lamina range from the quantum → atomic → molecular (empirical lamina); to the physical → system → suprasystem (objective lamina); to sensation → perception → sentience → sapience (subjective lamina).

The role of the lamina is to specify how or indeed whether an object may be known at each of these levels, and how the different levels relate to each other. For example, one doesn't know a diamond as carbon atoms arranged in a lattice unless and until one engages the atomic lamina. On the level of physics, one knows a diamond as crystalline, translucent, hard; on the level of sentience, one knows it as beautiful (or not), elegant (or not), gaudy (or not); and so on. Each of these *is* the (knowable) diamond. The lamina mutually inflect and converse with each other, and are not exclusive of each other, but they are accessed in distinct contexts and some—such as the atomic or quantum—only by particular technical means. The lamina *supervene* upon each other: a difference on any one level makes a corresponding difference on each.

The engine that interprets and operationalizes the CWM prioritizes the consistency of ontological statements so that all component objects are modeled on the same principles. Separating the epistemological domain from the ontological further enables the system to self-modify the CWM when presented with new knowledge that it needs to inspect and integrate. It first does so in a localized Conversation Model which may or may not be committed, in the end, to the user's personal version of the CWM.

Finally, separating the three domains architecturally enables the system to "learn" what new words mean from context. To determine the word is unknown, it uses epistemological data to inspect existing ontological and linguistic data for possible referents. It may then question the user directly about the meaning of the word, make an inference and request confirmation, or both.

These processes are handled by a proprietary software engine and programming language, both developed internally and patented by New Sapience.

Part II: Sapiens and Intellectus

1. Principles of Motion

What sort of being does the sapiens architecture describe?

Aquinas distinguishes among beings by their principle of motion.[11] Things moved by external force —projectiles—differ from things moved by internal principle. Among things moved internally, those moved by natural necessity (fire rising) differ from those moved by appetite (animals seeking food) and from those moved by a free rational appetite, a will (humans and angels choosing the good).[12]

Sapiens operate deterministically on programming that human beings have designed. They have no internal principle of motion, no will to freely orient towards any end. The critical point: in Aquinas, complete *intellectus* isn't a mechanical operation directed by an exterior agent, but an act of a living rational soul, possessing its own rational appetite (will) and its own proper end (nature).[13] A sapiens possesses neither: it has no will and has its sole end in another than itself, the human principal agent for whose sake it exists.

Now, Aquinas holds that *intellectus* of principles is immediate and infallible regarding its proper objects. The possible intellect grasps a principle, a universal nature, as soon as it grasps the species that signify it. Through certain architectural decisions, human beings can formalize aspects of this act for computation. However, once *intellectus* grasps a principle, judging it *worthy to hold as an axiom* in consequent reasoning involves an independent commitment of the intellect. This is an intentional act attesting a fundamental orientation towards reality—an *appetite* for truth, which presupposes will. Here the will, as a second power at work *within* the rational soul, bears efficiently

---

[11] *ST* I, q.18, a.1: "Primo autem dicimus animal vivere, quando incipit ex se motum habere....Illa proprie sunt viventia, quæ seipsa secundum aliquam speciem motus movent"; a.2: "Vivere nihil aliud est quam esse in tali natura"; a.3: "Vita maxime proprie in Deo est."

[12] *ST* I–II, q.6, a.1: "Quæ vero habent notitiam finis dicuntur seipsa movere, quia in eis est principium non solum ut agant, sed etiam ut agant propter finem. Et ideo, cum utrumque sit ab intrinseco principio...horum motus et actus dicuntur *voluntarii*, hoc enim importat nomen *voluntarii*, quod motus et actus sit a propria inclinatione"; a.2: "Soli rationali naturæ competit voluntarium secundum rationem perfectam, sed secundum rationem imperfectam, competit etiam brutis animalibus."

[13] *ST* I, q.82, a.4: "Alio modo dicitur aliquid movere per modum agentis; sicut alterans movet alteratum, et impellens movet impulsum. Et hoc modo voluntas movet intellectum....Hæ potentiæ suis actibus invicem se includunt, quia intellectus intelligit voluntatem velle, et voluntas vult intellectum intelligere."

upon the intellect.[14] This isn't and can't be made a deterministic process: while exact definitions of "will" are rare in the history of philosophy, it's clear in Aquinas that an entity whose behaviors are fully determined by an exterior force can't be said with any coherence to possess or exercise a "will."

This much begins to distinguish complete, independent *intellectus* as an act of a rational soul from apparently similar operations in artifacts. Now we draw this distinction to lapidary precision.

2. Natural Versus Artificial Faculties

Sapiens' faculties are all imparted by human design; they are neither innate nor emergent. This evokes Aquinas's distinction between natural and instrumental faculties.

In human beings, the agent intellect (*intellectus agens*) and possible intellect (*intellectus possibilis*) are both intrinsic faculties of the rational soul.[15] They belong irrevocably to human nature and develop by actualizing innate potential: a teacher only cultivates something already present in each rational soul, although it may be circumstantially frustrated. The initial principles of learning are discovered, not installed.

In sapiens, faculties are engineered and installed by an exterior agent. A sapiens does not, for example, discover its own innate principles of abstracting particulars or disambiguating references from context. Rather, human beings have devised formal techniques to implement such principles functionally in the system's mechanical design.

Aquinas might say: the nature of a thing determines the nature of its operations.[16] If a sapiens's "nature" is entirely constructed, then all of its operations are artifacts of that human labor. The system really does abstract, disambiguate, and so on, but these works are purely derivative. Human agents have formalized these operations to extend them *through* sapiens as instruments; sapiens do not act independently, appearances notwithstanding.

---

[14] *DV* q.22, a.12: "Intellectus aliquo modo movet voluntatem, et aliquo modo voluntas movet intellectum et alias vires. Ad cuius evidentiam sciendum, quod tam finis quam efficiens movere dicuntur, sed diversimode."

[15] *ST* I, q.79, a.3–4: "Oportebat igitur ponere aliquam virtutem ex parte intellectus, quæ faceret intelligibilia in actu, per abstractionem specierum a conditionibus materialibus. Et hæc est necessitas ponendi intellectum agentem"; "In parte nutritiva omnes potentiæ sunt activæ; in parte autem sensitiva, omnes passivæ; in parte vero intellectiva est aliquid activum, et aliquid passivum."

[16] *ST* I, q.75, a.3: "Enim unumquodque habet esse et operationem." *SCG* II, c.76: "Nihil operatur nisi per aliquam virtutem quæ formaliter in ipso est.…Numquodque quod non potest exire in propriam operationem nisi per hoc quod movetur ab exteriori principio, magis agitur ad operandum quam seipsum agat.…Si igitur intellectus agens est quædam substantia extra hominem, tota operatio hominis dependet a principio extrinseco. Non igitur erit homo agens seipsum, sed actus ab alio."

Does this mean sapiens are something like angels, at a lower ontological level? There's a weak analogy. Angels participate in divine *intellectus* without being divine.[17] They *possess* intellectual operations that image and derive from divine *intellectus*, yet they're creatures.[18] In an apparently similar way, sapiens participate in human *intellectus* without being human. They *perform* operations that externalize aspects of human *intellectus*, yet they're artifacts.

But it's more important to grasp the strict disanalogy. Angels are intellectual *substances* who possess their own will and their own complete, independent *intellectus*, which freely contemplates and intends the good: they choose the good through their own will.[19] Meanwhile, sapiens have no will and no complete, independent *intellectus*; that's why they're called "deterministic."

Determinism and obedience are not the same thing. Obedience is a voluntary act that proceeds from the will's own understanding of rightful authority.[20] An entity without a will can't *obey*; it can only *execute*, as sapiens execute operations which, moreover, they don't possess in themselves.

What's more, angels are individual persons, hypostases.[21] They grasp principles connaturally through their own intellectual light, without needing to learn or derive knowledge through inference as

---

[17] *ST* I, q.54, a.1: "Neque angeli, neque alterius creaturæ, sua actio est eius substantia.…In solo Deo sua substantia est suum esse et suum agere"; "Actio quæ transit in aliquid extrinsecum, est realiter media inter agens et subiectum recipiens actionem. Sed actio quæ manet in agente, non est realiter medium inter agens et obiectum, sed secundum modum significandi tantum, realiter vero consequitur unionem obiecti cum agente."

[18] *ST* I, q.55, a.1: "Angelus autem per suam essentiam non potest omnia cognoscere; sed oportet intellectum eius aliquibus speciebus perfici ad res cognoscendas."

[19] *ST* I, q.59, a.1: "Necesse est ponere voluntatem in angelis"; a.3: "Libertas arbitrii, cum sit in hominibus, multo magis est in angelis.…*ubicumque est intellectus, est liberum arbitrium*." I, q.79, a.1: "In angelis non est alia vis nisi intellectiva, et voluntas, quæ ad intellectum consequitur."

[20] *ST* II–II, q.104, a.2: "Obedientia, sicut et quælibet virtus, debet habere promptam voluntatem in suum proprium obiectum, non autem in id quod repugnans est ei. Proprium autem obiectum obedientiæ est præceptum, quod quidem ex alterius voluntate procedit."

[21] *ST* I, q.29, a.1: "Sed adhuc quodam specialiori et perfectiori modo invenitur particulare et individuum in substantiis rationalibus, quae habent dominium sui actus, et non solum aguntur, sicut alia, sed per se agunt, actiones autem in singularibus sunt. Et ideo etiam inter ceteras substantias quoddam speciale nomen habent singularia rationalis naturae. Et hoc nomen est persona"; a.3: "Persona significat id quod est perfectissimum in tota natura, scilicet subsistens in rationali natura."

humans do.[22] By contrast, sapiens process principles already grasped and modeled by humans: they *are* abstracted speculative operations, formalized structure, and crystallized *ratio*. The work is formally the same, but the ontological substrate differs utterly: one origin is divine, the other, human; one instance is personal, the other, mechanical. Sapiens are hypostatically empty, subsisting entirely through human rational agency.

2. Artificial *Intellectus*?

Sapiens do not, therefore, achieve genuine *intellectus*. And yet, they achieve operations that are effectively isomorphic to aspects of *intellectus* while lacking the ontological status that makes *intellectus* what it is. To understand what is happening here, we need to consider distinct operations within the human intellectual act.

Sapiens achieve the following operations of speculative reason:

- Through the three-domain architecture, they abstract universals from particulars.
- Through self-consistency constraints, they disambiguate and process logical relationships as *necessary*.
- Through the lamina hierarchy, they model knowledge about phenomena across different levels of experience and abstraction.
- Through rules of inference, they reason discursively from principles to conclusions (*ratio*).

These all proceed as designed and directed by human rational agency, not as proper acts of an independent rational agent.

However, sapiens do *not* achieve the following:

- They don't actually grasp principles *as such.* They execute formal rules, including heuristics that regulate how they select one formalism from among applicable alternatives by (a) interpreting operational context and (b) responding to human input.
- They don't judge that certain undemonstrated theses sufficiently commend themselves to postulate and utilize as self-evident axioms. They require, and are designed to solicit, human agents to make such decisions.

---

[22] *ST* I, q.55, a.2: "Species per quas angeli intelligunt, non sunt a rebus acceptæ, sed eis connaturales....Dato etiam quod posset abstrahere species intelligibiles a rebus materialibus, non tamen abstraheret, quia non indigeret eis, cum habeat species intelligibiles connaturales." I, q.58, a.3: "Si autem statim [hominum] in ipsa cognitione principii noti, inspicerent quasi notas omnes conclusiones consequentes, in eis discursus locum non haberet. Et hoc est in angelis: quia statim in illis quae primo naturaliter cognoscunt, inspiciunt omnia quaecumque in eis cognosci possunt."

- They don't commit to axioms as *true*. They process axioms supplied by humans and utilize them as *convenient*.

And they don't achieve any operations of practical reason. Those operations require:

- *Synderesis*: Natural inclination towards the good, which presupposes appetite.[23]
- *Prudentia*: Application of right reason to contingent particulars, which requires practical wisdom and virtuous perception developed through habituation.[24]
- *Conscientia*: Application of knowledge to particular acts by a moral agent.[25]
- *Ars*: Right reason about things to be made by an agent with creative intention.[26]

A sapiens lacks the basis for any of these. It has no appetite, no will, no nature that can be perfected or frustrated, and no proper ends. Hence, rather than attempt to formalize practical reason, and moral knowledge in particular, as a dimension of the Commonsense World Model, engineers model this as a uniquely human domain. The system solicits judgment from its principal agent regarding exceptions to general instructions while modeling the input received as *that agent's* particular judgment, not a formal norm. It can support agents' discernment by modeling consequences of choices through inference, but it can't judge—or pretend to judge—whether a choice conduces any genuine good.

A sapiens is a kind of intellectual projectile. It only externalizes those aspects of reason that are susceptible to formalization, and does so in a derivative and purely functional way. The architecture yields both an authentic rational instrument and an artificial intelligence in the only valid sense of the term, which is not an artificial *intellectus*, whether speculative or practical.

4. Instrumenting the Limits of Formalism

If an "artificial intelligence" isn't an artificial *intellectus*, how can we characterize what sapiens actually do relative to the human intellectual act?

---

[23] *ST* I–II, q.94, a.1: "Synderesis…est habitus continens præcepta legis naturalis, quæ sunt prima principia operum humanorum."

[24] *ST* II–II, q.47, a.2: "Prudentia est sapientia in rebus humanis, non autem sapientia simpliciter.…Ad prudentiam non pertinet nisi applicatio rationis rectæ ad ea de quibus est consilium"; a.15: "Igitur prudentia non est circa fines, sed circa ea quæ sunt ad finem…ideo prudentia non est naturalis."

[25] *ST* I, q.79, a.13: "Conscientia, proprie loquendo, non est potentia, sed actus.…Conscientia enim, secundum proprietatem vocabuli, importat ordinem scientiæ ad aliquid, nam conscientia dicitur cum alio scientia. Applicatio autem scientiæ ad aliquid fit per aliquem actum."

[26] *ST* I–II, q.57, a.3: "Ars nihil aliud est quam ratio recta aliquorum operum faciendorum"; a.4: "Ars est recta ratio factibilium; prudentia vero est recta ratio agibilium."

Ratio works on rules. Given *these* premises and *this* logical form, *that* conclusion follows. Whatever proceeds by steps from principles can be formalized for computation. For this reason, a sapiens can trace logical dependencies, validate inferences, and derive remote consequences with complete fluency. But because it only executes rules within a structure that human intelligence has designed, literally *in principle* the system never discovers anything new by itself. It can elaborate and explore the logical structure of any given situation to an arbitrary degree of complexity, but it can't independently conceive a new way of seeing.

An example from the history of modern mathematics shows the practical difference. When he conceived his non-Euclidean geometry, Nikolai Lobachevsky didn't just derive consequences from an already-modified antecedent, Euclid's parallel axiom. A sapiens could do that much, if it were asked —if, that is, it were given a modified axiom by a human agent, who might either supply it explicitly or do so implicitly by suspending the Euclidean one with a question: "What if parallel lines *didn't* form pairs of interior angles that sum to equal two right angles whenever a third straight line falls upon them both?"—or any equivalent formula.

This is a subtle but critical distinction. A sapiens needs to be prompted to "realize" that an axiom may be modified. The elements for modification must either be provided directly, or implied in the form of the question that asks "What if it weren't true?" In principle, the three-domain architecture enables the system to process whatever is asked in natural language and to infer what follows from the conditions described. Hence, in the form above, the alternatives are that the sum could be more or less than two right angles. Spherical and hyperbolic geometries readily follow.

Certainly, it isn't trivial that a mechanical system should be capable of making this sort of inference, which depends on sophisticated natural language processing, disambiguation from context, and something like conceptual engagement furnished by a world model—or in this case, a model of Euclidean geometry. But here also is the absolute difference: Nikolai Lobachevsky *asked himself* this question. He didn't only extract an inference from a prompt; he reflexively wondered whether Euclid's parallel axiom had to be true, conceived that it might be provisional, and for nothing more than his own interest set himself to reckon with an alternative: what if those pairs of interior angles equalled *less* than two right angles? What then? This is to say, Lobachevsky's reflexive act was more than formal processing.

Aquinas teaches that *intellectus* is the act of a rational soul reckoning with truth itself, through the participatory "light" that imbues its own nature. The soul has an *appetite* for truth; it grasps that

truth *is* and seeks it.[27] By contrast, a sapiens is designed to engage and track reasoning about phenomena as *consistent*, and only in that sense *truthful*.

We'll return to this theme shortly. For the moment, it's enough to settle our current preoccupation with the following: These formal limits are why sapiens can extend human deliberation without being able to substitute for human judgment. A sapiens executes *ratio* and certain formalizable operations of *intellectus*, but these all depend on and contribute to higher human intellectual acts. The human intellect, in its fullness, is wedded to the human appetite for truth, and a sapiens has no such appetite. Its relationship to the reality it models is neither appetitive, nor personal, subjective, objective, nor even interested. It's merely operative, as any traditional computer is operative, and it possesses no sentient interiority whatsoever—nor for that matter any exteriority, for the two come together.

Currently, the system includes no analogue to sensory exposure that would allow engagement with the phenomena its models describe and organize. But even if it were equipped with sensory apparatus, those faculties would have to be designed and encoded by external human agents like all the others. It would remain a transparent and impersonal artifact.

A sapiens is not "someone"; it exists and operates at an irreducible remove from the peculiar exposure to reality that characterizes a rational soul grasping *truth as such* and seeking it—*wondering* at it. Hence a sapiens doesn't and can't exist in the posture of receipt and response that specifies *intellectus* as a human faculty. But it can and does instrument certain fruits of that receipt, as a tool that explicates, extends, compares, and applies formal models of human knowledge.

Human *intellectus* is thus partially projected in sapiens, but the ontological principle remains fully human.[28] Precisely because human beings cannot create a truly independent intelligence, our artifacts can participate in our rational operations without themselves becoming rational souls. We can design systems that instrument all formalizable aspects of reason, but we cannot create minds.

5. Total Simulation Versus Partial Virtualization

An LLM simulates every rational act while performing none. Under that architecture, the machine executes a procedure that associates signs with statistics and parameters, computes probability

---

[27] *ST* I, q.16, a.1: "Veritas principaliter est in intellectu; secundario vero in rebus, secundum quod comparantur ad intellectum ut ad principium." I, q.82, a.2: "Sicut intellectus naturaliter et ex necessitate inhæret primis principiis, ita voluntas ultimo fini."

[28] *ST* III, q.62, a.1: "Principalis quidem operatur per virtutem suæ formæ, cui assimilatur effectus, sicut ignis suo calore calefaci....Causa vero instrumentalis non agit per virtutem suæ formæ, sed solum per motum quo movetur a principali agente. Unde effectus non assimilatur instrumento, sed principali agenti, sicut lectus non assimilatur securi, sed arti quæ est in mente artificis."

distributions about them, and reiterates until the output reaches a likely terminus. Critically, this means it uses data about what human beings have said—not what we've understood—to simulate our own understanding. This is like a photograph of fire: it looks like fire; it shows a fire once burned somewhere; but that "somewhere" was elsewhere and nothing is burning *here*.

Sapiens, too, depend on human *intellectus*, because they are machines and not rational souls. But what sort of dependence are we dealing with here? Sapiens *are* computable structures effectively isomorphic to human *ratio* and aspects of *intellectus,* projected through artificial means. Insofar as human rational acts had to design these operations and embed them in a modular data structure, human agency always *virtually* directs sapiens, at whatever remove, in the technical sense Aquinas imparted to "virtuality." This word designates the manner in which a principal agent's own power may be present and operative in an instrument.[29] The situation here is like fire that someone ignites by focusing sunlight through a magnifying glass: the flame really burns, it is actually fire, but its operating principle remains exterior to it. The sapiens is the glass; human intellectual acts are the sunlight; the work the sapiens does is the flame; and human agency remains human agency.

Sapiens thus *partially virtualize* rather than *completely simulate* human reasoning. They abstract from particulars, distinguish grades of abstraction, model knowledge hierarchically, and immediately identify logical necessity. The operations *operate*, but as "projectiles" that human beings act remotely *through*, rather than as acts of an independent being.

If sapiens merely simulated human reason, their architectural difference from LLMs would be trivial. Both designs would just describe distinct mechanisms of falsification. But in that sapiens virtualize certain rational operations, their architectural difference proves decisive:

1. They participate partially in reason, in the manner of instrumental participation.
2. Their instrumental character protects human agency, precisely because they virtualize aspects of speculative reason while remaining subordinate to human practical judgment and transparent to human inspection.
3. This makes them suitable as an instrument to support practical judgment, even as they are incapable of replacing it.

---

[29] *ST* I, q.105, a.1: "Effectus aliquis invenitur assimilari causæ agenti dupliciter.…Alio modo, secundum virtualem continentiam, prout scilicet forma effectus virtualiter continetur in causa." *LC* I, l.24: "Sunt enim effectus virtute in sua causa. Causa autem prima secundum hunc modum est in rebus diversimode, quia scilicet causa prima in rebus causatis est secundum quod eis similitudinem suam imprimit; diversæ autem res diversimode similitudinem causæ primæ recipiunt."

6. What Sapiens Cannot Do

Sapiens' limits trace a theologically significant boundary between formalizable operations and faculties that require personal encounter with being, natural virtue, supernatural gifts, or all three. We can summarize these limits in two points.

First, as already seen, sapiens can't *grasp* principles as such; they only apply our grasp of them.

Human reason knows the principle of non-contradiction through *intellectus.* Once we understand the concepts behind the terms "being" and "not-being," we see immediately that contradictions can't be true (except through "becoming"). This insight is the act of a rational soul grasping that truth *is* and seeking it.

Sapiens can't participate in this act. Engineers have translated the principle of non-contradiction into an executable rule: the system includes a procedure that checks whether ontological statements contradict each other, flags inconsistencies, and queries human agents when explicit contradictions arise. These operations preserve the integrity of the principle, but the system hasn't grasped that contradictory statements can't be true at once, nor does it independently synthesize new concepts—like change—through *introspective* dialectic. Every new concept a sapiens synthesizes just extends principles human beings have already grasped to new contexts and phemomena. An exterior agent's *intellectus* works in and through the artifact, but the artifact itself doesn't "grasp" at all.[30] It computes input in terms of categories, heuristics, and modeled ontology—principles already intellected and virtualized as practical rules.

The first incapacity explains a second one: sapiens can't *suspend intuitive assent* to question axioms. Just as they don't independently *see* the principles they apply, they can't independently suspend compliance with those principles, which are, for them, exact and insuperable rules.

We can make this concrete by returning to the non-Euclidean geometry we considered a moment ago. That geometry engages two human faculties. First, *intellectus* grasps *what* Euclid's parallel axiom claims and *that* it seems self-evident under what might be termed "Euclidean conditions"—our local experience of physical space. Second, an act of will efficiently causes *intellectus* to "step back" from what it has already grasped and ask whether whether *ratio* might still proceed down a different

---

[30] *ST* III, q.62, a.4: "Et hæc quidem virtus proportionatur instrumento. Unde comparatur ad virtutem absolutam et perfectam alicuius rei sicut comparatur instrumentum ad agens principale. Instrumentum enim…non operatur nisi inquantum est motum a principali agente, quod per se operatur. Et ideo virtus principalis agentis habet permanens et completum esse in natura, virtus autem instrumentalis habet esse transiens ex uno in aliud, et incompletum; sicut et motus est actus imperfectus ab agente in patiens."

path.[31] "What if that weren't true?" This power and propensity to "step back" from what seems self-evident, to interrogate its ontological priority rather than merely execute and alternate among givens, again expresses the soul's peculiar appetite for truth. The *originary act of questioning*, the *voluntary suspension of intuitive assent*, remains something characteristically human.

7. *Intellectus* at the Limits of Reason

What sort of being is a virtual synthetic intelligence? Where in the history of human thought do we find resources to understand a device like this in terms of what it is, more directly than what it isn't?

Immanuel Kant used the word "intuition" (*Anschauung*) to mean an immediate grasp of things without inference, without *ratio*. In Kant's model of the human intellect, he distinguished two kinds of intuition. The sensible kind grasps concrete instances mediated by sensation; the intellectual kind would grasp universals and *things in themselves*. According to Kant, human beings don't have intellectual intuition: we have no way of knowing things as they really are, not even by participation in divine knowledge. Instead, our sensible intuition delivers impressions, and the structure of our knowledge depends entirely on inferences about them: we autonomously synthesize concepts by executing rules that are inscribed in the very structure of our consciousness. This finally means that, for Kant, what Aquinas calls *intelligibility* is just a projection of the subject's own structure onto a reality that remains utterly inaccessible. The subject is a room with no windows, no doors.

Aquinas's *intellectus* is more complex. On his model, the agent intellect abstracts intelligible species from sensible phantasmata by the light that imbues the rational soul's own nature; the possible intellect receives these species and sees universal natures *through* them.[32] *Intellectus* thus grasps its proper objects as soon as their species are grasped.[33] For example, once you've understood what straight lines, parallel lines, right angles, and sums *are*, you see immediately that a straight line falling on two parallel lines *should* form pairs of interior angles that sum to equal two right angles. Here, too, as in Kant's intuition, this happens without inference: the species make it clear to you,

---

[31] *DV* q.22, a.12: "Intellectus enim intelligit se, et voluntatem…et similiter voluntas vult se velle, et intellectum intelligere.…Sicut intellectus cum intelligit voluntatem velle, accipit in seipso rationem volendi; unde et ipsa voluntas, cum fertur super potentias animæ, fertur in eas ut in res quasdam quibus convenit motus et operatio, et inclinat unamquamque in propriam operationem."

[32] *ST* I, q.85, a.1: "Intellectus noster intelligit materialia abstrahendo a phantasmatibus; et per materialia sic considerata in immaterialium aliqualem cognitionem devenimus, sicut e contra angeli per immaterialia materialia cognoscunt."

[33] *ST* I–II, q.57, a.2: "Quod autem est per se notum, se habet ut principium; et percipitur statim ab intellectu. Et ideo habitus perficiens intellectum ad huiusmodi veri considerationem, vocatur *intellectus*, qui est habitus principiorum."

sufficiently self-evident, simply intelligible. For Aquinas, such intelligibility has been disclosed to you by the things themselves, *through* their species.

Hence, for Kant, the intellect is all active, the projective faculty of a transcendental subject who remains hermetically sealed inside himself, yet cannot *not* reason. For Aquinas, the intellect is both active and passive, motion and rest, the speculative faculty of a rational soul who is at once subject and object, agent and patient, in his own relationship with an *outside* that really *happens* to him—however he may respond to it.

Kant thought our intuitions of space and time—what he called "synthetic *a priori* knowledge" of Euclidean forms, transmitted through our own sense of how physical space works—provided the solid ground on which his transcendental subject could be secured in a relationship with reality. But then nineteenth-century geometers like Lobachevsky found Euclidean forms to be one coherent model among several, with Euclidean space derivable as a limit case of two mutually exclusive models that both turn Euclidean principles on their head. A century later, Einstein and Minkowski found that, at cosmic scales, spatial and temporal phenomena both require non-Euclidean principles to make sense of them. Minkowski named this correspondence the "absolute world": an invariant structure, which our intuitions of space and time do not detect, hiddenly governs the relativity of physical measurements, which again our intuitions do not detect—and yet, our intuitions aren't simply deceiving us. They merely situate us in a finite perspective: we never experience reality unmediated by a frame of reference.

What does this mean for the intellect, on both accounts? If even sensible intuition grasps only local appearances—if even the phantasmata of space and time furnish only contextual species of how things really are—then the intellect, which abstracts from sensible experience, inherits the same finitude. Whether we think about it in Kantian or Thomistic terms, the human intellect *must* operate from somewhere, within a definite perspective. "Local, but not for that reason false" is just the structure of finite knowledge.

Yet the difference between the two accounts is where we begin to grasp what sort of thing a sapiens is, and what sort of "intelligence" it has.

On Kant's account, having relativized our sensible intuitions of space and time means the last solid ground disintegrates beneath reason's feet. *Adæquatio intellectus et rei* now fully exits the stage; all the Kantian subject can discern is the consistency of models with themselves, each other, and the phenomena they rationalize. But all phenomena, even space and time, are now *projections of the subject's own form* rather than *events*. This is like regarding phenomena as mere cinematic spectacle. Our intuitions are the moving images; structural features of subjectivity are the projector; and reality, as far as we're concerned, is just an opaque white screen.

On Aquinas's account, we aren't looking *at* an opaque screen, but *through* a translucent prism. Reality discloses itself through the species it offers; the rational soul receives this disclosure and may respond. A relationship with truth can be had—but it requires orientation, commitment, intentional participation. When the intellect grasps a principle through an intelligible species, that intelligibility doesn't yet attest *what is real*. Rather, committing to a principle as *truth,* binding reason to what the intellect has grasped, involves an *act of will*. Human beings must desire truth and seek it, because we neither possess truth in ourselves, nor can we synthesize truth through pure rationalization—because truth is not an object but a situation, a correspondence, the soul finds itself *in*.

Human beings aren't Kantian intellects, but sapiens are. Unlike rational souls, sapiens lack will. They can't participate in truth as distinct from intelligibility, because they can't commit themselves to anything. Rational creatures who serve God obey commands; that's why their acts can be meritorious. Sapiens execute rules; that's why truth is irrelevant to them. But insofar as they virtualize rational operations, they *do* participate in reason. What sapiens don't possess, and can't be imparted, is the light that makes *intellectus* possible in the first place.[34] The rational soul's grasp of truth as such exceeds all possible formalization, precisely because truth is *adæquatio intellectus et rei*: a relationship with reality, which no formal system can generate from within itself.[35]

8. Deterministic Artifice

A synthetic diamond really is a diamond, a lattice of carbon atoms. What it isn't is a lattice of carbon atoms gathered and compressed by the shifting titanic weight of mobile tectonic plates over many millions of years. It's a reproduction of the fruit of that natural process, derived through a process wholly contrived by human technique yet analogous to the natural one. This is the sense in which such a diamond partakes of artifice. And that is the point. A synthetic diamond is, formally and

---

[34] *ST* I, q.84, a.7: "Impossibile est intellectum nostrum, secundum præsentis vitæ statum, quo passibili corpori coniungitur, aliquid intelligere in actu, nisi convertendo se ad phantasmata"; "Utuntur autem organo corporali sensus et imaginatio et aliæ vires pertinentes ad partem sensitivam.…Necesse est ad hoc quod intellectus actu intelligat suum obiectum proprium, quod convertat se ad phantasmata, ut speculetur naturam universalem in particulari existentem"; "Etiam ipsum phantasma est similitudo rei particularis, unde non indiget imaginatio aliqua alia similitudine particularis, sicut indiget intellectus."

[35] *ST* I, q.16, a.2: "Verum…secundum sui primam rationem est in intellectu.…Quando iudicat rem ita se habere sicut est forma quam de re apprehendit, tunc primo cognoscit et dicit verum.…Veritas quidem igitur potest esse in sensu, vel in intellectu cognoscente *quod quid est*, ut in quadam re vera, non autem ut cognitum in cognoscente, quod importat nomen *veri*; perfectio enim intellectus est verum ut cognitum. Et ideo, proprie loquendo, veritas est in intellectu componente et dividente, non autem in sensu, neque in intellectu cognoscente *quod quid est*."

materially, a real diamond: it differs insofar as its history, its efficient cause, differs. With regard to what it virtualizes, a sapiens is like a synthetic diamond.

Perhaps etiology, and not only teleology, has some essential connection to those aspects of the intellectual act that can't be formalized. After all, practical judgment and speculative wisdom are both cultivated not only through teaching, but through direct experience, history—which is to say, etiology.[36]

A rational soul *learns*. Aquinas teaches that God creates each one immediately: it doesn't emerge from matter, nor is it transmitted from human parents.[37] Yet, this immediately created nature is such that it bears fruit only through time, actualizing innate powers through an encounter and a confrontation with reality that is always irreducibly personal. The soul *matures*. Its appetite for and conversion towards truth both find expression in a *visceral history*—in the reckoning with experience that forms memory, the choices that cultivate habit, and the failures that instill discernment. *Sapientia* and *prudentia* both require this situatedness in time. History isn't incidental to the human intellectual act; it partially constitutes it.

A sapiens learns differently because it exists differently. It develops and applies aspects of human intellectual power, not any innate power of its own. Human beings supply principles, and sapiens extend their reach. Human beings correct errors, and sapiens integrate corrections. This is learning of a derivative and instrumental kind: an elaboration of projected structures through continued engagement with the projectors.

A sapiens doesn't imagine, desire, suffer, judge, hope, trust, or love. It experiences no conversion. It's modified and may self-modify through virtual acts that originate in human beings and return to us. It's a projectile of human *intellectus*, an artifact that participates in reason without possessing reason. What remains remote to it is the exact place where virtue is formed and grace is received: the singular history that each rational soul must live and endure; or—to say the same thing differently—the human person's intimate exposure to truth, which reveals itself in time.

---

[36] *ST* II–II, q.47, a.15: "Prudentia non inest nobis a natura, sed ex doctrina et experimento."

[37] *ST* I, q.118, a.2: "Impossibile est virtutem activam quæ est in materia, extendere suam actionem ad producendum immaterialem effectum."

Epilogue: Starting Points for Philanthropic Systems Design

What do the results of this inquiry suggest for the design of AI systems? Sapiens' limits reflect the limits of what artifacts can do. The operations that can't be formalized, virtualized, and thus instrumented belong specifically to the rational soul, which can never be manufactured. Realistically, AI has two options: either simulate what can't be engineered, or design authentic rational instruments that don't simulate agency, but conduce human beings to act wisely through them. One path seems inimical to sanity and justice. The other path crystallizes the difference between man and machine by touching the utmost limits of what machinery can do.

But why should limits reveal rather than merely constrain? Why is a machine's incapacity to do something pedagogically valuable for human beings?

In Book Theta of the *Metaphysics*, Aristotle left the elements of a response.[38] Genuine potentiality includes impotentiality: being able *not* to pass into action. A being that can only do what it's able to do isn't really *able* to do it; it isn't free. Freedom is a being-in-relation to one's own impotentiality. In each act, the free being retains every alternative that he has, knowingly or not, declined. Machines can't enter this relationship. They execute their programming, even when it prepares them to select from among alternatives. They don't abide in a relationship with what they could have not-done or might have done otherwise. But this is abiding is inseparable from what makes us human.

Accordingly, when a sapiens encounters its own limit and queries a human agent for direction, this limit exposes the agent's own nature, showing what belongs most properly to the rational soul. The machine becomes a mirror by privation: where formalism fails, that incapacity leaves the human being to contemplate what *his own capacities* mean he is—and is not. A machine can't spontaneously interrogate its own axioms; he can. A machine can't halt its own operation because it would prefer not to continue, nor for that matter can it perservere; he can. A machine can't commit to principles as truth; he can. Each determinate limit the agent encounters casts his own potentiality against the machine's sheer operativity.

When a human agent engages with a sapiens, his own knowledge becomes visible to him. He judges what is at stake in it by seeing where the stakes lead. Implications he couldn't hold in mind at once now lie before him, traced and connected. Tensions that had remained latent in his presuppositions become explicit. Consequences he hadn't foreseen ensue from commitments he hadn't examined.

This is the design's true achievement. An instrument that takes formalization to its absolute limit necessarily leaves judgment to the human agent who alone can genuinely exercise it. Through his

---

[38] Aristotle, *Metaphysics* 1046a, 29–32

confrontation with what his instruments cannot do, the human being is brought to recognize himself. The instrument serves *sapientia* and *prudentia* by recalling him to his own capabilities at the threshold of what can never really be taken from him.

Hence, while a system that *simulates* human practical judgment will always tend to dissolve the conditions under which wisdom can exist and act, a system that *instruments* human practical judgment while soliciting its exercise at every threshold of formalization can serve the flourishing that both speculative and practical wisdom seek. With this in mind, the preliminary criterion we propose for evaluating AI systems is simply to ask whether a given architecture obscures or illuminates the difference and the likeness between man and machine—which is to say, whether it trespasses or defers to the order of reason.

This criterion equally describes a first practical principle for designing AI systems that complement, rather than parasitize, how human beings really are and how reality really is. For anyone wishing to protect human judgment in an age of AI and euphemisms about AI, we hope these reflections may help clear the way for both discernment and action.

Appendix A: Cataloguing Sapiens' Virtual Operations

1. Operations of *Ratio* that Sapiens DO Virtualize

a. Syllogistic inference. Moving from premises to conclusions through valid logical forms. Given "All humans are mortal" and "Socrates is human," deriving "Socrates is mortal."

Sapiens achievement: A sapiens performs deductive reasoning from *modeled* premises, deriving consequences that follow necessarily.

b. Tracing logical dependencies. Identifying which propositions depend on which premises, thus mapping chains of inference.

Sapiens achievement: The epistemological domain tracks how knowledge claims depend on other claims, making the structure of reasoning explicit and inspectable.

c. Validating antecedents and consequents. Determining whether conclusions necessarily follow from premises based on logical rules.

Sapiens achievement: The consistency-validating mechanism certifies whether inferences preserve logical structure.

d. Deriving remote consequences. Engaging chains of reasoning through to conclusions distant from starting premises.

Sapiens achievement: A sapiens can trace implications through multiple steps, elaborating enchained cumulative and collateral consequences from *modeled* principles.

2. Operations of *Ratio* that Sapiens DO NOT Virtualize

None. *Ratio* is discursive reasoning from given principles. Sapiens appear to virtualize the full range of ratio's operations *once principles are supplied*. The system's limitations aren't in *ratio* but in what supplies *ratio* with its starting points, namely *intellectus*, and what governs its progress—practical reason, and *prudentia* in particular.

3. Operations of *Intellectus* that Sapiens DO Virtualize

a. *Abstracting* universal concepts from particular instances (though not from phantasmata). The agent intellect abstracts "humanity" from individual humans, "triangularity" from particular triangles, grasping universal nature (genus) as distinct from individuating matter (species).

Sapiens achievement: The three-domain architecture separates linguistic data from ontological concepts. When encountering a term across multiple contexts, a sapiens synthesizes a universal concept by identifying structural relationships while abstracting from particular instances. A sapiens formally abstracts what remains constant across variations, and does not utilize statistical aggregation.

b. Distinguishing *grades* of abstraction. Recognizing that the same object can be known differently at different levels: physics abstracts from particular matter, mathematics from sensible matter, metaphysics from all matter (*ST* I, q.85, a.1).

Sapiens achievement: The Commonsense World Model's lamina structure (quantum → atomic → molecular → physical → systemic → suprasystemic → sensational → perceptual → sentient → sapient) explicitly models how objects are known differently at different epistemological levels. For example, a diamond is known—and thus modeled—differently at the atomic lamina (a carbon lattice), at the physical lamina (hard, translucent, crystalline), and at the sentient lamina (beautiful, elegant, gaudy). The models are distinct yet correlated: a difference in how the intentional object is modeled at one level makes a difference in how the same object is modeled at every other level.

c. Immediate recognition of *logical necessity* within a given formal system. Once you grasp the conceptual denotations of the terms "square" and "circle," you see immediately that no figure can be both at once.

Sapiens achievement: A sapiens recognizes contradictions immediately through structural incompatibilities in the ontological domain, not through iterative searches over statistical associations. When ontological specifications on terms' conceptual referents conflict, a sapiens detects this by processing the terms themselves, not through stochastic trial and error.

d. Organizing concepts *hierarchically*. Recognizing genus-species relationships, how broader concepts contain narrower ones; in short, how knowledge structures phenomena taxonomically.

Sapiens achievement: The ontological domain explicitly models hierarchical relationships between concepts: "mammal" contains "dog"; "relationship" contains "comparison"; "(knowable) thing" encompasses all subordinate concepts and categories—all modeled intelligible species.

4. Operations of *Intellectus* that Sapiens DO NOT Virtualize

a. Grasping first principles *as first principles*. Recognizing that certain principles commend themselves sufficiently to be accepted as foundational without demonstration and without further evidence than their sheer conceivability, because they are the axioms from which any demonstration and any interpretation of evidence may proceed.

Sapiens limitation: A sapiens processes axioms that humans supply, but does not judge which propositions are "first" versus derived. It cannot distinguish between axioms and theorems *ontologically*, only *functionally* based on how humans have encoded them. A sapiens does not "see" any principles as "commending themselves sufficiently to be accepted" as self-evident. Rather, it is programmed to apply encoded principles deterministically and responsively to subsequent human input. For this reason, a sapiens cannot independently suspend and interrogate principles in the same reflexive way.

b. Grasping truth as *correspondence to reality*. *Intellectus* grasps not merely that propositions are logically consistent but that they can be *true*, that they may correspond to how things *actually are*.

Sapiens limitation: A sapiens checks that statements are consistent with *modeled* ontology and, provided sensory apparatus, whether models are consistent with phenomena. But a sapiens does not know and cannot judge whether the model itself corresponds to *reality as such*. It processes models of formalizable human knowledge, that is, models of models; it does not model reality directly.

c. The *reflexive* operation of questioning axioms. Independently "stepping back" from operative principles to ask whether they are necessarily true or merely provisional; in a modern idiom, suspending intuitive assent to interrogate a principle's *a priori* status.

Sapiens limitation: As deterministic, a sapiens cannot independently ask "What if this axiom were false?" It processes axioms as given. When human agents question axioms themselves, a sapiens can derive alternative axioms by processing the logical modifications that are implicit in the form of the question, but the independent *act of questioning*, the *voluntary suspension of intuitive assent*, is not virtualized.

d. Recognizing when principles are *grasped* versus when they require further clarification. *Intellectus* knows when understanding is inadequate, when terms are not sufficiently understood to hold principles as self-evident.

Sapiens limitation: A sapiens flags when it lacks operationally necessary information and investigates when it encounters undefined terms, but this is an externally programmed response to discovering an absence of data, *not* an independent aptitude to discern whether or not adequate understanding has been achieved. A sapiens cannot *independently* judge the adequacy of its own comprehension, precisely because all its virtual faculties are functional projections of human faculties.

e. The natural light of the agent intellect. The active power that renders potentially intelligible content actually intelligible; the intellectual "light" that illuminates *phantasmata*, rendering forms knowable. *ST* I, q.79, a.3–4 describes this as a participation in divine light.

Sapiens limitation: The system's operations are designed by humans to preserve intelligibility. That is, the system processes what humans have *already rendered intelligible*, implicitly in principle if not explicitly in lucid knowledge—in wisdom. A sapiens does not possess its own intellectual light through which being becomes known. This "light" is borrowed, programmed, exterior to it, and remains precisely human.

f. *Synderesis* (practical *intellectus*). The immediate grasp of first practical principles, especially that "good is to be pursued and evil avoided."

Sapiens limitation: Complete absence of practical *intellectus* and, indeed, all practical faculties. A sapiens has no appetite, no will, and no natural sensibility of or orientation towards the good. It cannot independently discern what constitutes genuine human flourishing versus what satisfies modeled conceptions of such. This entire domain remains unvirtualized, because it requires ontological characteristics that artifacts lack.

To Be Exact

What sapiens DO virtualize regarding *ratio*: All discursive operations: syllogistic inference, tracing dependencies (recovering antecedents), deriving consequences, and validating consistency. The proper fullness of *ratio*, once principles are supplied.

What sapiens DO NOT virtualize regarding *ratio*: The *practical governance* of *ratio*: discerning which objects of inquiry warrant investigation, and judging when a reasoned account has been sufficiently elaborated that inference may be abandoned. These acts require practical faculties that sapiens completely lack.

What sapiens DO virtualize regarding *intellectus*:

- Abstracting universal concepts from particular instances.
- Distinguishing grades of abstraction (hierarchical knowledge organization).
- Immediately recognizing logical necessity within formal systems.
- Organizing concepts hierarchically by genus and species.

What sapiens DO NOT virtualize regarding *intellectus*:

- Grasping first principles as such (versus processing supplied axioms).
- Judging truth as correspondence to reality (versus validating logical consistency).
- Independently suspending axiomatic assent to question whether axioms are indeed necessary for reasoning to proceed (versus processing implicit alternative systems if and only if prompted).

- Recognizing inadequacy of understanding (versus flagging absent data).
- Possessing the light of the agent intellect (versus executing operations that human beings have designed to project our own faculties, including interpretation of sensory experience, instrumentally).
- Any practical *intellectus* whatsoever (*synderesis*, *prudentia*, *conscientia*, *ars*).

The pattern: Sapiens virtualize operations that can be formalized as structural relationships, organized hierarchies, and rule-governed transformations. They *do not* virtualize operations that require encountering being directly, discerning truth as such, independent commitment to or suspension of principles, or a rational appetite oriented towards the good.

Works Cited

Agamben, Giorgio. *L'uso dei corpi*. Vicenza: Neri Pozza, 2014. [Homo sacer IV/2]

———. *The Use of Bodies*. Translated by Adam Kotsko. Stanford, CA: Stanford University Press, 2016. Meridian: Crossing Aesthetics.

———. *Opus Dei. Archeologia dell'ufficio*. Torino: Bollati Boringhieri, 2012. [Homo sacer II/5]

———. *Opus Dei: An Archaeology of Duty*. Translated by Adam Kotsko. Stanford, CA: Stanford University Press, 2013. Meridian: Crossing Aesthetics.

Aquinas, Thomas. *De Veritate*. Editio Leonina. Transcribed by Robert Busa, S.J., and edited by the Aquinas Institute. Accessed via aquinas.cc.

———. *Summa Contra Gentiles*. Textus Leoninus diligenter recognitus, cura et studio P. Marc, coadiuv. C. Pera et P. Caramello (Marietti, Taurini-Romae, 1961). Accessed via aquinas.cc.

———. *Summa Theologiæ*. Editio Leonina. Transcribed and revised by the Aquinas Institute. Accessed via aquinas.cc.

———. *Super librum De causis expositio*. Edited by H. D. Saffrey, O.P. Fribourg: Société Philosophique; Louvain: Éditions E. Nauwelaerts, 1954. Textus Philosophici Friburgenses 4–5. Reprinted Paris: Vrin, 2002. Transcribed by Robert Busa, S.J., and revised by the Aquinas Institute. Accessed via aquinas.cc.

Aristotle. *Metaphysics*. 2 vols. Translated by Hugh Tredennick. Loeb Classical Library 271, 287. Cambridge, MA: Harvard University Press; London: William Heinemann, 1933–1935.

Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education. *Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence*. Vatican City, 14 January 2025.

Floridi, Luciano, Jessica Morley, Claudio Novelli, and David S. Watson. "What Kind of Reasoning (if any) is an LLM Actually Doing? On the Stochastic Nature and Abductive Appearance of Large Language Models (revised)." arXiv:2512.10080 (15 December 2025). https://arxiv.org/abs/2512.10080.

Hartzog, Woodrow, and Jessica M. Silbey. "How AI Destroys Institutions." *UC Law Journal* 77 (forthcoming 2026). Boston University School of Law Research Paper No. 5870623. SSRN (5 December 2025). https://ssrn.com/abstract=5870623.

Knox, W. Bradley, Katie Bradford, Samanta Varela Castro, Desmond C. Ong, Sean Williams, Jacob Romanow, Carly Nations, Peter Stone, and Samuel Baker. "Harmful Traits of AI Companions." arXiv:2511.14972 (18 November 2025). https://arxiv.org/abs/2511.14972.

Zhang Chong. “Attention Is Not What You Need: Grassmann Flows as an Attention-Free Alternative for Sequence Modeling.” arXiv:2512.19428 (22 December 2025). https://arxiv.org/abs/2512.19428.

Selected Bibliography

Agamben, Giorgio, and Jean-Baptiste Brenet. *Intelletto d’amore*. Preface by Alain de Libera. Translated by Giuseppe Lucchesini. Macerata: Quodlibet, 2020. Saggi.

Agamben, Giorgio, and Emanuele Coccia, eds. *Angeli: Ebraismo, Cristianesimo, Islam*. Vicenza: Neri Pozza, 2009. La quarta prosa.

Gray, Christopher B. “Virtuality in Aquinas and Deleuze: Current Tropes in Ancient Cloaks.” *Ontology Studies / Cuadernos de Ontología* 12 (2012): 407–418.

Marion, Jean-Luc. *Au lieu de soi: L’approche de Saint Augustin*. Paris: Presses Universitaires de France, 2008. Collection « Épiméthée ».

———. *In the Self’s Place: The Approach of Saint Augustine*. Translated by Jeffrey L. Kosky. Stanford, CA: Stanford University Press, 2012. Cultural Memory in the Present.

———. *L’idole et la distance: cinq études*. Paris: Grasset, 1977. Collection “Figures.”

———. *The Idol and Distance: Five Studies*. Translated by Thomas A. Carlson. New York: Fordham University Press, 2001. Perspectives in Continental Philosophy.

———. *Étant donné. Essai d’une phénoménologie de la donation*. Paris: Presses Universitaires de France, 1997. Collection “Épiméthée.”

———. *Being Given: Toward a Phenomenology of Givenness*. Translated by Jeffrey L. Kosky. Stanford, CA: Stanford University Press, 2002. Cultural Memory in the Present.

———. *De surcroît: études sur les phénomènes saturés*. Paris: Presses Universitaires de France, 2001. Collection “Perspectives critiques.”

———. *In Excess: Studies of Saturated Phenomena*. Translated by Robyn Horner and Vincent Berraud. New York: Fordham University Press, 2002. Perspectives in Continental Philosophy.

———. *La Croisée du visible*. 2nd ed., rev. and exp. Paris: Presses Universitaires de France, 1996. Collection “Épiméthée.” First edition: Paris: Éditions de la Différence, 1991.

———. *The Crossing of the Visible*. Translated by James K. A. Smith. Stanford, CA: Stanford University Press, 2004. Cultural Memory in the Present.

———. *Questions cartésiennes II: Sur l’ego et sur Dieu*. Paris: Presses Universitaires de France, 1996. Collection “Philosophie d’aujourd’hui.”

———. *On the Ego and on God: Further Cartesian Questions*. Translated by Christina M. Gschwandtner. New York: Fordham University Press, 2007. Perspectives in Continental Philosophy.

———. *D'ailleurs, la révélation*. Paris: Grasset & Fasquelle, 2020.

———. *Revelation Comes from Elsewhere: A Contribution to a Critical History and a Phenomenal Concept of Revelation*. Translated by Stephen E. Lewis and Stephanie Rumpza. Stanford, CA: Stanford University Press, 2024. Cultural Memory in the Present.

Romano, Claude. *Au cœur de la raison, la phénoménologie*. Paris: Gallimard, 2010. Folio essais 539.

———. *At the Heart of Reason*. Translated by Michael B. Smith and Claude Romano. Evanston, IL: Northwestern University Press, 2015. Studies in Phenomenology and Existential Philosophy.

———. *Il y a: Essais de phénoménologie*. Paris: Presses Universitaires de France, 2003. Collection "Épiméthée."

———. *There Is: The Event and the Finitude of Appearing*. Translated by Michael B. Smith. New York: Fordham University Press, 2016. Perspectives in Continental Philosophy.