Journal of Moral Theology, Vol. 11, Special Issue 1 (2022): 157–181

Metaphysics, Meaning, and Morality: A Theological Reflection on AI[1]

Jordan Joseph Wales

ARTIFICIAL INTELLIGENCE is an increasingly pervasive, if hidden, factor in our daily lives. While "general" AI[2] remains, for the present, an aspiration rather than a reality, so-called "narrow" AI techniques now answer questions on our phones, translate between languages, select the advertisements that we see, recommend our next purchase or musical selection, identify potential hot-spots for crime, flag tumors in brain scans, and soon will drive us to work. Theology can and ought to say much about the ethical implications of artificial intelligence and our use of it, but I wish to ask: what may theology say about contemporary AI *in itself*? Some suggest that the answer is, "relatively little." Theologian David Bentley Hart contends:

> The operations of a computer are merely physical events devoid of meaning….[A] computer does not even really compute. *We* compute, using it as a tool….[I]ts operations are not determined by any semantic content but only by binary sequences that mean nothing in themselves. The visible figures that appear on the computer's screen are only the electronic traces of sets of binary correlates, and they serve as symbols

[1] For the development of this paper, I am indebted to too many persons to list, but I must thank the patience of the JMT editors, the insight of the two anonymous reviewers, as well as John Cavadini, Thomas Clemmons, Matthew Gaetano, Brian P. Green, Andrew Kuiper, Dwight Lindley, Anselm Ramelow, David C. Schindler, John Sehorn, Ezra Sullivan, Marga Vega, Marius Dorobantu, and John Seiffert. Each contributed important insights or commented on portions of the paper. The rest of you know who you are. The deficiencies of the final product have only me for their author.

[2] So-called "general" AI, the "ultimate goal of AI research," would be human-level or superhuman not in the sense of being conscious or having any sort of interior life—indeed, that is highly unlikely—but in being "applicable across all problem types." It would "[work] effectively for large and difficult instances while making very few assumptions." Needing "no problem-specific engineering," such a (now-hypothetical) system "can simply be asked to teach a molecular biology class or run a government. It would learn what it needs to learn from all the available resources, ask questions when necessary, and begin formulating and executing plans that work." Its success, then, would be in its behaviorally measured omnicompetence with respect to the goals that we appoint for it. See Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*, Reprint (New York: Penguin Books, 2020), 46.

> only when we represent them as such and assign them intelligible significances.[3]

It would seem, thus, that not only is it impossible for a programmed computer ever to constitute a mind of the sort that humans have; computers themselves are naught but systems of signs that exist *as* signs only at the whim of the beholder. Or, as Hart urges, they "have meanings only so long as they are objects of the representing mind's attention."[4]

Rhetorically, at least, this view is not without its difficulties. The claim that it is *we* who compute seems stretched to breaking by AI systems that convert Swedish into English or identify faces and fingerprints by self-generated formulae that even the systems' designers cannot comprehend. How can something have a merely observer-dependent meaning when it seems reliably tuned to the world in ways unfathomable to us? I argue that Hart's position—while true so far as it goes—is not so threatening to the reality or meaningfulness of computation as it may seem. As with printed text, we assign both functions and semantic content to tools and computers based on culturally shared intentional frames ("intentionality" here refers not to voluntariness but to "aboutness"). These framings determine the design (and our interpretation) of, for instance, a screwdriver's handle, a computer's output images, and the printed characters on a page. As observer-dependent realities, our artifacts are contingent, but they are not arbitrary.[5]

---

[3] David Bentley Hart, "Consciousness (Chit)," in *The Experience of God: Being, Consciousness, Bliss* (New Haven, CT: Yale University Press, 2013), 219.

[4] Hart, "Consciousness (Chit)," 218.

[5] Amie L. Thomasson, "Artifacts and Mind-Independence: Comments on Lynne Rudder Baker's 'The Shrinking Difference between Artifacts and Natural Objects,'" *APA Newsletter on Philosophy and Computers* 8, no. 1 (2008): 25–26. This is the case even for Piccinini's "mechanistic" account of computation, in which computation is defined not by any semantic content but by the manipulation of non-semantic machine states in line with some mechanistically specified rule. Semantic content can be assigned, of course, but it is not necessary to the definition of computation itself; see Gualtiero Piccinini, "Computers," *Pacific Philosophical Quarterly* 89, no. 1 (2008): 32–73, doi.org/10.1111/j.1468-0114.2008.00309.x; Gualtiero Piccinini, *Physical Computation: A Mechanistic Account* (New York: Oxford University Press, 2015). Even Piccinini, however, acknowledges that shared human purposes (in which we participate by our use of input/output devices) are necessary to fix the level of description wherein the mechanism is defined. Paul Schweizer therefore urges that even a mechanistic account is ultimately observer-dependent although not, therefore, arbitrary; see Paul Schweizer, "Computation in Physical Systems: A Normative Mapping Account," in *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence: Themes from IACAP 2016*, ed. Don Berkich and Matteo Vincenzo d'Alfonso (Cham: Springer International, 2019), 27–47, doi.org/10.1007/978-3-030-01800-9_2. Other positions might be taken, but at this time I find Piccinini, as modified by Schweizer, persuasive enough to move forward.

The fact of intentional framing indicates the ground upon which we may consider AI theologically. Observer-attributed meanings are tied up with the device's service to our purposes; an AI system mediates between our goals and the world with which it is engaged. Exploring these observer-dependent and world-attuned dimensions of AI in light of theological loci both moral (the spiritual life) and metaphysical (the doctrine of creation), I hope to facilitate further explorations of topics that, heretofore, have received comparatively little theological attention.

For this project I draw especially on Augustine of Hippo (lived 354–430 CE), who attended to human interpretation of the world within a Christian understanding of reality. His writings are respected by many Western Christian traditions and, on points relevant to my enterprise, are in broad agreement even with those Eastern traditions by which he is less esteemed. I make two claims:

First, with its metaphysics of *rationes seminales*, Augustine's theology of creation makes sense of the failures and successes of different AI methods by explaining the world as God's self-expression, a kaleidoscopic refraction of his Wisdom rather than a collection of discrete objects standing in crisp relations.

Second, these ontological considerations can be united to Augustine's account of interpretive judgment as a moral act bound up with love, in order to reveal the "deep neural network," contemporary AI's most powerful tool, as a kind of "memory" that maps the world to human purposes, without in itself accommodating the transcendent framing of the spiritual life. As such a "memory," the network may draw us to reduce reality to the measurable scope of this-worldly ambitions; or, as a *pointer* to reality, it may perhaps serve one's regathering of the created echoes of divine Wisdom as one journeys into the Trinity.

NATURAL WISDOM, OR AI'S CHALLENGE TO METAPHYSICS: WHAT IS THE WORLD?

"Symbolic" AI and its Ontological and Epistemological Failures

What computer scientists have called "artificial intelligence" has always reflected something of how their times have interpreted both human beings and the world. Somewhat following Thomas Hobbes, the dominant AI of the 1950s through the 1980s[6]—now called

[6] On the history of AI, see Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (New York: Cambridge University Press, 2010). Or, popularly, see Luke Dormehl, *Thinking Machines: The Quest for Artificial Intelligence—and Where It's Taking Us Next* (New York: TarcherPerigee, 2017). The most widely used introductory textbook on AI is Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. (Upper Saddle River: Pearson, 2009). For clarity, I sometimes pass over distinctions that can be drawn between AI as human-like action (e.g., the Turing Test in the 50's), AI as human-like thought (e.g., Newell

"symbolic" or "Good Old-Fashioned" AI (GOFAI)—was philosophically founded on the "computationalist" hypothesis that thinking simply *is* the logical manipulation of symbolically represented information.[7] Accomplishing this task, a properly programmed computer would in fact *be* "thinking;" "a computer running a program that models a human cognitive process is itself engaged in that cognitive process."[8] Under this paradigm, a computer program that diagrammed a sentence and constructed a plausible response could be said to have *understood* that sentence.[9] Symbolic AI's greatest achievement was in "expert systems"—great structures of linked rules that, when queried, would generate a list of possible answers, perhaps posing further

---

and Simon's early work with symbolic representation in the 60's, leading to the field of cognitive modeling), AI as rational deliberation (e.g., logicism and expert systems in the 80's), and AI as rational agency (e.g., intelligent robots); on which see Russell and Norvig, *Artificial Intelligence*, 1–33.

[7] This intuition, a species of the Computational Theory of Mind (or "Computationalism"), was formalized as the "Physical Symbol Systems Hypothesis," seminally described in Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," *Communications of the ACM* 19, no. 3 (March 1976): 113–26, doi.org/10.1145/360018.360022. The authors conclude: "Intelligence resides in physical symbol systems. This is computer science's most basic law of qualitative structure. Symbol systems are collections of patterns and processes, the latter being capable of producing, destroying and modifying the former. The most important properties of patterns is [sic] that they can designate objects, processes, or other patterns, and that, when they designate processes, they can be interpreted. Interpretation means carrying out the designated process. The two most significant classes of symbol systems with which we are acquainted are humans and computers" (Newell and Simon, "Computer Science as Empirical Inquiry," 125). For a recent assessment, see Nils J. Nilsson, "The Physical Symbol System Hypothesis: Status and Prospects," in *50 Years of Artificial Intelligence*, ed. Max Lungarella, Fumiya Iida, Josh Bongard, Rolf Pfeifer, vol. 4850 (Berlin: Springer, 2007), 9–17, doi.org/10.1007/978-3-540-77296-5_2.

[8] Jaegwon Kim, *Philosophy of Mind*, 3rd ed. (Boulder, CO: Routledge, 2010), 160. "Computationalism, or the computational theory of mind, is the view that cognition, human or otherwise, is information processing, and that information processing is computation over symbolic representations according to syntactic rules, rules that are sensitive only to the shapes of these representations. On this view …there is nothing more to a cognitive process than what is captured in a computer program successfully modeling it." Prominent advocates of some form of computationalism include Daniel Dennett and Steven Pinker; see Daniel C. Dennett, *From Bacteria to Bach and Back: The Evolution of Minds*, 1st ed. (New York: Norton, 2017); Steven Pinker, *How the Mind Works* (New York: Norton, 1997); and Steven Pinker, "So How Does the Mind Work?," *Mind & Language* 20, no. 1 (February 2005): 1–24.

[9] See Bertram Raphael, *SIR: A Computer Program for Semantic Information Retrieval*, AI Technical Reports (AITR-220) (Cambridge, MA: MIT, 1964), 42, hdl.handle.net/1721.1/6904. Even more comfortably asserting the identity of the computer's functioning with true understanding is Roger C. Schank and Robert P. Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures* (Hillsdale, NJ: Erlbaum, 1977). This book occasioned John Searle's much-discussed rebuttal, the "Chinese Room" argument, in "Minds, Brains, and Programs," *The Behavioral and Brain Sciences* 3 (1980): 417–57.

questions to the user in order to prune the tree of possible resolutions. The most thrilling application of such systems was the Deep Blue chess computer that in 1997 defeated reigning world champion Gary Kasparov by winning two out of six games and playing to a draw in the other three.[10]

With time, however, symbolic AI came up against practical limits that suggested philosophical problems, particularly in the paradigm's underlying ontological and epistemological assumptions. Ontologically, symbolic AI worked with pre-defined sets of discrete categories standing in definite relations with one another. This diluted rationalism of innate ideas could easily implement Aristotelian syllogisms[11]—e.g., I wish to be dry in the rain; an umbrella will keep me dry in the rain; I will use my umbrella when it rains—but it did not yield a generalized capacity to deal directly with the world and human knowledge of it. Expert systems could break down in subtle situations wherein the interactions of tens of thousands of rules yielded unexpected and incorrect behaviors.[12] The incompletely understood congeries of factors bearing on the interpretation of a phrase or the outcome of an action made symbolic AI difficult to apply beyond constrained situations.[13] Nor could it represent or reason effectively about knowledge less precisely defined or more naturally contoured such as, for instance, one's sense of propriety in a social situation or one's route through a tangled wood.

Today, many problems of explosive scale in symbolic reasoning have been resolved or circumvented.[14] Guided by heuristic rules of

---

[10] On expert systems, see Nilsson, *Quest for Artificial Intelligence*, 229–40, 481–84. On Deep Blue, see Murray Campbell, A. Joseph Hoane, and Feng-hsiung Hsu, "Deep Blue," *Artificial Intelligence* 134, no. 1 (January 1, 2002): 57–83, doi.org/10.1016/S0004-3702(01)00129-1.

[11] See Aristotle, *Analytica Priora* (Selections), in *The Basic Works of Aristotle*, ed. Richard McKeon, Reprint (New York: Modern Library, 2001), I.2, 24b18–20.

[12] Nilsson, *Quest for Artificial Intelligence*, 326.

[13] Such problems embrace both "combinatorial explosion" (the intractable multiplication of factors in a rule-governed and search-based AI such as an expert systems) and the "qualification problem" (the impossibility of listing all preconditions for successful action). Combinatorial explosion was a special focus of the infamous Lighthill report, seen as responsible for a raft of funding cuts throughout Europe in the 1970s; see James Lighthill, "Artificial Intelligence: A General Survey," in *Artificial Intelligence: A Paper Symposium* (Science Research Council, 1973), www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm. These are related, in turn, to the "frame problem" (the impossibility of knowing *which* information is relevant and which can be ignored in the prediction of an action's effects).

[14] John McCarthy pioneered approaches to combinatorial explosion, the qualification problem, and the frame problem with his "Circumscription: A Form of Non-Monotonic Reasoning," *Artificial Intelligence*, Special Issue on Non-Monotonic Logic, 13, no. 1 (April 1980): 27–39, doi.org/10.1016/0004-3702(80)90011-9. Murray Shanahan wrote, recently: "Although improvements and extensions continue to be found, it is fair to say that the dust has settled, and that the frame problem, in its technical

thumb, logic engines like Doug Lenat's "Cyc" selectively traverse vast datasets of discrete categories and relations to analyze business practices and anticipate terrorist activity.[15] Still, the fundamental weaknesses of symbolic methods remain: They falter wherever discrete categories are difficult to detect, unknown, or too subtly intertwined. This is true for problems that humans handle poorly (e.g., weather prediction) and for those they solve well (e.g., behavioral prediction; language interpretation and translation; face recognition). Especially hard are those tasks in which humans attain to refined and effective sensibilities that, nonetheless, are difficult to articulate conceptually (e.g., aesthetics, improvisation, humor, and Go). In the words of Deep Blue architect Murray Campbell, human intelligence "is very pattern recognition-based and intuition-based," unlike symbolic AI's "search intensive" methods, which can require checking "billions of possibilities."[16]

Computer scientist and philosopher Brian Cantwell Smith argues that symbolic AI cannot provide a complete solution because its assumed ontology is inaccurate. The world, he writes, does not come "chopped up into neat, ontologically discrete objects" at human scale,

---

guise, is more-or-less solved" ("The Frame Problem," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2016 [Metaphysics Research Lab, Stanford University, 2016], plato.stanford.edu/archives/spr2016/entries/frame-problem/). The article cites Murray Shanahan, *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia* (Cambridge, MA: MIT Press, 1997); and Vladimir Lifschitz, "The Dramatic True Story of the Frame Default," *Journal of Philosophical Logic* 44, no. 2 (April 2015): 163–76, doi.org/10.1007/s10992-014-9332-8.

[15] Now deployed as Lucid.ai, developed by Cycorp Inc. See popular accounts in Cade Metz, "One Genius' Lonely Crusade to Teach a Computer Common Sense," *Wired*, March 24, 2016, www.wired.com/2016/03/doug-lenat-artificial-intelligence-common-sense-engine/; Doug Lenat, "Not Good as Gold: Today's AI's Are Dangerously Lacking in AU (Artificial Understanding)," *Forbes*, February 18, 2019, www.forbes.com/sites/cognitiveworld/2019/02/18/not-good-as-gold-todays-ais-are-dangerously-lacking-in-au-artificial-understanding/. For scholarly literature, see Douglas B. Lenat, "CYC: A Large-Scale Investment in Knowledge Infrastructure," *Communications of the ACM* 38, no. 11 (November 1, 1995): 33–38, doi.org/10.1145/219717.219745; Abhishek Sharma, Michael J. Witbrock, and Keith M. Goolsbey, "Controlling Search in Very Large Commonsense Knowledge Bases: A Machine Learning Approach," *Advances in Cognitive Systems* 4 (June 2016): 1–12; and Abhishek Sharma and Keith M. Goolsbey, "Simulation-Based Approach to Efficient Commonsense Reasoning in Very Large Knowledge Bases," *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (July 17, 2019): 1360–67, doi.org/10.1609/aaai.v33i01.33011360.

[16] Larry Greenemeier and Murray Campbell, "20 Years after Deep Blue: How AI Has Advanced since Conquering Chess," *Scientific American*, June 2, 2017, www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/.

"standing in unambiguous relations."[17] The world *seems* that way only because its ontological messiness has been made tractable by our human epistemology. Our ability to register the world, to apprehend it richly while coming to know individual objects, passing easily from sensation to conceptual thought, is something prior to the syllogism. Aristotle called this "abstraction."[18] In abstraction, something apprehended through the senses (e.g., this round taut-skinned tart-tasting misshapen sphere), comes to be understood consciously[19] as an instance of some more general category (e.g., plum)—that is, from sensation one comes to understand some *thing*. We do this easily, both recognizing objects and sensing their relations to one another, but it is ill accounted-for by the methods of symbolic AI, which proved clumsy and brittle when it came to distinguishing and identifying objects captured on camera or human words recorded through a microphone—tasks once expected to be easy in comparison to supposedly higher-level activities such as playing chess.

Crucially, according to Smith, our conceptualization of objects in the world is a form of *judgment*—not false but still deeply contingent and partial. We can meaningfully engage in discursive logical reasoning only because the abstractions flowing from our judgments remain grounded by our sense for their situatedness in a world not fungible with any finite set of symbols. More than a rule of thumb, this contextualization is necessary for true reasoning. Otherwise, as Smith says of symbolic AI, our conceptual symbolizations will "float free of

---

[17] Brian Cantwell Smith, *The Promise of Artificial Intelligence: Reckoning and Judgment* (Cambridge, MA: MIT Press, 2019), 28, 8. See also Smith, 34–35. I do not embrace all of Smith's metaphysical positions, but he is an exciting interlocutor.

[18] Aristotle, *De Anima*, in *The Basic Works of Aristotle*, III.4; *Metaphysica*, in *The Basic Works of Aristotle*, I.1; *Physica*, in *The Basic Works of Aristotle*, I.1. See also Allan Bäck, "The Conception of Abstraction," in *Aristotle's Theory of Abstraction*, New Synthese Historical Library (Cham: Springer International, 2014), 7–26, www.springer.com/us/book/9783319047584.

[19] "Consciously," i.e., as a conscious experience. While not agreeing with all of his positions, I will take philosopher John Searle's stab at a popularly accessible definition: "The central feature of consciousness is that for any conscious state there is something that it feels like to be in that state, some qualitative character to the state. For example, the qualitative character of drinking beer is different from that of listening to music or thinking about your income tax. This qualitative character is subjective in that it only exists as experienced by a human or animal subject. It has a subjective or first-person existence (or "ontology"), unlike mountains, molecules, and tectonic plates that have an objective or third-person existence. Furthermore, qualitative subjectivity always comes to us as part of a unified conscious field. At any moment you do not just experience the sound of the music and the taste of the beer, but you have both as part of a single, unified conscious field, a subjective awareness of the total conscious experience. So the feature we are trying to explain is qualitative, unified subjectivity" (John R. Searle, "Can Information Theory Explain Consciousness?," *New York Review of Books*, January 10, 2013, www.nybooks.com/articles/2013/01/10/can-information-theory-explain-consciousness/).

reality," potentially "devolv[ing]…into an endless play of signifiers, signifying nothing."[20] Not only do we need context; the world's richness and our sensitivity to it always exceed our explicitly stated concepts. We cannot, Smith argues, define a finite set of discrete categories—let alone define and detect in the real world the finite set of discrete features by which to identify something as belonging to those categories—that would lead to consistent and reliable performance for purely symbolic AI. There is more to the world, and more to thinking, than symbolic AI assumed.

"NON-SYMBOLIC" OR "STATISTICAL" AI

The problems cited above, along with immense advances in computing power, have brought recent eminence to so-called "non-symbolic" or "statistical" AI, a set of methods among which artificial neural networks hold greatest fame.[21] An artificial neural network is a computer program that mathematically simulates an interconnected set of simplified brain neurons. As an AI technique, then, it begins less from an interpretation of what human thinking *is* than from an analogy with its biological aspects. The goal of such networks is not so much human-like reasoning as it is neuron-like data-processing.[22] Having

[20] Smith, *Promise of Artificial Intelligence*, 73.

[21] The artificial neural network (ANN) was given its original form in Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5 (1943): 115–33. For a time, this technique was neglected after Marvin Minsky and Seymour Papert's critique of single-layer networks' inability to perform certain elementary logical functions (e.g., XOR); see *Perceptrons: An Introduction to Computational Geometry*, 1st ed. (Cambridge, MA: MIT Press, 1969). Fifteen years later, a method for training multi-layer networks was described in David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, "Learning Representations by Back-Propagating Errors," *Nature* 323 (October 1986): 533–36, doi.org/10.1038/323533a0. Nonetheless, Minsky and Papert released an "Expanded Edition" of their book in 1987, refining and restating the limitations of ANNs. *Perceptrons* is often accorded a causal role in the "AI winter" of the 70s through the 90s, a decline of research in light of the perceived limits of both "symbolic" methods and ANNs; see Mikel Olazaran, "A Sociological Study of the Official History of the Perceptrons Controversy," *Social Studies of Science* 26, no. 3 (1996): 611–59, www.jstor.org/stable/285702. The current renaissance of ANN techniques, specifically "Deep Learning" (neural networks with many layers) began in 2012 with AlexNet, a deep convolutional network capable of amazing feats of image recognition; Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM* 60, no. 6 (May 24, 2017): 84–90, doi.org/10.1145/3065386; Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature* 521 (May 28, 2015), www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf.

[22] To say that ANNs are non-symbolic does not mean they are irreconcilable with computationalism. Moreover, perhaps they could even be considered "symbolic" at the appropriate scale. See discussion of these two controverted issues in Michael Rescorla, "The Computational Theory of Mind," in *The Stanford Encyclopedia of Philosophy*, plato.stanford.edu/archives/spr2017/entries/computational-mind/.

some affinity with the British empiricist tradition, these methods are far less beholden to assumptions about either ontology or epistemology than are the techniques of symbolic AI.[23]

An artificial neural network receives a pattern of information as numerical values at its input nodes, which are connected with various strengths to layer upon layer of further nodes. At each node, when the sum of incoming connections exceeds some pre-set threshold, that node will fire and its own signal will be transmitted variously to nodes on a further layer, and so on. If you put in a pattern at the beginning, it is transformed as its elements are recombined and processed until something else comes out on the final layer of the network. A network can be "trained" to produce desired responses—say, to predict travel patterns or to recognize faces—by adjusting the strengths of its connections, thus tuning the contribution made by each node to each recombination and, in due course, to the final result. A piano offers a poor analogy but a useful image. If you have ever shouted into the instrument with its sustaining pedal held down, then you have heard its tuned strings resonate with the different frequencies of your shout. One receives back a sort of echo, not of one's words but of the tones of one's voice. Similarly, as a neural network is tuned (i.e., as its connection strengths are adjusted), it begins to resonate with the entangled relations implicit in our world, including relations not easily discerned or logically represented by human investigators. Moreover, by its training, the network does not just echo; it transforms input data in order to make explicit the relations that are of interest to the trainer.

Neural networks and other statistical methods subserve the AI that underlies self-driving cars, programs that beat world champions in the games of Go and chess,[24] the voice recognition of Siri and Alexa,[25]

---

[23] They are not wholly empiricist, but have certain predetermined architectural features, with the debate centering on whether these are domain-general features (as empiricists would claim to be the case in the human brain) or domain-specific, which would entail some "nativist" or quasi-rationalist innateness in their "interpretive" action; thus Cameron Buckner, "Deep Learning: A Philosophical Introduction," *Philosophy Compass* 14, no. 10 (2019): 11–12, doi.org/10.1111/phc3.12625.

[24] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis, "Mastering the Game of Go without Human Knowledge," *Nature* 550, no. 7676 (October 19, 2017): 354–59, doi.org/10.1038/nature24270; David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis, "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm," *ArXiv:1712.01815 [Cs]*, December 5, 2017, arxiv.org/abs/1712.01815.

[25] Sree Hari Krishnan Parthasarathi and Nikko Strom, "Lessons from Building Acoustic Models with a Million Hours of Speech," *ArXiv*, no. 1904.01624 (Cs, Eess, Stat), April 2, 2019, arxiv.org/abs/1904.01624; Brian Barrett, "Alexa's Had a Big Year,

Google Translate,[26] webmail autocomplete functions,[27] and the "curated" recommendations delivered by Spotify, Netflix, and Amazon.[28] Many problems that bedevil symbolic methods can be solved handily by a neural network because, in a manner of speaking, the network is receptive to, imprinted by the structure of the world as presented to it. We might say that it develops a point of view: not a conscious experience, but something like the classical notion of the mind's conformity to a thing[29]—although here that conformity is always constrained by the task for which the AI is trained. But *to what* is it conformed? To answer that question, we need a richer ontology.

CONCEPT AND CONTEXT

Symbolic AI's treatment of the world has a long pedigree that finds analogues in certain streams of ancient Greek thought, for which to

---

Mostly Thanks to Machine Learning," *Wired*, December 19, 2018, www.wired.com/story/amazon-alexa-2018-machine-learning/.

[26] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean, "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," *ArXiv*, no. 1609.08144v2, September 26, 2016, arxiv.org/abs/1609.08144; Cade Metz, "An Infusion of AI Makes Google Translate More Powerful Than Ever," *Wired*, September 27, 2016, www.wired.com/2016/09/google-claims-ai-breakthrough-machine-translation/; Gideon Lewis-Kraus, "The Great A.I. Awakening," *The New York Times Magazine*, December 14, 2016, www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html; Douglas Hofstadter, "The Shallowness of Google Translate," *The Atlantic*, January 30, 2018, www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/.

[27] Yonghui Wu, "Smart Compose: Using Neural Networks to Help Write Emails," *Google AI Blog* (blog), May 16, 2018, ai.googleblog.com/2018/05/smart-compose-using-neural-networks-to.html.

[28] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, and Hemal Shah, "Wide & Deep Learning for Recommender Systems," in *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems - DLRS 2016* (Boston: ACM, 2016), 7–10, doi.org/10.1145/2988450.2988454; Faisal Siddiqi, "Machine Learning Platform Meetup: Recap of the Oct 2017 ML Platform Meetup at Netflix HQ," *Netflix TechBlog* (blog), October 18, 2017, medium.com/netflix-techblog/machine-learning-platform-meetup-ddec090f3c17.

[29] E.g., Thomas Aquinas, ST I, q. 16, a. 1, co.: "Knowledge is according as the thing known is in the knower" and the "truth [of one's own thoughts] is the equation of thought and thing." See also ST I, q. 16, a. 3; translation from *Truth: A Translation of Quaestiones Disputatae De Veritate*, trans. Robert W. Mulligan (Chicago: Regnery, 1952), 1.1. One's apprehension of the world is not just a symbolic representation of an account of it but is a world-conformed habit of mind from which such accounts and their representations are generated. One's capacity for understanding is shaped by one's experience and one's memory and accompanies one in every experience.

understand a thing—be it a natural object or a human-made artifact—was to apprehend rationally its "form" or "idea," that is, its configuration toward some activity or use. According to Langdon Gilkey, for such thinkers,

> Once the scientist has uncovered this form of the object….he really knows all he can or need know about it; he has penetrated to the very heart of reality. He does not need to experience or describe any further its external characteristics or patterns of behavior, since, with its form in his mind, he can predict all that is important about its activities and powers.[30]

Indeed, all "sensible characteristics" (e.g., a knife's gleam) beyond those necessitated by the form (e.g., its cutting edge) are but byproducts of the "necessary but distorting…substratum [that has been] arranged according to the guiding principle" of the form.[31] They may be discarded from consideration as meaningless "result[s] of unpredictable flaws in the material and so quite beyond rational explanation."[32]

This (perhaps unnuanced) rendering of ancient Greek science strikingly anticipates the formalistic world-model assumed by symbolic AI, which Smith finds inadequate both to physical realities and to how we apprehend them: "Taking the world to consist of discrete intelligible mesoscale objects is an *achievement* of intelligence, not a premise on top of which intelligence runs."[33] The concepts by which the world is discretized (i.e., Gilkey's forms) *do* represent reality but they are engagements *with* it rather than separable simulations of it. They are instruments by which we interact with the world at a particular but non-exhaustive level of description. Only as points of contact with their real-world context do they remain true to it. Therefore, especially in "long chains of articulated reasoning" about realities unavailable to immediate experience, our highly abstracted formal concepts must remain habitually "embedded" in their underlying "sub-conceptual webs" so that, from this context, we may draw the "subtleties, adjustments, and so on" that will give "nuance and inflection" to inferences both immediate and distant.[34] By this embedding, articulated reasoning can be a true engagement with rather than a reduction of the world.

Smith derives his understanding of the "sub-conceptual" from the success of today's "deep" neural networks, which have broad input layers and dozens of interior layers. "When fed with data obtained

---

[30] Langdon Gilkey, *Maker of Heaven and Earth: A Study of the Christian Doctrine of Creation* (New York: Doubleday, 1965), 124.
[31] Gilkey, 123–24.
[32] Gilkey, 126–27.
[33] Smith, *Promise of Artificial Intelligence*, 35.
[34] Smith, *Promise of Artificial Intelligence*, 74–75. On these sub-conceptual webs, see also Smith, *Promise of Artificial Intelligence*, 34–35.

directly from low-level sensors" such as cameras, the "high-dimensionality" of the network's layers enables it "to 'encode' all kinds of subtlety and nuance …[without] hav[ing] to categorize and discretize their inputs at the outset." Its self-adjusting weightings, which "store and work with" the input, come subtly to reflect relations between phenomena in the data. Like the cultivation of a sommelier's palate, a progressive attunement to the world's "'sub-conceptual' terrain" renders the network effective in a way that pre-categorization by formal ontologies would never have permitted.[35]

The nature of this attunement, however, is difficult to explore because, as physicist Judea Pearl writes, neural networks are "opaque." Even when tuned to their training data and able to generalize to new data, their interior sensitivities are not at all easily interpreted.[36] Eminent computer scientist Peter Norvig argues that their statistical attunement "describes what *does* happen" but—"mak[ing] no claim to correspond to the generative process used by nature"—it "doesn't answer the question of *why*."[37] Norvig's statement is of ambiguous value. True, networks do not develop theoretical models, but if Smith is right, then the network may say quite a bit, even if obscurely. How could nature's "sub-conceptual" *not* be somehow related to the deep flow of its "generative process[es]," especially if this sub-conceptual gives *our* theoretical concepts their success as engagements with nature itself?

To develop a joint account of nature and networks, let us reflect on what the "sub-conceptual" might be. Consider a hypothetical (but technologically realistic) neural network, trained to distinguish grasses, wildflowers, and trees with fidelity to distinct scientific categories.[38] In "earlier" layers, we might observe activity quite out of keeping with these hierarchical classifications as, for instance, if certain areas were to be equally highly activated by the subtle ridging on a blade of grass, the stalk of a valerian wildflower, and the needles of certain conifers. Were this but a matter of surface-level similarity with no more conceptual heft than the redness of a coral snake grouped with that of a red panda, then we might agree that the network's activity "bear[s] no relation" to nature's "generative process[es]." But what if

---

[35] Smith, *Promise of Artificial Intelligence*, 58–59. It has been proposed that deep networks find inherent symmetries in the data manifolds, to yield useful relations and to encode a large amount of this data. See Buckner, "Deep Learning," 9–11.

[36] Judea Pearl, "The Limitations of Opaque Learning Machines," in *Possible Minds: Twenty-Five Ways of Looking at AI*, ed. John Brockman, 1st ed. (New York: Penguin, 2019), 18.

[37] Peter Norvig, "On Chomsky and the Two Cultures of Statistical Learning," 2011, norvig.com/chomsky.html. Emphasis original.

[38] As critics rightly point out, at no point does the network learn to see these plants as wholes; see Gary Smith, *The AI Delusion* (New York: Oxford University Press, 2018), 50–51. Still, my point concerns that to which it is attuned within wholes.

the network has captured features that give us access to Smith's "subconceptual web"? The visual similarities of grass, stalk, and needle are not mere surface coincidence. A ridged configuration strengthens these plants' narrow structures, which are mostly hollow but must remain stiff to perform their function. In grouping these three, then, the neural network is attuned to what *we* know to be a manifested harmonization of gravity, force of wind, capillary action, and—in the movement of fluid—a hint of metabolism. These three plants are not at all closely related nor all common to the same environment; yet, as expressions of this kingdom of life, they have settled into a groove that expresses something not only about these particular organisms or even about their local environment, but also about the natural harmony of earth as a whole.

All this is taken into the absorbent mind of the attentive child; it is forgotten amidst the classifications by seed, climate, and species in an introductory biology class; and it is rediscovered by the botanist and the gardener. It is like the sounding of a piano note, which bears witness not only to the struck key but also, in its overtones, to the shape of the piano, the species of wood from which it is constructed, and even—perhaps discernible only by a neural network—the orientation of the grain and the history that imparted to that particular tree its distinctive physical quirks.

For a reductively formalistic science, the three plants' ridges mislead because we would define their forms better by macroscale physical characteristics, climates, and modes of nutrition and reproduction. For the same science, *all* that I have written of the piano falls among the "sensible characteristics" that may be discarded upon grasping the intelligible form of the keyed instrument. Against such a view, I contend that *this* sort of thing is what we have meant by "piano" all along; and it is why the classically trained pianist finds something lacking in the finest electronic instrument, as a matter not of snobbery nor of tradition only, but of the full meaning of the piano's "form." Like the commonalities of the three plants, the distinctiveness of the instrument *cannot* be captured by abstracting from its sensible characteristics because its *truest* concept—the concept that we hold—is adequately transmitted only by the experience of the piano itself as a transduction of the world from which it is drawn. It is not that our formal concepts fall short of experience; it is that, as we see in the sensitivity of the neural network, they embrace much more of reality than our way of speaking may lead us to believe.

The Rationes Seminales, an Augustinian Ontology Commensurate with AI

If our attempts to "purify" the conceptual from its sensuous matrix lead to a parody rather than to a more precise grasp of reality, and if concepts engage a thing's form, then I propose to think of particular

forms as including, rather than abstracting from, the harmonics of nature. Augustine of Hippo has in mind such an inherently dynamic form when he speaks of a thing's *ratio* (in Greek, *logos*). His ontology of *rationes* (*Gen. ad litt*. 4–6)[39] makes sense both of symbolic AI's failure and of neural networks' successes by describing how particular things are inextricably at home in the world. Meanwhile, as I will discuss later, his teaching on interpretive judgment (*Trin*. 9.6.11–9.11.16, 15.10.17–15.11.21) clarifies how concepts are "achievements"[40] that are truest engagements when they do not detach things from that matrix.

The key for Augustine is contingency. Plato and the tradition emanating from him sought a fixed non-contingent world—i.e., the forms or ideas—in light of which the contingent and the shifting might be explained. For Augustine, however, there is no world of the forms. The only non-contingent reality is God himself, a simple being, alive in love. Construed as the archetype of all things, his transcendent and inexhaustible life is the divine Wisdom (Prov 8), in which are the non-contingent prototypes or "eternal reasons" (*aeternas rationes*) of all contingent things—not as distinct forms but as identical with his simple life (*Trin.* 12.2.2). "God would not make creatures unless he knew them before he made them; nor would he know them unless he saw them; nor would he see them unless he possessed them; nor would he possess what had not yet been made except as uncreated being, as he is himself" (*Gen. ad litt.* 5.16.34).[41] God's uncreated life is single and simple. Therefore, the *aeternas rationes* are distinguishable only from our point of view, being aspects of God seen as "simply multiple and uniformly multiform" through the prisms of his contingent *created* expressions here below (*Ciu.* 12.19).[42]

---

[39] See, among the secondary literature, Gerald P. Boersma, "The *Rationes Seminales* in Augustine's Theology of Creation," *Nova et Vetera* 18, no. 2 (2020): 413–41, doi.org/10.1353/nov.2020.0030; Christina Hoenig, "Augustine," in *Plato's Timaeus and the Latin Tradition* (Cambridge: Cambridge University Press, 2018), 242–51; Luigi Gioia, *The Theological Epistemology of Augustine's* De Trinitate (New York: Oxford University Press, 2016), 262–69. Also, setting the *rationes* in wider context for today's inquiries, see John C. Cavadini, "Augustine and Science," in *T&T Clark Handbook of Christian Theology and the Modern Sciences*, ed. John P. Slattery (London: T. & T. Clark, 2020), 59–66.

[40] Already quoted from Smith, *Promise of Artificial Intelligence*, 35. See this paper, note 33.

[41] Augustine of Hippo, *The Literal Meaning of Genesis* (401–415), trans. John Hammond Taylor, vol. 1, ACW 41 (New York: Paulist, 1982), 167.

[42] Augustine of Hippo, *Concerning the City of God Against the Pagans* (413-427), trans. Henry Scowcroft Bettenson, Penguin Classics (London: Penguin Books, 2003). See also *Gen. ad litt.* 5.13.29–5.15.33, especially 5.15.33, translated in Augustine, *Literal Meaning of Genesis*, 1:166: "What has been made through Him is understood to be 'life' in Him, the life in which He sees all things when He makes them. He has made them as He has seen them, not looking beyond Himself, but He has numbered within Himself all that He has made. His vision and that of the Father are not different:

In the contingent world, divine Wisdom is expressed doubly: in the kinds of things created according to the *rationes*; and in God's providential governance of the whole, whereby the ways of these things are expressed in interaction with one another (*Gen. ad litt.* 5.12.28).[43] As for kinds, the nature and capacities displayed in the life of each created thing reflect, facet-like, the goodness and wisdom of God:

> Through Wisdom, all things were made; and the motion we now see in creatures, measured by the lapse of time as each one fulfills its proper function, comes to creatures from causal reasons [*rationes*] implanted in them, which God scattered as seeds at the moment of creation when *He spoke and they were made; he commanded and they were created* [Ps. 32:9] (*Gen. ad litt.* 4.33.51).[44]

While distinguishable, created kinds are not isolatable. Somewhat as all created *rationes* are found archetypically in the one divine Wisdom, each plant and animal has its common origin in the causality of the one earth that "received the power of bringing them forth" (*Gen. ad litt.* 5.4.11, see Gen 1:12).[45] From the very beginning, the universe has contained, *in nuce*, the meaningfulness that historically has unfolded into the distinction of contingent creatures. Thus, God made "all things together" (Sirach 18:1; *Gen. ad litt.* 5.23.44–46).[46]

The second contingency—which Augustine honors as no purely platonic thinker could—is history itself.[47] This is the sphere of God's

---

there is one vision, as there is one substance"; citing Job 28:12–13, 22–25. See also, plainly showing that these are not "moments" in God's life, but the eternal life that is God's existence, *Trin.* 4.3. Augustine affirms that nothing is "irregular or unforeseen" by God, because the "*rationes* for all things created and about to be created are contained in the mind of God," "eternal and…unchangeable;" *Ciu.* 12.19. See also John C. Cavadini, "God's Eternal Knowledge According to St. Augustine," in *Cambridge Companion to Augustine*, ed. David Vincent Meconi and Eleonore Stump, 2nd ed. (New York: Cambridge University Press, 2014), 37–59.

[43] Augustine distinguishes the unchangeable *rationes* from God's work from which he rested (i.e., creatures, with their immanent *rationes*) and the things he produces from these works—that is, material things and their motions under providence according to their particular *rationes*.

[44] Cited by Cavadini, "Augustine and Science," 64.

[45] Augustine, *Literal Meaning of Genesis*, 1:153. See also *Gen. ad litt.* 5.4.11; 6.6.10–11; 6.10.17; 6.14.25; Ernan McMullin, "Evolution as a Christian Theme" (Herbert Reynolds Lecture Series, Baylor University, 2004), 7–8.

[46] Augustine does not seem to think that the distinct *rationes* are contingent *within* our historical frame. In other words, while his theory is ripe for development into a theology of biological evolution, he himself does not fully anticipate it.

[47] The Christian belief in God's progressive self-revelation culminating in the Incarnation would have sensitized Augustine to history. We find this even in his early and supposedly neoplatonic treatise *De uera religione*; on which see recently Thomas Clemmons, "The Common, History, and the Whole: Guiding Themes in *De Vera Religione*," *Augustinianum* 58, no. 1 (June 28, 2018): 125–54, doi.org/10.5840/agstm20185816.

providential governance, not a marionette-like foisting of the divine will upon otherwise-free creatures, but an elicitation of their interacting harmonies by the unfolding of temporal events (*Gen. ad litt.* 5.11.17; 5.20.41; *Trin.* 3.5–6.11).[48] Not only does the general developmental and behavioral history of squirrels manifest more fully their *ratio* as a refraction of divine Wisdom; Wisdom is further manifested through particular things' contingent histories of interaction—e.g., *this* squirrel in *this* forest, scrambling up *this* oak tree away from *that* fox. Even turbulent micro-particle systems, the "deep pools [that] seethe with tumbling waterfalls," speak to harmonies moved rather than transgressed by the power of God. The whole of it thrums with the one uncreated *ratio* of divine Wisdom himself (*Gen. ad litt.* 5.20.41)[49] because the *aeternas rationes*, in their simple unity within Wisdom, have an intrinsic order (*ordo*) that is "hidden from us rather than…lacking to universal nature" (5.21.42).[50] We cannot skip past history to access this order because "our knowledge…depends upon the governance in time of creatures already made, inasmuch as God, in the unfolding of his creatures…is working still" (5.4.10).[51]

Thus, the world is not a collection of static essences defined by distinct forms existing on a different plane. Because there is no distinct world of forms, but only this world or the very mind of God himself, we know *rationes* as distinct only through historical and material existence (*Gen. ad litt.* 5.16.34).[52] A creature's *ratio* is not a functional form obscured by material conditions, but a trajectory that works itself out *through* material conditions and in relation to other *rationes*. When we take up one *ratio*, we take up a knot in the whole tapestry. In this kaleidoscopic theophany of things, their ways, and their histories, "the

---

[48] See also Cavadini, "Augustine and Science," 64. Also, *Gen. ad litt.* 5.21.42; *Literal Meaning of Genesis*, 1:172: "Creatures shaped and born in time should teach us how we ought to regard them. For it is not without reason that Scripture says of Wisdom, that *she graciously appears* to her lovers *in their paths and meets them with unfailing providence* (Wis. 6:17)."

[49] Augustine, *Literal Meaning of Genesis*, 1:171–72: "God moves his whole creation by a hidden power, and all creatures are subject to this movement: the angels carry out his commands, the stars move in their courses, the winds blow now this way, now that, deep pools seethe with tumbling waterfalls and mists forming above them, meadows come to life as their seeds put forth the grass, animals are born and live their lives according to their proper instincts, the evil are permitted to try the just. It is thus that God unfolds the generations that he laid up in creation when first he founded it; and they would not be sent forth to run their course if he who made creatures ceased to exercise his provident rule over them."

[50] Augustine, *Literal Meaning of Genesis*, 1:172.

[51] Augustine, *Literal Meaning of Genesis*, 1:153.

[52] Augustine, *Literal Meaning of Genesis*, 1:167: "*In him we live and move and have our being* [Acts 17:28]; but most creatures…being corporeal, are of a different nature, and our mind is unable to see them in God, [that is,] in the archetypes according to which they were made. [Otherwise,] we should know their number, size, and nature, even without seeing them by means of the senses of our body."

world itself in all its ordered change and movement and in all the beauty it presents to our sight," "bears a kind of silent testimony to the fact of its creation, and proclaims that its maker could have been none other than God, the ineffably and invisibly great, the ineffably and invisibly beautiful" (*Ciu.* 11.4.2).[53]

The contingency of a particular *ratio*—that God did not have to make the creature that way, or express himself under that particular economy—is recapitulated in the contingent conditions under which we encounter it materially. The *rationes* imply and are disclosed in the messiness of nature. The *ratio* of a worm is not captured by some definition such as "fleshy flexible moving linear metabolizer," if that definition does not live in the worm's "silent testimony" by its writhing in *this* physical environment, in *this* soil. The "sub-conceptual" detail of the clumping earth after a light rain and the way it shapes the worm's progress—a behavior adapted *to* that sort of soil—is not inconsequential but is entailed in the *ratio* of that creature.

Here, then, is a fitting metaphysical account of the world's diversity and of the inexpressible interior relations held in our concepts—concepts properly used when creatures and their *rationes* are known in their coherence with one another. Herein is Wisdom apparent; herein, the measure and harmony of the whole draws forth our awe and wonder and praise—all of which, John Cavadini writes, would be denied if that whole were formalistically "reduc[ed] to our rationality" in the sense critiqued by Gilkey and Smith.[54] For an Augustinian metaphysics of creation, the neural network at its best is attuned not to happy accidents but to the *rationes*; the network's mathematics do not refute so much as deepen our notion of the concept as an engagement with those *rationes*.[55]

The metaphysical and epistemological challenge of the neural network has led us not to abandon concepts, but to see them as rooted in the historical outworking of reality through the activities of particular things, with histories and their kinds understood as refractions of a simple and unitary divine Wisdom. What, then, has the network captured when it is trained to distinguish images of plants according to a scientific taxonomy? Philosopher Cameron Buckner writes: "The exact boundaries of each category's manifolds," that is, each category's geography in the data-space,

---

[53] Augustine of Hippo, *City of God*, 432.

[54] Cavadini, "Augustine and Science," 64–65, discussing Augustine, *Gen. ad litt.* 5.22.43.

[55] I do not here have space to deal with networks that discover false correlations or biased shortcuts in data, but will say that such problems do not defeat the claim that, when attuned to data causally linked to nature, the network is attuned to the *rationes* in some measure.

> are inaccessible to networks during training; the "goal" of training a neural network for classification can then be understood as discovering a global output function—composed of individual nodes' activation functions and associated link weights—that can draw boundaries between the manifolds of categories that need to be discriminated.[56]

Certainly, the categories may be determined by the human designer, but a network that maintains robust performance when tested has, in a certain way, accomplished a mapping from the *rationes* (manifested in the world-data presented to the network) to the interests of the network's trainer. It has to *work*; and so it must preserve the *rationes* as much as our own categories do in experience even if not in explicit definition—but this means that it can also distort these *rationes*.

KNOWLEDGE, WISDOM, AND ARTIFACTUAL MEMORY

Having inquired into the neural network's relation to nature's *rationes*, we may now consider its relation to human understanding. The network's outputs are human-designated classifications, categories, and purposes to which the network is trained to map its input data. This is how semantics are attributed to the AI program. The strengths and potential deficiencies of this mapping can make it the subject of a deeper moral-theological reflection:

First, for Augustine, concepts in the human mind are begotten by intentional and moral judgments that, in turn, form the very fabric of our understanding as an engagement with the sub-conceptual web—much like the neural network.

Second, while knowledge (*scientia*) directs these understandings toward our own purposes, true wisdom (*sapientia*) receives the *rationes* contemplatively, allowing them to exceed the scope of any purpose. Always developed for a particular task, AI points beyond itself but, in pointing toward *us* before it points toward the world, it seems unable to transcend a utilitarian frame.

Third, I conclude that, as AI cannot escape the morally infused nature of all human thought, we must develop a "spirituality" of AI wherein we do not permit it to stand between us and the world—lest we remain self-imprisoned in the knowledge of our own designs.

MEANING AS MORAL: VERBUM AND MEMORIA

Every act of understanding involves an act of the will. For Augustine, our acts interpreting natural things, conventional signs, and artifacts all follow the same fundamental sequence: we apprehend something through the senses; we judge it as good (i.e., as real)[57] with

---

[56] Buckner, "Deep Learning," 10. Dashes added for clarity.

[57] Moral evils like murder are "good" only in, say, involving voluntary motion. The act itself forestalls any goodness beyond the bare fact of this motion, in intentionally

respect to something else; then, as we cling to that goodness with our approbation or love, we conceive a "mental word" (*verbum mentis*)—i.e., a conceptual understanding (*Trin.* 9.6.11–9.11.16, 15.10.17–15.11.21).[58] The *verbum mentis* is not a spoken word, nor an autonomous form in Gilkey's sense, but a particular embrace by the mind of some facet of reality according to its *ratio*—an embrace, however, that is shaped by the knower's own assessment of its goodness.[59] For Augustine, in the words of Luigi Gioia, "Intellectual knowledge is not the result of an 'infusion' in our mind of a pre-existing reality, but the production of a new reality."[60]

This *verbum* is truer as it approaches an embrace of the *ratio* as that *ratio* is; and this means that one's own desire and love must conform to reality rather than plucking out from reality only that which is congenial to the stance that one has brought with oneself. Even our recognitions of a narwhal or a "no parking" sign are not neutral because our judgments of meaning issue within the general frame of our cultural, societal, and personal values and position within the world. Every act of understanding entails a moral judgment; habitual moral judgments of this sort form our habit of seeing the world.

Augustine calls the ground of this habitual vision our "memory" (*memoria*).[61] While corporeal things cannot be kept uninterruptedly before the physical eyes, *memoria* makes present the object of the mind's striving, such that God and corporeal objects alike can be present uninterruptedly. The *memoria* is not, however, a movie-screen or data repository (*Conf.* 10.17.26). Rather, it is an implicit knowledge of objects and experiences, a fabric of varyingly accurate *rationes* built up from apprehensions in *verba mentis*. Contained implicitly in this fabric, objects can be said to be present to the mind even without conscious cognition. Desire or love—the will's implicit judgment concerning the thing known—draws the object anew into explicit thought as a *verbum mentis* in the intellect. As Augustine writes in the

---

extinguishing the goodness of one personal life by the agent's ugly inter-personal attempt at absolute domination.

[58] Luigi Gioia writes: "The process of knowledge is set off by desire for the object to be known and is completed only through union with the object known through love" (*The Theological Epistemology of Augustine's* De Trinitate, 200).

[59] John C. Cavadini, "The Quest for Truth in Augustine's *De Trinitate*," *Theological Studies* 58, no. 3 (September 1, 1997): 429–40, doi.org/10.1177/004056399705800302.

[60] Gioia, *Theological Epistemology of Augustine's* De Trinitate, 200.

[61] The texts of Augustine dealing most prominently with *memoria* include: *Conf.* 10; *Trin.* 9, 14, 15.19–20. On the *verbum mentis* see *Trin.* 9.11–12; 15.11.20. See also Nello Cipriani, "Memory," trans. Matthew O'Connell, in *Augustine Through the Ages: An Encyclopedia*, ed. Allan Fitzgerald and John C. Cavadini (Grand Rapids, MI: Eerdmans, 1999); Matthew L. Lamb, "St. Augustine on *Memoria* and *Commemoratio*," in *Philosophy and Theology in the Long Middle Ages*, ed. Kent Emery (Boston, MA: Brill, 2011), 237–47.

Confessions, "I hid in my memory not the images but the realities"—that is, the *rationes* as construed in the *verba mentis* (*Conf.* 10.10.17).

More than permitting implicit presence, the *memoria* is a kind of ground *for* thought, formed *by* thought. One's past apprehensions become the fibers from which one's present intuitive leaps are made and concepts past and present (re)woven (*Trin.* 12.14.23). To use a mathematical analogy, the *memoria* is a set of basis vectors, more or less approaching the true principal components of the vector space that is reality. As built up from the *verba mentis* shaped by the will, the *memoria* constitutes the deep substructure of understanding wherein the *verba* of past and future subsist. Like the palate, the *memoria* is cultivated by the things one tastes attentively and potentiates what and how one is able to taste: past judgments shape the *memoria* and the *memoria* is also the substrate wherein the resulting *verba* are sustained. It is our sensitivity to reality, the primary colors of thought, our way of seeing the world, a habit of mind shaping the judgments that will come readily to us, and a sort of sedimentary aggregate of the *verba* begotten over one's lifetime. If a particular *ratio* is a knot in the tapestry of reality, then *memoria* is a corresponding tapestry of mind from which the *verbum mentis* comes forth. Finally, inasmuch as the will and the affections are susceptible of reformation, the *memoria* is malleable as well.

Artifactual Memoria: Knowledge (Scientia) but not Wisdom (Sapientia)

As an artifact, the programmed computer receives its semantics from the meaning-making intentional frame constituted by the judgments of those who share that frame. Now, if the trained neural network maps the *rationes* of some dataset to categories of interest to the system designer; and if this mapping preserves those *rationes* to the extent that they can be transduced without loss into the designer's moral and conceptual engagement with reality; then the neural network is an artifactual *memoria*, its learned weights preserving the mapping of *rationes*. As with a network sensitive to ridged stalks, this *memoria* is not transparently interpretable in terms of scientific classifications or formal conceptual relationships, but nonetheless it encodes the *rationes* of its input data as shaped by the wills of its designers and users, mapping reality to human interest and utility. Thus, its meaning as used in the world involves both the *verba mentis* that shape the system's architecture and especially its trained outputs, and the moral stances implicit in the goals and purposes to which its users put the AI.[62] When it is read as a standard, taken as a prompt for action, or

[62] This remains the case even for apparently purely "scientific" uses. Weather prediction has goals and valences embedded in it—what we deem important, what is the difference between light and heavy rain, what effects are worth singling out for

contemplated for what it reveals, it thus has ultimate reference not to God (as with the *rationes* of the natural world) but to ourselves.

The artifactual *memoria* therefore subserves what Augustine calls "knowledge" (*scientia*) as distinct from "wisdom" (*sapientia*). *Scientia* apprehends things according to their *rationes* for the sake of "action" in the "good use of temporal things." The moral stances and judgments that beget a right *scientia* build up the "virtues that make for right living" along the way to eternal life. The neural network cannot mirror, however, the higher form of knowledge that is "wisdom" (*sapientia*). *Sapientia* engages the *rationes* not according to their usefulness but as they echo the *aeternas rationes* that are one in divine Wisdom; thus, it reaches toward the "contemplation of eternal things" in God himself (*Trin.* 12.14.21–22).

Whereas the *verba* of *scientia* are begotten by a morally oriented will, *sapientia* enters a transcendent frame because it is begotten by the higher love of "charity" (*caritas*), "poured forth in the heart by the Holy Spirit who is given to us" (Rom 5:5). By charity, one participates in God's own life,[63] and so by a long apprenticeship, the Christian's loves may be brought into this frame so that *scientia* will flow seamlessly into *sapientia* as one refers the goodness of all things to the originating goodness of God, loving them *in* God, with him rather than our temporal purposes being the horizon of their meaning (*Doctr. Chr.* 1.3.3–1.4.4). To be truly wise, in Augustine's sense, is to live from within the life of God according to the self-donative love that is the life of the Trinity, and to know according to Wisdom by finding in each created thing a glimmer of the *aeternas rationes*, which are one in God's eternal Wisdom—i.e., Christ, the second person of the Trinity. *Sapientia*, then, is not simply a matter of having a connected view of things, nor only of knowing the causes of things; it is a configuration of the mind according to God, actualized in a relationship *with* God. In this, one lives fully as God's image by remembering, understanding, and loving the Trinity in direct relationship (*Trin.* 14). By this active participation in God's own life, the "mapping" of human *memoria* becomes a living sign and image not first of the world nor of one's worldly purposes, but of Christ, who is God's own self-knowledge. The wisdom begotten in this *memoria* is a vision beyond words, a contemplation beyond representation, received in

identification; all of these have to do with the human scale of life in the world and the interest that we have in it. We must delineate the concepts else how can it enter our web of meaning? Language translation is a particularly knotty case that I hope to address in a future paper.

[63] David Vincent Meconi, "Augustine's Doctrine of Deification," in *Cambridge Companion to Augustine*, 208–28, universitypublishingonline.org/ref/id/companions/CCO9781139178044A023; Ron Haflidson, "We Shall Be That Seventh Day," in *Deification in the Latin Patristic Tradition*, ed. Jared Ortiz (Washington, DC: Catholic University of America Press, 2019), 169–89.

relationship. This is what it is most fully to eat of the Tree of Life (Prov 3:18; John 17:3; *Gen. ad litt.* 8.5.9–11).

Here we come to what I tentatively propose as a fundamental limitation of the neural network. While *scientia* can be uplifted into *sapientia* within the human mind, I would argue—tentatively—that the interior structure of the neural network, considered as artifactual *memoria*, cannot. This is because, simply, the outputs of the network—when they dictate human action—do so in terms of exterior acts (sell stock), or world-associated categories (thunderstorm). This is why the network's performance is objectively measurable, generating the learning signal by which its weights are adjusted. Such an artifact cannot represent or point to *caritas* because *caritas* is a reality measured not firstly by world-definable ends and actions but by an interior embrace of God as one's friend, even spouse. *Caritas* dictates concrete dispositions in the world but it is not captured by classifications and action decisions. Such a transcendent frame can be declared (i.e., we could train a network to infer and comment upon one's ordering of loves) but it cannot be captured except in terms of its effects in the world. Networks do not have real and subjectively alive relationships in Augustine's sense.

Without a sensitivity to *caritas*, the network cannot become an artifactual *sapientia*. The human observer might reframe the network's meaning beyond its original instrumentality. One might even develop a network to facilitate contemplation of the natural world as a refraction of divine Wisdom. However, as *memoria*—that is, within the intentional frame by which it is trained to map from the world to world-measurable human purposes—it could not be said to represent the *rationes* of created things *as* referring to God. On the part of the human being, the meaning of the network could perhaps be held open to something more, but here it would not be *memoria* but only a sign incomplete in itself because it is unable to accommodate a transcendent frame in the trace of its interior.[64]

THE TWO TREES: OUR CHOICE IN USING AI

It is fitting to conclude this paper by recalling Peter Norvig's suggestion, that we might rightly be satisfied with statistical AI, which "describes what *does* happen" but "doesn't answer the question of *why*," even to the point of bearing no relation to the natural generative processes that give rise to the predicted phenomenon.[65] I suggest that

[64] The question of the network as a predictor and hence a representer of human behavior is intriguing. The love of human beings that frames the meaning of their behavior and their artifacts is ambiguous, almost outwardly incoherent, in that it is shaped both by *caritas* and covetousness, by humble love and the self-defeating autonomy of pride.

[65] Norvig, "On Chomsky." Emphasis original.

such satisfaction would be dangerous, because the opacity of the network's "reasoning" lends itself to becoming a replacement for rather than an invitation to the world, reducing our engagement with the world to the scope of our desires and intentions—to the point that we risk rewriting the world itself as a resource for the accomplishment of our designs, with ourselves rather than divine Wisdom as its ultimate *ratio*. This paper cannot fully expound these familiar themes; here I but gesture to a landscape that demands re-exploration in light of AI.

We deal with the network in terms of outputs for which our own goals are the necessary framing. The network's interior—even as an echo of *memoria*—is recondite. It is manifested first to us by the network's responses to various inputs, somewhat as an animal's instincts are manifested in its behavior. Like these instincts, which must be studied and tested and even then not fully understood, the network—seen from without—suggests its implicit "concepts" but hides them from our view. An animal does not judge the world; it does not theorize about but works *within* the reality with which it interacts. Similarly, the neural network, for all its sophisticated ability to predict data correlations that we might never have imagined, remains in this sense at the level of the animal.

We, on the other hand, ask about reality because it is by judgment that we come to understand. We can see the animal as an invitation to judgment: acting according to its own *ratio*, the animal *is* itself a mapping of the world that we might judge. But the neural network invites our judgment especially because it ultimately concerns us, who have determined the outputs of interest. The network is not a theory or explanation of the world, but *itself* something to be theorized and conceptualized. At best, it may help us to interrogate our own purposes, or it may redirect us (as with the ridged stalks) to the meaning of the world. At worst, it may hide the world from us by hiding its own workings except for its efficacious aid to our own goals.

The interior behavior of artificial neural networks suggests a rich ontology well accounted-for by Augustine's *rationes*. On the other hand, if the opaque network becomes a buffer *between* us and the world, then we risk a very different assumed ontology. The ancient Babylonians, un-wondering masters of data-fitting, sought exacting astrological forecasts but showed no interest in astronomical mechanisms.[66] Their cultural orientation matched their cosmogony, in which the world was fashioned from the bisected carcass of a primordial chaos dragon, slain by her own descendants and held together by ever-

[66] Philip Ball, "Stop Calling the Babylonians Scientists," *The Atlantic*, February 10, 2016, www.theatlantic.com/science/archive/2016/02/babylonians-scientists/462150/; Pearl, "Limitations."

vigilant heavenly guardians.[67] It is not a world that invites science because it is not a world that requites wonder. Redolent of menace, it is not to be understood but only to be controlled.[68]

Our stance of will toward the world shapes our own implicit practical ontology, a reading of the world's *rationes* and a particular actualization of its capacity for meaning. If statistical AI is used as an unexamined instrument of reduction, harnessing the world without understanding, then we shall become practical Babylonians, the instrumental framing of the network dominating our framing of the world itself. As John Cavadini puts it, for Augustine "the sign systems we create are no better than the love in which they were ultimately begotten."[69] A love that values the world merely for its amenability to AI-driven mastery is a love closed to *sapientia*. Such a love will fast decline from *scientia* into mere *superbia* ("pride"), the fatuous science of false autonomy that reduces all to the scope of our perceived desires, so as to live the lie of self-complete dependence on nothing—as if we were gods (Gen 3:5). The artificial neural network can serve our sapiential tasting of the tree of life (although it cannot capture that Wisdom); but, if permitted to delimit our relationship to the world, it will become the Tree of Knowledge, denying to us all that cannot be represented by the instrumental structures by which we have cultivated the network's activity and rendered it intelligible. Is this not the basic dynamic of AI bias? A network tells us what we already "know" because we train it to reduce the world as we do; or it reduces the world in ways we do not notice because our purposes are shaped by the reductive character of our own biases.

What, then, must we do? We must inquire of the world—and we must let the network lead us back to it by inquiring into the network, by striving to understand its working and refusing the easy claim that it bears no relation to the generative processes of nature itself. In that it must reckon in some sense with the *rationes* that have divine Wisdom as their source, a network that cannot accommodate the fullness of the world can still perhaps lead us to it by routes unexpected.

This leads us to the use of artificial intelligence as a spiritual activity. Our behavior and goals are the inescapable framing of the network itself; and so our deployment of these artifacts must imitate God's providential governance of the universe—arranging and further elucidating the *rationes* to yield meanings that they cannot possess simply on their own. As we undertake this godlike activity, will we seek

---

[67] James B. Pritchard, ed., *The Creation Epic* [Enuma Elish], in *The Ancient Near East: An Anthology of Texts and Pictures* (Princeton, NJ: Princeton University Press, 1958), 31–39.

[68] On the political theology of this situation, see Joseph Ratzinger, *"In the Beginning…": A Catholic Understanding of the Story of Creation and the Fall*, trans. Boniface Ramsey (Grand Rapids, MI: Eerdmans, 1995).

[69] Cavadini, "Quest for Truth in Augustine's *De Trinitate*," 436.

greater understanding or only greater efficacy? Let us not be harsh imperators of an unstable world. Instead, let us seek an unveiling of the dynamics of creation for use according to their intrinsic goodness and meaning. The network maps natural things to conventional meanings, but if we return to the world, we prevent those meanings from merely signifying ourselves. The right use of AI does not depend merely on the architecture of our systems, nor even on the ethics that we attempt to embed in them, but on the ultimate stance of will that we adopt—be it *superbia* or *caritas*, unto a false knowledge or a true *scientia* and, finally, wisdom. This is the challenge of AI, our moral framing of which will determine what of reality we permit ourselves to see. M

Jordan Joseph Wales is Associate Professor and the John and Helen Kuczmarski Chair in Theology at Hillsdale College. His scholarship—which has appeared in the journals *AI & Society*, *Augustinian Studies*, and *Cistercian Studies Quarterly*, among others—focuses on early Christianity as well as contemporary questions relating to theology and artificial intelligence. He currently is working on two books—one on the theology of Gregory the Great and another on AI and theology. Holding degrees in Engineering (BS), Cognitive Science (MSc), and Theology (MTS, PhD), he is an advisor to the Holy See's new Center for Digital Culture, under the Pontifical Council for Culture; and he is an affiliated scholar with the Centre for Humanity and the Common Good at Regent College, University of British Columbia.