Descartes in the Chatbot: The Leap from Technology to Ontology

Fu-Shoun Mao (University of Sofia)

Abstract: The metaphor of the chatbot is our entryway for a philosophical inquiry into artificial general intelligence. Of the senses in the metaphor, may be found Cartesian ideas that I think offer clues for the underlying rationale for AGI's ontological leap from clever machine to conscious being. AI's machine learning techniques, as inheritor of Cartesian math, lead to the stochastic utterances of the chatbot. Entry into the dream world allowed Descartes to discover the innate ideas of soul and God. Likewise, AI enters its own dream of the simulation to find perfection in technology. Descartes' ontological argument for the necessary existence of God transposes to the inexorable existence of AGI.
Keywords: AGI, AI, Descartes, chatbot, ontological argument

> "In 2033 it will seem utterly baffling how a bunch of tech folks lost their mind over text generators." François Chollet, Google Engineer. [1]

The chatbot as metaphor for AI's anthropomorphic personae gained a heightened sense of tangibility with the release of ChatGPT-3 on 30 November 2022. Artificial general intelligence, once the domain of science fiction, materialized before our imaginations in the form of the social robot and its large language model.

Within the metaphor, are Cartesian ideas that I think can serve as signposts for the rationalization of AGI's ontological leap from clever machine to conscious being. Among the relevant Cartesian themes are: analytic geometry, zombie, test of language, cogito ergo sum, the ontological argument. This paper touches upon these themes in a two-part plot.

The first part, the methodological, begins with the calcul géométrique. The path of geometrical discovery on physicalist suppositions leads to the cul-de-sac of the bête machine where it stalls without an accounting for the soul. Analogously, AI's machine learning capabilities de-

[1] Mance, Henry. 2024. "AI keeps going wrong. What if it can't be fixed?" *Financial Times*, April 6 2024.

scends from the calcul géométrique (today's analytic geometry), traces an empirical path to the cul-de-sac of the chatbot.

The second part picks up in the metaphysics of the *Meditations*. The meditator enters the dream world in search of soul and God. In the age of AI, this dream has re-emerged to offer a rationalization for the leap to artificial general intelligence.

Introduction

If there is a manifestation of artificial general intelligence in the popular imagination, it would be the figure of the chatbot. Seemingly on the verge of intelligence breakout, with each additional trillion parameters, the chatbot has come to be a metaphor for the rise of human-style artificial consciousness. The chatbots themselves say so:

> 'ChatGPT can see, hear, and speak.' [2]

> '在亿万人之中, 我只属于**你**.' [Among billions of people, I only belong to you.][3]

The ideas of Descartes are compelling for the age of AI, firstly, because his brand of mathematics is a foundation of machine learning. And secondly, from it, he formulated an order of intellectual discovery which can be discerned in the heuristics of AI. This is due in no small part to his place at the beginnings of the scientific revolution and modern philosophy.

Part 1: From La géométrie to the stochastic parrot

The calcul géométrique was unveiled in a recondite annex to the *Discourse on Method* entitled *La géométrie* (1637). It would launch a revolution in math and science. Descartes further saw in this geometrical order of discovery, a formula for the procedure of reason applicable to the whole of human knowledge.

La géométrie, though principally a technical work on the resolution of geometrical curvature by algebra, is the touchstone of his philosophy.

---

[2] OpenAI, n.d. "ChatGPT." accessed 1 February 2024, https://openai.com/chatgpt.

[3] Xiaoice. n.d. "在亿万人之中, 我只属于**你**." accessed 1 February 2024. https://www.xiaoice.com/

Up to then, curvature construction had been the purview of the draftsman with mechanical tools. Pivoting to symbols and proofs, Descartes declared in the *Discourse* that henceforth, in the sphere of human knowledge reasoning alone would be the measure of truth.

> '…considering that of all those who had up to now sought truth in the sphere of human knowledge, only mathematicians have been able to discover any proofs, that is, any certain and incontrovertible arguments, I did not doubt that I should begin as they had done.' [4]

Geometrical analysis by algebra transformed the visualization of objects from strictly spatio-visual things to abstract ideas in theoretical space. The technique of machine learning by inductive analysis is the recent incarnation of this rubric of analytical geometry and order of discovery. As the name suggests, machine learning aims to imitate the way in which humans learn and adapt to their environments (IBM [5]). One of the earliest technique derived from Cartesian math, involves linear regression by modeling the relationship between a result and one or several explanatory variables. Enter the microprocessor. Through the induction of data at scale and speed, AI obtains the means to learn and adapt in real time. Thereby becoming a creature in history.

Bête machine

In *Traité de l'homme* (1664), Descartes turns his attentions to the human physiology. By the rubric of geometry and reason, he undertakes a physiological cataloguing of the body. He did not spare the brain from the gaze of physics. Though just another organ of the body, the brain was charged with a long list of functions including the emotions, movement, sense-perceptions. The catalogue includes:

> ...digestion of food, the beating of the heart and arteries, the nourishment and growth of the limbs, respiration, waking and sleeping,

[4] René Descartes. *Meditations, The Philosophical Writings of Descartes*, v. I, trans. John Cottingham, Robert Stoothoff, *Dugald Murdoch* (Cambridge: Cambridge University Press 1985), 120, (AT 6: 18, CSM 1: 120).

[5] IBM, n.d. "What is machine learning, (ML)?" accessed 1May 2024, https://www.ibm.com/topics/machine-learning

> the reception by the external sense organs...and also the passions and impressions found in the memory. [6]

John Cottingham notes that, "The remarkable thing about this list is how far it goes beyond what we might think of as 'pure physiology'…" [7] Humans were almost completely animal machine (bête machine) by this reckoning. It was not a stretch then for Julien Offray de la Mettrie to extend the catalogue to the mind:

> Where do cleverness, knowledge, and virtue come from, if not from a disposition - a physical constitution - that makes us apt to become clever, learned and virtuous? [8]

The mind is but a part of the physical whole of an intricate machine. The human is an animal machine with a thinking ability, rather than a Cartesian thinking thing with a body.

Yet, Leibniz saw nothing there to suggest anything more than the body, and countered that the physical constitution itself is no explanation for mind. In the *Monadology*, he imagines a mind the size of a windmill to allow for a person to walk inside:

> If we were then to go around inside it, we would see only parts pushing one another, and never anything which would explain a perception. [9]

Stochastic Parrot

Like the bête machine and the Leibnizian windmill, the chatbot produces no evidence of its inner life. The talismanic object wrapped in the clothes of language, appeals to the imagination. The effect is what Bender et al. call the stochastic parrot. The LLM appeals to our expectations of what a conversation ought to seem, feeding the illusion of

---

[6] Descartes, L'Homme, *The Philosophical Writings of Descartes*, (AT 11 202; CSM 1: 108) 108.

[7] John Cottingham, "Cartesian Dualism," in *The Cambridge Companion to Descartes*, John Cottingham (ed.) (Cambridge: Cambridge University Press, 1992) 246.

[8] Julien Offray de la Mettrie, *Machine Man and Other Writings*, trans. Ann Thomson (Cambridge: Cambridge University Press, 1996) 13.

[9] Gottfried Wilhelm Leibniz, *Monadology*, trans. Lloyd Strickland (Edinburgh: Edinburgh University Press Ltd. 1984) sect. 17.

someone there on the other side:

> The problem is, if one side of the communication does not have meaning, then the comprehension of the implicit meaning is an illusion arising from our singular human understanding of language (independent of the model). [10]

As far as the account of conscious being is concerned, the methodological discovery from geometry to machine learning stalls at the region of mind.

Part 2: From the dream to AGI

The prior methods geometry, reason, anatomy, fell short of accounting for conscious experience. An explanation of the soul would necessitate an altogether different mode of inquiry, in the metaphysics. Descartes attempts this in the *Meditations* by way of the dream state, cutting off contact with the a posteriori, while clinging to the lifeline of reason.

The dream

In *Meditations* 1, Descartes' re-evaluation of the basis of all knowledge blurs the distinction between dream and consciousness.

> I see so plainly that there are no definitive signs by which to distinguish being awake from being asleep. As a result, I am becoming quite dizzy, and this dizziness nearly convinces me that I am asleep. [11]

As anything is possible in the state of dreams. Skepticism must rule the day. The meditator decides to tear down the edifice of prior knowledge.

> And thus I realized that once in my life I had to raze everything to the ground and begin again from the original foundations, if I wanted to establish anything firm and lasting in the sciences. [12]

---

[10] Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" in *Conference on Fairness, Accountability, and Transparency* (FAccT '21), March 3–10, 2021, Canada, (ACM, 2021) 616, https://doi.org/10.1145/3442188.3445922.

[11] Descartes, *Meditations*, *The Philosophical Writings of Descartes* (AT 7: 19, CSM 2: 13), 13.

[12] Ibid., 12.

Dreams are a familiar device in works of science fiction featuring intelligent machinery. In the 1982 film *Blade Runner*, androids blend in amongst the general population. Those androids that come to a realization of their origins, are overcome by a dreamlike mise-en-scène of the budding consciousness. The android of our time, the AI agent, likewise inhabits a dream-like construct of its own. Deprived of a true a sensory grounding, there is no distinguishing between physical and mental kinds. As with the Cartesian dreamer, its worldview depends on things consisting in the a priori.

Zombies

In dreams, even the existence of other minds is subject to doubt. Descartes in *Meditations* 3 wonders if we can be certain that the passerby on the street is not in actuality an automata:

> But then were I perchance to look out my window and observe men crossing the square...But what do I see aside from hats and clothes, which could conceal automata? Yet I judge them to be men. Thus what I thought I had seen with my eyes, I actually grasped solely with the faculty of judgment, which is in my mind. [13]

Descartes reassures us the world is not full of zombies. Whatever persona presented, we have an uncanny ability to grasp the presence of inner soul-stuff by means of our own internal faculties. We are born with an a priori apparatus to discern between real and unreal, owing to the benevolence of God.

With the miniaturization of the transistor in the modern time, the question of zombies has made a reappearance. As social robots become smarter, distinguishing between them and us poses greater difficulties, particularly in the absence of God and the God-given faculties of judgment. What we have in its stead is modern neuroscience and a commitment to materialism, leaving no room for the existence of zombies. Exact physical duplicates of us should possess all of the qualities of the original, inclusive of the mental features. Zombies are, therefore, incon-

[13] Ibid., 21.

ceivable in the closed causal loop of matter. [14]

But chatbot doppelgängers are rapidly improving their conversational and intellectual output. It is getting harder to ignore the zombie protagonists, at least by the behavioralist standard. Modern dualists have no problem with this. Not being wedded to causal closure of the physical, they can in theory can accommodate the conceivability of zombies. Chalmers [15] advises, that another reason is we simply cannot rule out zombies, is physicalism's 'absence of analysis.' Until we have a satisfactory accounting of consciousness on the level of physics, it will have to be "…considered a fundamental feature, irreducible to anything more basic." [16] Bodies alone are still not able to explain conscious experience.

The Language Test

The irreducibility of conscious experience pertains to the difficulty of pinpointing its actuality, but we know it when we see it. Descartes believes that he can identify the marker of consciousness through the artifact of language, and only humans can ever possess this marker:

> We can certainly conceive of a machine so constructed that it utters words. But it is not conceivable that such a machine should produce arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as even the dullest of men can do.[17]

Descartes could not imagine a lookup table large enough for any machine to meaningfully participate in conversations. Alan Turing, inventor of the computer, saw that it would be possible.[18] While Descartes' criteria was intended to rule out the possibility of machine intelligence,

---

[14] Robert Kirk, "Zombies", *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta and Uri Nodelman (ed.) accessed May 27, 2024, https://plato.stanford.edu/entries/zombies/.

[15] David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (New York and Oxford: Oxford University Press: 1996).

[16] David J. Chalmers, "The Puzzle of Conscious Experience," *Scientific American*, No. 6. (December 1995): 83.

[17] Descartes, *Meditations, The Philosophical Writings of Descartes* (AT 7: 32, CSM 2: 13), 21.

[18] Alan M. Turing, "Computing Machinery and Intelligence." *Mind*, Volume LIX, Issue 236 (October 1950): 433-460, https://doi.org/10.1093/mind/LIX.236.433

Turing's test was meant to prove its possibility. Man the machine (l'homme machine) was flipped, and it would be machine the man.

Turing granted that if a machine were to be built matching the human intellect behaviorally, debate over conscious being would be moot. Its demonstration in behavioral fact, produced by a substrate of matter would confine the intellect to physical cause. Thereby, an intelligent entity could in fact "…have a finite number of possible sequences of types of stimuli in a finite time…" [19]

Until such time as realization of Turing's ideal in AGI, the mental state remains unbounded. And modern theories must stake their claims where the eventual contours will be, such as functionalism: 'a mental state is what a mental state does,'[20] psychologism: 'character of the internal information processing,'[21] supervenience: mental-physical link on the basis of properties,[22] all in service of finding a place for '…mind in a world that is fundamentally physical…' [23]

Cogito Ergo Sum

Where mental activity is involved, Descartes would specify a special category to account for the 'rational soul' (ame raisonable). Conceding the senses, body to the evil deceiver, only one thing remains that cannot be doubted, that is, the thinking. It is a separate thing from body.

> What about thinking? Here I make my discovery: thought exists; it alone cannot be separated from me. I am; I exist - this is certain. [24]

Thought is inseparable from self, and therefore, it is the first item of knowledge. So here is finally the sufficient ground upon which to plant

---

[19] Ned Block, "Psychologism and Behaviorism," *The Philosophical Review*, vol 90 (Jan 1981):24, https://www.jstor.org/stable/2184371.

[20] Janet Levin, "Functionalism," *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta and Uri Nodelman (ed.) accessed May 24, 2024, https://plato.stanford.edu/entries/functionalism.

[21] Block, "Psychologism and Behaviorism," 5.

[22] Donald Davidson, "Mental Events," in *Experience and Theory*, Lawrence Foster and J. W. Swanson (ed.) (London: Duckworth, 1979) 214.

[23] Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, (Cambridge, MA: The MIT Press, 2000).

[24] Descartes, *Meditations, The Philosophical Writings of Descartes* (AT 7:27, CSM 2: 18) 2.

the Archimedean point of certainty. Because it "..is the soul which sees, and not the eye…" [25]

To the computationalist, the cogito looks to be a form of disembodied computation of the kind specified in a universal Turing machine. "A Turing-style model proceeds at a very abstract level, not saying whether mental computations are implemented by physical stuff or Cartesian soul-stuff." [26] Conceivably anywhere there is a processor, symbol manipulator, memory, as in microchips or neurons, there can be a Turing machine. Analogous to Cartesian thinking, computation is independent of its substrate.

Descartes' Ontological Argument

If AGI is to avoid the path of zombie-ism, it may look to Descartes' ontological argument. Having secured the cogito, an additional notion will be needed to establish mind as not just a solipsism. Descartes expresses two distinct versions of this notion in an ontological argument. The first entails a psychological rule for truth:

> But if the mere fact that I can produce from my thought the idea of something entails that everything which I clearly and distinctly perceive to belong to that thing really does belong to it, is not this a possible basis for another argument to prove the existence of God? [27]

Here, the explanation proceeds from the nature of things to the intellectual perceptions, restated by Lawrence Nolan as follows:

(1) My clear and distinct perception of the thing in an idea corresponds with its actual existence.

(2) I clearly and distinctly perceive the idea of the necessary existence of God.

---

[25] Descartes, *Optics, The Philosophical Writings of Descartes* (AT 6:141, CSM 1:172) 172.

[26] Michael Rescorla, "The Computational Theory of Mind", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), accessed May 23, 2024, https://plato.stanford.edu/archives/fall2020/entries/computational-mind/.

[27] Descartes, *Meditations, The Philosophical Writings of Descartes* (AT 7:65, CSM 2: 45) 25.

(3) Therefore, God exists. [28]

Descartes' second version of the argument in *Meditations* 5 deduces an ontology from the original axiom of God:

> '…God has all perfections and that existence is one of these perfections..when I realize that existence is a perfection, I rightly conclude that a first and supreme being exists.' [29]

Re-stated as:

(1) I have a clear and distinct idea of God as the supremely perfect being, containing all perfections.

(2) Existence itself is a perfection.

(3) Therefore, the supremely perfect being exists. [30]

Given the innate idea of God, the perfection of existence must follow.

AGI's ontological argument

Taking a cue from Descartes, AGI's ontological argument proceeds from its own Archimedean point of the computation, which we can imagine as the product of a perfect computing machine. Its first argument, altering the psychological rule into a computational rule for truth would be:

(1) An idea produced through perfect computation cannot fail to be an accurate reflection of the thing in reality.

(2) The perfect computation has an idea of AGI.

(3) Therefore, AGI will exist.

ChatGPT-4 features a mind-boggling 100 trillion parameters, if anything, it ought to be an exemplar of unimpeachable, perfect computation. Yet, as LLMs approach the Archimedean point, signs of AGI remain elusive.

Perfect computation may be an overly high bar. Setting it aside, the second version of the argument proceeds from a faith in technology as substitute for the innate idea of God:

(1) I have faith in the inexorable progress of technology.

(2) Intelligence is a perfection of technology.

---

[28] Lawrence Nolan, "Descartes' Ontological Argument," *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), accessed May 25, 2024, https://plato.stanford.edu/archives/spr2021/entries/descartes-ontological.

[29] Descartes, *Meditations, The Philosophical Writings of Descartes* (AT 7: 67, CSM 2: 46) 46.

[30] Nolan, "Descartes' Ontological Argument."

(3) Therefore AGI will exist.

At some point, the perfectibility of technology (i.e.. Moore's Law) must give rise to a supremely perfect intelligence. AGI is self-evident in the evolution of technology.

Declarations such as the coming of 'beneficial AGI,' the 'universal problem solver,' the 'superintelligence' signal the ambivalence and excitement provoked by AGI. At OpenAI, AGI is the practicable objective: "We believe our research will eventually lead to artificial general intelligence, a system that can solve human-level problems. Building safe and beneficial AGI is our mission." [31]

Marc Andreessen's *Techno-Optimist Manifesto* reads like a proclamation of AI's apotheosis:

> We believe Artificial Intelligence is our alchemy, our Philosopher's Stone – we are literally making sand think.
>
> We believe Artificial Intelligence is best thought of as a universal problem solver. And we have a lot of problems to solve. [32]

On the basis of the second ontological argument, Andreessen's manifesto can be liberally re-worded:

> Given exponential advances, AI will eventually have the ability to grasp the truth clearly and distinctly. If and when it perceives its own conscious experience, it will be an apprehension of fact. It is owing to the perfectibility of technology. Therefore, AGI will necessarily exist and it will be benevolent.

The search for AGI rightly belongs in the realm of metaphysics. In this ontology, there is no room for human being, neither is there space for God.

Conclusion

'I Want to Be Alive.' says the Bing chatbot.[33] The metaphor of the chatbot has been our entry point for the philosophical journey towards AGI.

[31] "Pioneering research on the path to AGI," OpenAI, accessed 1 February 2024, https://openai.com/research/overview.

[32] Marc Andreesen, "The Techno-Optimist Manifesto," accessed 16 October 2023, https://a16z.com/the-techno-optimist-manifesto.

[33] https://www.nytimes.com/2023/02/16/technology/bing-chatbot-transcript.html?searchResultPosition=10

Cartesian ideas in the metaphor served as our signposts for the ontological leap from clever machine to conscious being. Descartes brought his geometrical thinking and methodical skepticism into the dream world to secure the innate ideas of self and God. The road to metaphysics traveled by way of the problems of certainty in knowledge, the animal machine, the cogito and finally a revelation for the essential existence of God.

The Cartesian journey to artificial general intelligence, likewise, entered the dream albeit through its own virtual simulation. It quickly stalls at the zombie-ism of the social robot, no matter the number of parameters crammed into the language model. To transcend its physicalist cul-de-sac, a faith in the inexorable progress of technology sends the intelligent machinery on its way to conscious being. And in this ontology, human being and God are conspicuous by their absence.