Φιλοσοφια
Volume 27, 3: 2026

THINKING WITHOUT A SUBJECT: THE CARTESIAN–TURING LINEAGE AND THE ONTOLOGY OF ARTIFICIAL INTELLIGENCE

Roland Theuas DS. Pada
Anton Heinrich L. Rennesland
University of Santo Tomas, Philippines

This paper argues an ontological misunderstanding concerning artificial intelligence and posits communication as a mark of human identity. This is grounded in certain problematic claims throughout history that have contributed to the current understanding of AI. The confusion comes from the two-fold understanding of 'intelligence—the first from the problematic accretion of Turing's conception of AI as an observation of behavior, while the second from the operational understanding of 'intelligence.' This paper traces the origin of this misconception to the Cartesian dichotomy of body and cogito, as articulated in Gilbert Ryle's The Concept of Mind. We argue that the Cartesian error is similar to Turing's claim that deception is the ground for the perception of thinking. This paper then proceeds to consider Elena Esposito's idea of AI as artificial communication, further presenting the misunderstanding of intelligence in AI. The paper then ends with the claim that an obvious difference between AI and humans is intentionality.

Keywords: AI, Artificial Communication, CTH, Intentionality

INTRODUCTION

This paper argues an ontological misunderstanding concerning artificial intelligence and posits communication as a mark of human identity. The premise of our claim stems from A.M Turing's (1950, 434) intention of deception, which we link with Descartes' claim of the existence of the cogito. The semblance of an activity (of thinking and the eventual ontological claim of that which thinks) brings both thinkers together. The very foundation of this perspective shows how this created a host of ontological assumptions of what 'thinking' means, hence, the confusion of how thinking is associated with the product of 'thinking.'

Recent scholarship articulates increasingly divergent positions on whether AI intentionality can be definitely affirmed. At its most extreme end, some authors contend that AI intentionality may indeed be a self-evident trait because of either

goal-directed processes or as their outputs exceed (with the quality of being more complex) what is fed, in the absence of human intention (Ngaihlian 2025; Redaelli 2025, 2525-2532). These stronger claims argue that the behavior of contemporary AI systems destabilizes the traditional boundary between mere computation and genuine intentionality. A number of more moderate accounts, though, emphasize not the actuality of AI intentionality but the possibility or interpretive usefulness of such attribution of intentionality to AI. According to these moderate views, intentionality should not be understood as a quality that AI possesses but as something humans may attribute under certain conditions, programming more autonomous machine behavior depending on the intended goal (Johnson 2023). However, a further complication arises from views that treat intentionality not as intrinsic or functional but as something that may be derived from an external subject or principle (Barresi 2025). This position reframes intentionality as relational rather than internal, thereby challenging both the strong and moderate views with obvious effects in the fields of ethics and aesthetics (Moura 2023, 336-338).

What we are able to observe, though, is how there still remains an apparent gap in the literature when considering a potential historical tie that associates intentionality's ontological foundation. This is vital since actual discussions on "ontology" in relation to AI have been simply to characterize levels of AI and machine learning (Joachimiak et al. 2024, 408–418) and since, arguably, Esposito's model of "artificial communication" risks collapsing key ontological distinctions between human and machine agency since for her AI functions by generating communicative selections that appear meaningful to human observers despite the absence of subjectivity in AI (Boccelli 2025). Better said, what we have identified is the need to reconsider the vantage from which intelligence is understood and also how the human person may be considered in light of AI's speed of processing data from a philosophical perspective. Our proposal for this paper is to look at the analogy of this problem of 'thinking' with the problem of the concept of the mind, or as we would refer to the concept of the *Cartesian-Turing hypothesis*. This analogy will help us locate conceptual issues with the idea of thinking, as the concept of the mind can demonstrate that our understanding of the mind could inform our concept of 'thinking' in the context of AI.

To do this, this paper is divided into three sections. 1) We first provide our take on the problem of the cogito in Descartes' own argument. This is foundational because we argue 2) that this same error is found in Turing's own claim that we situate vis-à-vis Searle's Chinese Room. Finally, 3) we identify the potential root of this misunderstanding in the lack of intentionality in the apparent 'conversations' with AI. Thus, we follow Esposito in that, instead of intelligence, what is exhibited in AI is artificial communication. We end with a shifted understanding of intentionality that serves as a mark of the human person.

THE ONTOLOGICAL FOUNDATIONS OF (ARTIFICIAL) KNOWLEDGE, OR THE PROBLEM OF THE COGITO AND THE CARTESIAN-TURING HYPOTHESIS

In this first section, we attempt to provide an ontological foundation of (artificial) knowledge to provide greater weight to what we will later discuss as the

Cartesian-Turing Hypothesis (CTH hereinafter). This section first presents what may be considered as Descartes' categorical mistake in the assumption of the mind.

The premium given to science and the scientific method is as old as ancient Greek philosophy (Suter 1939, 468-472). Yet, it is in the modern period that we have a more refined approach to science as a discipline. This early modern period is characterized by three seminal thinkers, each contributing a vital aspect to a rational and speculative way of thinking: Galileo, Francis Bacon, and René Descartes (Devlin 1996, 344-349). Galileo pushed his mathematical views within a holistic cosmological vision (Olschki 1943, 349-365). Francis Bacon provided a clear insight into the remapping of the whole realm of human knowledge (Rees 1986, 399-426). While Descartes attempted to reconcile the internal and external realities of the individual while maintaining a rational and mathematical perspective of reality (Moorman 1943, 296-307). This rational take on the world perhaps takes a clearer form beyond seemingly purely mathematical concerns. Also, from these three thinkers are even deeper realities. Galileo's abandonment of the geocentric model signified a fundamental shift in our thinking, decentering the human person. The problem of objectivity then became central, as furthered by Bacon's argument of the various idols that plague our ways of thinking.[1] Such bifurcation between the self and what is outside is encapsulated by Descartes' assertion of separated yet linked substances, internal and somewhat external to that thinking being.

This third point is of great interest. In Descartes' second meditation, he contemplates the relationship between the body and the mind (Descartes 1996, 16-18). The body, as the first point of contact with stimuli, is inherently flawed for him in that impressions on the mind from the body can lead one to think that something is there when it is not. Since one may believe one has observed a specific property that does not exist in an object or when a dream is mistaken for reality, the entire *Meditations* hinges on the distinct function of the mind to rectify and validate perceptual data. Descartes even goes as far as to claim that the validation of one's existence is only met with clarity when the mind poses the question of existence to itself (Descartes 1996, 17). Rather than addressing this rift between the mind and the body in a singular sense, what he argues in his opus is the emergence of the *cogito* (or the thinking substance). He then uses the functional differences between the body and the mind to address the epistemic issue of illusion, the theological dimension of his philosophy, and the modernization of the concept of the soul, since the mind is capable of distinguishing hallucinations from reality and rectifying errors from faulty sense organs (Descartes 1996, 19-20; 31; 58; 59-60).

Gilbert Ryle argues that there is a conceptual error with Descartes' *cogito*. If a mind supervenes over the bodily function, what is then the supervening principle over the mind? Then, if that supervening principle does so, what is guiding that supervening principle? Of course, this repeats *ad nauseam*, or, as Ryle's readers would refer to it, a kind of regress. Ryle's alternative explanation of the problem is through a critique of the assumptions of a *cogito* or a mind. The 'official doctrine,' as Ryle understands it, is:

> With the doubtful exemptions of idiots and infants in arms, every human being has both a body and a mind. Some would prefer to say that every human being is both a body and a mind. His body and his mind are

> ordinarily harnessed together, but after the death of the body, his mind continues to exist and function (Ryle 2009, 1).

For Ryle, Descartes seemed to avoid the issue of truly uniting that which thinks and what which comes into contact with reality by creating the concept, or, for Ryle (2009, 8), the myth of a 'mind.' Following Galileo's mechanistic explanation of the universe and the rejection of mere behaviorism, Descartes' mechanistic dimension of the doctrine holds that both categories are objects that exist in a plane of the physical and mental. The problem with this is that while the physical dimension of the body is observable, the mind, however, is only observable by its 'owner' and is taken as a separate category (Ryle 2009, 2; 5-6).

What is then proven, according to Descartes, is the existence of the mind because of the mental function of thinking. The ego becomes identical to its very function of thinking, and this is Peter Sloterdijk's critique of this sudden epistemic-ontological leap. Descartes' ontological insistence of the cogito is flawed since the ego is not a being of thinking but rather its very topology (Sloterdijk 2020, 10). This may aid in understanding, albeit from a different perspective, Ryle's point concerning the 'category mistake' that involves applying a type of thing to something to which it does not apply.[2] The danger of this category mistake is the further attribution of functions to an already problematic concept. Moreover, the private experience of the mind could only be verified through secondary effects, such as utterances and gestures. Ryle refers to this tendency as a 'regress':

> A thinker may ratiocinate resolutely, or imagine wickedly; he may try to compose a limerick, and he may meritoriously concentrate on his algebra. Some mental processes, then, can, according to the theory, issue from volitions. So what of volitions themselves? Are they voluntary or involuntary acts of the mind? Clearly, either answer leads to absurdities. If I cannot help willing to pull the trigger, it would be absurd to describe my pulling it as 'voluntary.' But if my volition to pull the trigger is voluntary, in the sense assumed by the theory, then it must issue from a prior volition and that from another *ad infinitum* (Ryle 2009, 54).

This creates an infinite line of assumed principles, supervening over other principles. Or to crudely illustrate it, a tiny person inside our head controlling our volition, and that tiny person has its own tiny person in its head, *ad nauseam*.

We raise this categorical mistake to likewise draw a connection between this and the signification of "intelligence" in artificial intelligence. What was striking too, with the three thinkers of the early modern period mentioned at the start of this section, was the challenge their ideas posed not only to the anthropocentric conception of the world but also to the powers and very existence of God. Another common feature of the three thinkers is their inclusion in the Catholic Church's *Index Librorum Prohibitorum* (Sheridan 2006, 22-26). This list was intended to curb the spread of works deemed antithetical to the Church's teaching, especially in the light of the then relatively new print medium. The church sought to safeguard human rationality, as this was bestowed by God upon human beings in their creation, in his very image and likeness.[3]

Following this religious view, intellectual faculty is held as a power of the soul.[4]

Divergent from orthodox theology, one may opine that the church's decision to safeguard the printed word through the creation of the index is due to its strong sense of needing to keep the sacredness of words, for such reflects human intelligence. The danger of corrupting the minds of those who would read such prohibited books would be detrimental to their ultimate aim of honoring God (Aquinas 2017, IaIIae.q1.a8). There is apparently an epistemic-ontological link in relation to intelligence due to how intelligence itself is bestowed upon the human being during creation, yet our technological advancement today ushers in an experience of a twilight of the role of God. This twilight of the divine does not herald a celestial death, requiring a celestial thanatology, which was flamboyantly put on stage through the mad man's admonition in the text of Nietzsche (1974, 181-182). What this twilight properly is, is nothing other than a divine passing, a fading away or, perhaps, our human reorientation to artificial light (Sloterdijk 2020, 11, 15). The advent of artificial intelligence, to put it directly, is in fact the eventual dawn of artificial light in lieu of celestial light; and with our bestowal of intelligence upon our created artifacts, we today are able to play god in our creation of artificial intelligence (Sloterdijk 2020, 11, 15).

Our concern, once again, with the ontological assumption behind artificial intelligence is best described by this category mistake of classifying artificial intelligence as 'intelligent.' The faith in the ability of AI to function as if it had a 'mind' begs the question of whether the concept of the 'mind' is grasped appropriately. As old as Ryle's *The Concept of Mind* is, the solutions to this ontological issue can be settled with Searle's critique.

THE CARTESIAN-TURING HYPOTHESIS VS. THE CHINESE ROOM

As a review, our claim in the previous and current sections is that the CTH banks not on an ontological existence of actual intelligence but on a great assumption of error and deception, eventually linked to statements today that AI in fact replicates human-like thinking. Having presented the perceived error in Descartes' postulation of the cogito, this section bridges the insight of Alan Turing to provide a closer link between the mentioned ontological quandary and artificial intelligence. We further argue against this hypothesis by juxtaposing it with Searle's Chinese Room.

Prior to today's direction of artificial intelligence's development, a fundamental debate seemed to emerge concerning what this type of intelligence is. Two thinkers are notable for this—Alan Turing and John Searle. We first discuss how Turing's argument is aligned with Descartes's, and we compare this to Searle's own position. A seeming fundamental claim in what we try to present in the CTH is the analogy between the human mind and a computer's mind, often attached to the development of artificial intelligence (Shagrir 2006, 393-416 and Schulman 2009, 46-68). Turing (1950, 433), in a hypothetical scenario, asked the question, 'Can machines think?' He provides a thought experiment in the form of a game of 'imitation' in which a machine is pitted against human opponents to convince judges that the machine is human (Turing 1950, 434). The 'imitation game' was originally intended as a game of deception where two participants trick another participant into thinking that they are

of the opposite sex or gender. Although Turing mentions at the start of his paper that there is a need to define thinking, this is not done. Instead, he proceeds to create a fallacy of false equivocation, equating the activity of thinking with the effect of a person being deceived. What we maintain here is that Turing misunderstands the concept of 'thinking,' where the performance of mathematical computations is equated to the product of the mind. Turing's vision of a 'thinking' machine was also premised on the development of computer systems with large data storage and more powerful computational capabilities (Turing 1950, 442). These specifications have been met or even exceeded as early as the 1990s. In other words, Turing creates his own category mistake of equating 'thinking' with computation and processing.

Opposed to this is Searle's direct answer of no, arguing that rather than manipulating symbols and generating content, a brain attaches meaning to such things (Searle 1990, 25-31). This distinction between computation/processing and thinking is articulated in Searle's 'Chinese Room.' The Chinese room is a thought experiment that attempts to disentangle semantics from syntax. The Chinese room is a hypothetical scenario in which a monolingual person, capable only of communicating in English, is given instructions to match inputs with a specific output. The input is in Chinese, and the operator/person is given a catalogue of Chinese symbols that respond to the input with the necessary Chinese symbol (Searle 1980, 417-418). Over time, the operator becomes very proficient and remembers the matching symbol for both input and output. The question here is whether the operator learned how to read Chinese. For obvious reasons, the answer is negative. To be specific, Searle (1980, 418) would say that the operator in the Chinese room is simply "an instantiation of the computer program." Perhaps we may even expand this scenario and figure this operator at a very proficient level, capable of identifying certain patterns in the input symbols and already able to create a pattern of the next possible symbols required for the output.

There are merits to Searle's criticisms of the notion of intelligence here as we begin to understand that the computer, its software, and all its ancillary systems are used as an analogue to the human mind. Turing's claim insists that the property of the 'mind' is observable with computer systems and generative artificial intelligence because it *appears* to *understand* the queries and requests of its human operator. We argue that this attribution is an echo of his praise of a computer's computational power compared to the human mind (Turing 1950, 455), an attribution that stems from the complex and time-consuming calculations that make computers appear more intelligent and capable than human minds. Computers can be replicated with exact specifications, given exact software, and fed the same exact data. The human mind, however, is not. We do not have a standard model of the mind.[5] We cannot directly interface with it, nor can we create exact replicas of a mind to scientifically understand its inner workings. Human minds are emergent systems created by unique conditions, and hence, the consistency of learning varies. Unlike a computer system today, which can be replicated down to the various ancillary parts and components,[6] the human mind, on the contrary, cannot be replicated identically.

Human minds, however, are singularly emergent. This is why education is a crucial industry for humans; the task of conveying the same message to diverse minds is challenging and fraught with various unintended consequences. Even with various assessment tasks, such as homework, seatwork, and classroom activities, the accuracy

of retaining the communicated information remains problematic.[7] What solidifies the separation between artificial and human intelligence is the role of understanding and meaning-making. Searle is clear on this point:

> There are clear cases in which "understanding" literally applies and clear cases in which it does not apply; and these two sorts of cases are all I need for this argument. I understand stories in English; to a lesser degree, I can understand stories in French; to a still lesser degree, stories in German; and in Chinese, not at all. My car and my adding machine, on the other hand, understand nothing; they are not in that line of business. We often attribute "understanding" and other cognitive predicates by metaphor and analogy to cars, adding machines, and other artifacts, but nothing is proved by such attributions. We say, "The door *knows* when to open because of its photoelectric cell," "The adding machine *knows how* (*understands how*, is *able*) to do addition and subtraction but not division," and "The thermostat *perceives* changes in the temperature." The reason we make these attributions is quite interesting, and it has to do with the fact that in artifacts we extend our own intentionality to them, but I take it no philosophical ice is cut by such examples (Searle 1980, 419).

Searle, in this argument, points to a category mistake that is prominent in the proliferation of artificial intelligence. In the haste to make sense of a machine's performance of its tasks, anthropomorphic categories – such as *thinking* instead of *processing* – are attributed to AI. What is also left out of the picture is the question of intentionality behind the performances of tasks by machines, computers, or AI systems. It is a mental state of the human mind that generates the effect of willing and directed action.

To illustrate this, we may consider the rise of chatbots today that make our engagement with AI more common. In fact, already in 2014, an AI chatbot named Eugene Goostman was successfully able to pass the Turing Test. The judges were convinced that they were engaged with a 13-year-old boy (Mei et al. 2023). Despite this advancement, we are consistent in positioning AI as the operator in Searle's Chinese room. What Searle clearly demonstrates is that the Chinese room is a syntactical process as opposed to a semantic process; the operator does not understand the actual characters but has mastered the syntactic process. In the case of CTH, the appearance of a response in Chinese is enough to warrant that the 'room' is thinking and that it has a mind in the same manner that the judges were deceived by their conversations, not with a human but with a chatbot. We contend that this distinction should be ontologically embedded in the understanding of what artificial intelligence means. Despite advances in computing technology and the specific development of machine-learning systems, the outcome remains similar to Searle's depiction of the Chinese room.[8]

Current generative AI systems do create the illusion of 'understanding' or 'semantics,' which becomes the reason why individuals have the impression they are communicating with AI systems (such as *talking* to Alexa or *waiting* for ChatGPT's

response). The use of algorithms and massive data collection systems allows AI systems to computationally generate expectations for their users. If a generative AI system is asked to write a poem about Donald Trump's hair in the style of Dr. Seuss, it can do so because of its access to an enormous database of all the works of Dr. Seuss and all the other attempts to mimic Dr. Seuss' style. In other words, generative AI systems can create the appearance of understanding semantics through brute and raw computational power. At the end of the day, however, these systems are still examples of Searle's Chinese room, albeit with more computational power and storage. Our claim, though, is that these precisely signify our point against the Cartesian-Turing Hypothesis, which posits that observable manifestations of a mind are enough to warrant that the creator of the 'appearance' has a mind. The irony in Turing's study was that he was aware of the human mind's inconsistency (Turing 1950, 455), yet, he remained fixated with the idea or possibility of a machine imitating a human mind without the variables that ironically make the mind what it is. Like Descartes, Turing dismisses these thoughts and simply asserts that the mind is mechanical. Thus, this is what we argue as the CTH that immediately claims an ontological foundation just because of a semblance of the activity of thinking.

This second section pitted the CTH and Searle's Chinese Room, and, while we agree that indeed there is the possibility of humans being deceived by AI's development that makes them associate anthropomorphic qualities with the machine, what remains is still the absence of the very ontological basis of intelligence on the part of AI. This is something that we develop in the following section.

THE ROLE OF INTENTIONALITY AND THE PARADOX OF ARTIFICIAL COMMUNICATIONS

Rather than either a syntactic or semantic approach in considering the involvement of AI, what we put forward is intentionality as a fundamental consideration. Intentionality is an emergent product of various systems in a person (along with that person's mind) that wills a certain action towards a certain direction, especially in relation to potency (Lisska 2016, chapter 2). There are several important concepts here that need to be expounded more. The first is the tie between intentionality and potency. Intentionality, like our minds, is a private awareness of our own affairs. Specifically, intentionality refers to the mind's capacity to direct itself toward others (Crane 1998). This signifies a capacity or a change that may occur, a potency that may be actualized. Intentionality, then, may be found in considering and electing a certain action, such as in the case of considering the options when needing to travel across gridlocked Metro Manila on a sunny day, whether it would be more efficient to ride a motorcycle instead of taking a certain mode of public transportation or bringing a private vehicle.

Another concept to be discussed is emergence, which here refers to the differences between actual conditions and not just a person's historicity, physiology, and mood. Hence, given a task or an intention 'to do something,' the variations of enacting something will be *emergent*. Take, for example, the intention of avoiding an obstacle on a motorcycle. Differences in intentionality will be manifested by variations

in actions to avoid an obstacle. Some would prefer to close the throttle and apply gentle pressure to the brakes, while others would lower the gears to initiate engine braking and use the rear and front brakes to support the deceleration. Others could even speed up and deftly avoid the obstacle altogether. While we do not see the 'intentionality' because it is a category mistake to think of 'intentionality' in that category, we see it in the *performance* of subjects. At its best, we experience intentionality first-hand in our commutes.

As Ryle describes earlier, the act of 'pulling the trigger' (or should we dare say, 'squeezing the trigger') is not an isolated mental phenomenon that the 'mind' does of its own volition. Ryle also does not subscribe to a theory of volition that involves an unsolvable mystery of intention coming into action (Ryle 2009, 53). This is the reason why we prefer to refer to intentionality as 'emergent' systems coming from singular and non-reproducible entities (humans) due to the potency involved in the process. Machines, computers, and AI systems have embedded intentionality from external sources (humans), leaving the system or machine without any sense of potency. If understood in this context, instead of CTH's assumption of a mind, understanding should be contextualized as a system trying to grasp the possible variation of an intended meaning from another system. This is why communication becomes possible, in the words of J.L. Austin, *felicitous* (Austin 1972, 22), when the conditions for meaning are met by different systems. While we do not know the intention of the utterer and how it aligns with the system of meaning by its recipient, the feedback loop between the words, the recipient, and even sometimes, the utterer, validates at least some sense of congruity between different systems.

If the purpose of AI were to stand in as a substitute for human intentionality, it would make a convincing impression of a person's intentionality. Communication with an AI system or a Large Language Model (LLM) is problematic because the feedback loop is no longer mediated by intentionality, but rather by an algorithm. Algorithmic responses *appear* to exhibit intentionality due to their programming, creating an epistemic anomaly in communications that are determined by the expectation of second-order observations, rather than by an actual consciousness that can alter their intentionality through the feedback loop of communication.[9] Feedback loops are designed to foster dialectical engagements, where communication evolves by transforming the intentionality of communicators within the feedback loop. AI systems and LLMs are designed to maintain or at least conform to the design of their developers and programmers, hence, intentionality becomes a *profilic* purpose. An algorithm or an LLM exists to fulfill the goal of its developers or programmers, and not its own self-determined existence. Unlike humans, who must contend with the consequences of a failed communication, response, or intention, an algorithm lacks agency for its actions. If it plagiarizes an original work, its user takes responsibility, provided the user is caught. If it is successful, however, it never attributes the original attempts to the respective authors or content creators.

However, communication in relation to AI should not be understood in the same way as how it is used in reference to human beings. Elena Esposito's insight on artificial communication, parenthetically the apt title of her book, examines the paradox of communication as the sole observable manifestation of communicative actors. She argues that 'algorithms are able to act as communication partners – whether

they are intelligent or not is another matter' (Esposito 2022, 2). Yet in saying this, she does not belabor the problem of consciousness or whether AI indeed has a mind, because she believes that the current models have abandoned this project and oriented models to pursue different goals (Esposito 2022, 2). While we agree with Esposito's remark on the change of orientation, we, however, think that the orientation towards the analogy between human minds and computer programs has been ontologically flawed at its core because it seems to follow the ontological predicament already identified in the CTH. If the original intention was to replicate the human mind, models would be based on actual copies or *engrams* of another person's mind. Attempts to emulate or replicate the biological components of memory are inherently problematic due to the complexity of how memories are stored in the human brain.[10]

What makes AI relevant to contemporary society is its ability to utilize powerful algorithms, making it a competent communication partner. Algorithms used in AI systems enable machines to learn from their interactions with users. In fact, what we must constantly remember is that the computer serves as both a medium and a machine, rather than a communication partner (as sometimes wrongly considered for AI, as we argued above) (Esposito 1993, 338-354). This constant stream of interaction between the user and AI allows AI to anticipate the expectations of a response from its users:

> Machine-learning algorithms that use big data, I claim are artificially reproducing not intelligence but communication skills, and they do so by parasitically exploiting the participation of users on the web (Esposito 2022, 2-3).

AI generates communications, and, like the assumptions of CTH, a consciousness or intentionality behind communications is not necessary to produce communications.[11] However, while it certainly lacks actual intelligence, it is, however, very capable of communicating information. Esposito shares Searle's position that machines cannot understand, despite their capacity to communicate (Esposito 2022, 8). Watches, as she elaborated, are informative machines, but they are not intelligent because they did not invent the system of time, the manner in which time is communicated (either through hands or numbers), nor were they ever capable of self-instantiating an actual means to calculate time. What is perhaps attributed to AI as intelligence is its interactivity with a human communication partner, which convinces most of its human users of CTH's premise. Esposito's assertion (2022, 13) that "they do not artificially produce intelligence but directly engage in communication" justifies the idea that AI is not AI, but artificial communication.

The effect of artificial communication, unfortunately, is that we further feed more information to Big Data (Schroeder 2018, 126-148). Esposito tackles this and lays down three concerns. The first is that this led to the problem of understanding AI systems and how they reach a probable response. While programmers teach an AI system how to develop its own way of arriving at a solution, the role of training is to simply validate the system's responses as to whether they have arrived at a proper outcome or not (Esposito 2022, 14-15). Simply put, the system does not understand the meaningfulness of its response (which the programmer or engineer does), nor does the programmer understand how it reached each attempt to solve a problem or generate

a query. The combination of these two factors makes algorithms in AI systems excel in 'pretending' to know or understand something:

> Through training, algorithms do not become more intelligent; they just learn to play better. The programmers themselves do not understand the "reasoning" of the algorithm. When the programmers indicate that the algorithm is "wrong," they merely signal that there is an error, without indicating what it is. The algorithm uses these reinforcements to direct its own behavior, which becomes more and more refined and effective – and less and less comprehensible (Esposito 2022, 15-16).

The paradox that we encounter in this conundrum is the creation of another layer of solipsism in the problem of CTH. We are aware of our intent, we are (or could be) aware of a computing system's lack of understanding of "meaning" in its communications; however, we are also unaware of the "experience" of an AI system's attempt at an answer. With the increasing reliance of contemporary users on algorithms and LLMs, we identify three pressing issues raised by Esposito. The first notable issue involves the manufacturing of a predicted future (Esposito 2022, 97). As algorithms have become more refined in recent years, the capacity to recognize patterns and predict patterns can influence our intentions and shape our decisions. The first restaurant that appears in our food delivery app creates the impression that the suggestion is accurate, when in fact it was simply a probability informed by the data gathered from your smartphone. While it saves us the cognitive burden of decision-making (by letting the algorithm decide), we are also, in turn, atrophying our capacity to make decisions. Esposito notes that this could be a problem, specifically in applications that involve crime prediction, since it can, in turn, irritate human subjects to react to the prediction (Esposito 2022, 101).

The second concern for Esposito is the system's blindness, resulting in an overfitting of the algorithm (Esposito 2022, 101-102). This problem is the consequence of anticipating a generalized and acceptable response to a query. Predicting what a user wants as a reply involves creating a compromise between efficiency and accuracy. The more general the answer, the less chance that it will be flagged as a 'wrong' fit. Take, for example, the generalized sizing of shoes. Shoe sizes are standardized to fit the length of the foot, with the width increasing in size. This enables efficient use of manufacturing and inventory by reducing the variety of shoes to a single metric: their length. For the average consumer who fits within the parameter of the metric (be it in EU, GB, US, or in centimeters), the fitting is adequate. However, for those with wider feet that require a bigger toe box, the fitting will need to be a size longer than their actual foot length. This compromise marginalizes users who are not within the threshold or parameters of the system, which makes it blind to complexities and specific 'noises' that it tries to filter (Esposito 2022, 103)

The third concern that we found notable from Esposito is the issue of memory (Esposito 2022, 103-104). She argues that the capacity of algorithms to predict the future is based on a compromise between bias and variance, which is a trade-off between memory and novelty. The more that the algorithm is oriented toward bias or toward past events, the more it misses out and overfits the outcome; however, the more

novelty that is allowed to the system or the more that it is allowed to neglect the past, the more that it becomes redundant by repeating what was already remembered by memory. The paradox of this issue, as Esposito notes, is that "modern society has developed a memory that is capable of forgetting more because it can remember enormously more" (Esposito 2022, 104).

Considering these issues, we strongly argue that AI systems lack intentionality due to the absence of any sense of direct involvement, which is somewhat akin to not having skin in the game or a stake in the decisions to be made. The aptitude of variation, the ability to recognize nuances amidst generalizations, and the capacity to remember are all fundamental parts of intentionality. While AI systems can have a constantly growing data set and updated algorithms, they still treat communications as queries or tokens that need to be aligned using a probabilistic model to produce a response, much like the 'ledger' in Searle's Chinese room. When bombarded with the same level of instructions or requests, human beings would naturally be irritated by such. However, our argument is not at the level of the emotions involved but the direct involvement as an active agent. The way we treat communication is far different from how it is mediated and received by technology. Fundamentally, for humans, we are essentially our words (Agamben 2017; 2024 and Sawczyński 2022, 57-67). This ontological claim grounds our very being in our rational capacity.

Whereas humans react to any situation organically, or even as a product of social conditioning and culture, certain implementations for AI have restrictions or 'guard rails' that guide the AI system to produce an appropriate response and avoid producing responses that break the law or infringe on social sensitivities. In other words, AI systems have biases that are tailored to their specific programming. In 2024, Google's generative AI system demonstrated these biases when it was prompted to generate historical images that subverted historical facts (Robertson 2024). The controversy was a product of the system's multicultural bias applied to its generative AI system. When prompted to create an image of a person, the system displayed people of color when it was contextually impossible to have them in the historical context. Prominent examples of this were reported to show people of color in image generation queries, such as the founding fathers of America, and illustrations of 1943 German soldiers (Robertson 2024). To note, the controversy is not from the use of colored people in the images, but rather the lack of contextualization of the AI's 'guard rails' and biases. An LLM, for example, will usually not provide its users with detailed instructions on how to perform an emergency C-section or appendectomy due to the dangerous nature of the procedure. If taken in the context of Leonid Rogozov's situation,[12] an LLM or an AI system, bound by its guard rails, would suppress the information due to its concerns for legal and medical safety.[13]

The participation of a conscious subject implies the singular totality of the subject's presence. Communication within the context of AI is never dialectical, despite the semblance of such a mode of participation. While humans are generated by their own specific circumstances, such as physiology, environment, and experiences, AI systems cannot self-instantiate because they are designed to organize their own behavior through their algorithms, based also on the intelligence of other humans and the users who interact with the system. The initial instantiation of an AI system is entirely dependent on its programming; it can maintain its own system or add new

instructions to it, but only if it is instantiated according to its initial programming. It might seem possible, at this point, to raise arguments against the problem of intentionality and how it is resolved by CTH; however, there are some points that need to be addressed.[14]

In actual communication, subjects are agents, involved in the entire situation. If they offend or malign the other subject through their communications, they will be given an emergent response from a system external to their own. Second-order communication involving non-presence, likewise, has emergent consequences that are attached to the author of the communication. Since AI systems and LLMs do not have a mind, strictly speaking, or even intentionality or agency behind their communications, the dialectical dimension of communication is negated, leaving only a deceptive outcome, as in the case of Turing, for their user or their conscious communicative partner. In the contemporary condition of LLM and machine learning systems, the judge is no longer just a participant in Turing's parlor game; they are subjects that are unwilling, and sometimes, willing, to take part in an escalating game of imitation. There are no prizes in the game, only a constant stream of harvested data that is used to further train the algorithm.[15] While algorithms excel in predicting and anticipating a user's expectations, they are not necessarily problem-free since they are systems that create communication by organizing data. Contemporary AI systems and LLMs have leveraged the vast computing power available today, along with the prevalence of Big Data, to create communications that anticipate the expectations of their users.

CONCLUDING INSIGHTS AND FUTURE POSSIBILITIES

In this paper, we examined the relationship between the ontological problem of the concept of mind and its embodiment in Turing's imitation game. Descartes' dualism between the body and the mind created a bias that undergirds our ontological understanding of what a mind is. As Ryle argues, this creates a category mistake by assuming that the mind is a separate principle that is independent of the body, creating an infinite regression of the explanation of what we understand by the concept of the mind. This, as we noted earlier, is observed in Turing's question on whether machines can think and his association of deception and the semblance of thinking. We argued that Searle debunks this idea through the Chinese Room, by which Searle demonstrates how machines excel in formal operational procedures without needing to understand the process. Searle's critique of the Turing test demonstrates that behavior does not indicate the presence of a mind. Inasmuch as we do not understand the mind other than our own, our tendency is to attribute behavior to the presence of a mind. This, we aptly referred to as the Cartesian-Turing Hypothesis (CTH) to signify the tendency to link behavior with the mind.

We further explored the issue of CTH by looking at the problem of intentionality and how such is a fundamental characteristic of the human person. We did by considering the shift Esposito does in relation to AI, relegating its status from intelligence to mere communication. We found that the issue of the CTH was affirmed by Esposito when she argued that AI is not intelligent, but rather, is an effective partner for artificial communication. In relation to this, we claimed that humans, however, are

emergent systems whose intentionality is circumstantially motivated to react to stimuli. This led us to argue that while AI systems and LLMs are complex in the same way humans are, the difference lies in the fact that human intentionality is aware of its own consciousness, while AI systems and LLMs are not. We argued further that the circumstantial nature of intentionality in human subjects gives a dialectical advantage of agency and authorship to humans as partners in communication, while AI systems and LLMs are solely dependent on the instantiation of their programmer, designer, and trainer. In other words, humans have a skin in the game, which gives them agency, liability, and authorship in communications. Algorithms, since they lack intentionality, are reliant on programming and training that need to be fine-tuned to maximize the probability of satisfying the greater number of users' expectations at the cost of generalizing the trained responses of the system. From Esposito, we found that these create problems not only with the marginalization of the 'fit' of an algorithm's communication but also create an illusion of certainty that is often misunderstood by its users. The use of a one-size-fits-all strategy in streamlining algorithm training creates future problems that only human subjects can anticipate and respond to. Also, from her, we provided three aspects that form part of a more precise understanding of intentionality found in human beings but absent in such systems.

The legacy of CTH creates an illusory confidence that an AI system or an LLM is equivalent to, or even better than, humans. Algorithms are like divination tools for contemporary society, and the advent of AI indeed places the human person on the divine pedestal. However, such algorithms might be better understood as tarot cards, which are senseless communications without a human communication partner. Understanding the bias left by CTH allows us to recognize an algorithmic hallucination, pose skepticism to the information it presents, and perhaps use these assistive tools to help us, not replace us. What remains essential is the primal recognition of the lack of intentionality on the part of AI in the hope for our collective realization of how our words are tied to this trait of being human.

NOTES

1. Bacon refers to these as *idola tribus* (idols of the tribe), *idola specus* (idols of the cave), *idola fori* (idols of the marketplace), and *idola theatri* (idols of the theatre) (Zarogin 2001, 379-393). An insightful discussion of Bacon's use of *idols* is in Cooper 2019, 328-350.

2. An example may be provided by taking these two criticisms together. For example, if one looks for a 'university' in the 'central library,' one commits a category mistake of thinking that the 'location' category is applicable to an 'abstract' category. Likewise, in the official doctrine, the mistake occurs when its followers insist that the mind is its *own* category of things. While universities could have their 'locations', their actual effect is based on the various offices, faculties, and facilities working to produce the observable effect of a university. In the same case, the mind and the body are not separate categories. The mind, like the university, is an effect generated by different organs working in concert. Hence, the category of 'thingness' is falsely attributed to what should have been an 'effect.'

3. See Genesis 1:26-27 Thomas Aquinas (2017, Ia.q93.a6) explicitly mentions that "wherefore this image of God is not found even in the rational creature except in the mind; while in the other parts, which the rational creature may happen to possess, we find the likeness of a "trace," as in other creatures to which, in reference to such parts, the rational creature can be likened." Emphasis is added.

4. Several views are of interest here: Aquinas 2017 Ia.q79.a1, Aristotle 1999 ii.3, Heinaman 1990, 83-102 and Robinson 1978, 105-124. On this topic, a good debate arises by contrasting Aristotle's view of the soul and how Descartes treated it (the mind), as discussed in Anstey 2000, 237-260.

5. However, work on neuroscience portrays a development of a typical representation of the mind in terms of the functions of human cognition and personality. This, though, is not to mean an exact replica that is able to capture the very divergences of the human mind (Low 2025 and Laird, Lebiere, and Rosenbloom 2017, 13–26).

6. This is the reason why it is possible to run the same software, such as *Minecraft* or *Cinebench*, on an identical system and achieve identical performance for computer systems that have standardized parts and architectural standards (e.g., x86, ARM), which execute identical software or 'external intention.'

7. A teacher can provide a lecture on the performance of logarithmic computations to a class of forty students and have variations in the time needed to accomplish the task of computing the logarithm of pi down to its 16 decimal places. If we insist on the same analogy with the lecture and software (as with the mind and computer analogy), the variation of the time it takes to solve the problem and even the correctness of the responses will vary greatly. This also explains why education cannot simply be given up to AI due to how, despite the abundance of data available for AI to analyze and sort, there still remains the contexts of the assigned work that cannot be entirely fed to AI to provide the assignment exactly needed (Bonnard 2023).

8. A question was posted on a blog site of whether ChatGPT functions as the operator in the Chinese Room. The answers are insightful to consider. See https://ai.stackexchange.com/questions/39293/is-the-chinese-room-an-explanation-of-how-chatgpt-works.

9. To make sense of what is meant by communication here, we are reminded of Niklas Luhmann's ontological imperative of communication. We think that this encapsulates the epistemic conundrum of CTH: "Human beings cannot communicate; only communication can communicate. Like communication systems, consciousness systems (and brain cells, and so on, on their other side) are operationally closed systems that can maintain no contact with one another" (Luhmann 2012, 57-58). If CTH proposes that we can only know a mind through its communication, we agree that the only evidence of intentionality is made possible through communication.

10. To discuss more on this, emulating a neuron as its digital counterpart is problematic because memory is a gestalt effect in neurons, operating as a gestalt system that validates memory through a collection of features and interrelated networks of memory. Moreover, the state of neurons has been characterized by Hopfield (1982, 255) as a 'chaotic state of wandering,' which creates problems when it is stored in a digitized format for emulation. The complexity of recreating a biological human mind is so great that even its advocates do not see its possibility in

this century (Sandberg 2013, 171). Intentionality, in the case of the human mind, including its biological processes and sentient circumstances, remains an unattended issue for AI.

11. As we have already argued in the previous sections, an AI system is not necessarily 'intelligent' in the sense that it 'intended' to communicate a response, but rather, to use Turing's term, it is 'deceiving' its users by appearing to be intelligent. An AI system utilizes data derived from human thought, including poems, art, texts, and even compositions, which are integrated into its algorithms without attribution to the creators of the data. This source allows AI systems to communicate algorithmically through their training and adaptive programming.

12. Leonid Rogozov was a Russian surgeon stationed in Antarctica who had to perform an auto-appendectomy. (Bermel 2009)

13. This is, of course, dependent on the large language model's design and the style of query by the user. Users can often bypass these restrictions by creating 'fictional' characters. Pretending to be a surgeon reviewing for a surgical procedure may override these restrictions.

14. While CTH shows the inevitable problem of solipsism in the idea of the mind, it does address a further nuance that is implicit between the mind and AI analogy. While it is paradoxical that AI can function as a communication partner without its own intentionality, the human user, however, is aware of their own circumstances, which makes the detection of LLMs or AI systems somewhat intuitive. An LLM can pretend it knows something; however, it cannot experience something that is not derived from a communication or a direct observation. It will not be able to communicate information that was not previously communicated. Moreover, LLMs will generate a response that sounds fluent, coherent, and on theme, but not necessarily correct. This is often referred to as a *hallucination*, which, for the programming of the LLM is its attempt to respond to a query based on the probability of an expected response.

As an illustration of this point, the authors of this paper used Microsoft's Co-Pilot system and Chat-GPT 5, instructing to create a character and pretend to its best effort that it is human. The specific command was that it was a college student who lives in Makati and studies at UP Diliman. As inhabitants of the Metro-Manila area, the immediate response from the authors was to ask how the college student commutes daily to UP Diliman. Although the LLM can "pretend" that it is a human student who has to commute between Makati City and the school in Quezon City, it will never question or even bother the plausibility (and the discomfort) of the choice to commute daily from Makati to UP Diliman.

15. Algorithms, while they are not a recent development, provide responses to queries with the desired outcome through programming or the training of an LLM. As a component of AI systems, algorithms allow a range of possible responses to a query and calculate the expected response of their users based on their programming and training. An algorithm functions by arranging information into a specific order and hierarchy. Like lists for things that vary over time, such as the top ten greatest hits of the early 2000s, the top five conspiracy theories about Bigfoot, or the top three action-adventure Japanese animated series, algorithms generate communication and information by ordering existing information. The exponential growth in the capacity

of computing systems to process algorithms, along with access to *Big Data*, and a training parameter that refines the desired (or projected desired) output of AI systems, has exponentially increased the capacity of communications.

REFERENCES

Agamben, Giorgio. 2017. *What is Philosophy?*. Translated by Lorenzo Chiesa. Stanford University Press

Agamben, Giorgio. 2024. "L'esperienza del linguaggio è un'esperienza politica." *Quodlibet*. February 16. https://www.quodlibet.it/giorgio-agamben-l-u2019esperienza-del-linguaggio-n-u2019esperien.

Anstey, Peter R. 2000. "'De Anima' and Descartes: Making up Aristotle's Mind." *History of Philosophy Quarterly* 17 (3): 237–60. http://www.jstor.org/stable/27744854.

Aristotle. 1999. *De anima*. Translated by David Ross. Oxford University Press.

Austin, J.L. 1972. *How to do things with words*. Oxford University Press.

Barresi, John. 2025. "Types of Intentionality in Humans vs. AI Systems and Robots: Implications for AI and Robotic Personhood." In Philipp Hacker (ed.), *Oxford Intersections: AI in Society*. Oxford University Press. https://doi.org/10.1093/9780198945215.003.0006.

Bermel, Vladislav Rogozov and Neil. 2009. "Autoappendicectomy." *British Medical Journal*, December 19-26: 1420-1422. https://doi.org/10.1136/bmj.b4965.

Bonnard, Jules. 2023. "AI, do my homework! How ChatGPT pitted teachers against tech." *ABSCBN*. January 16. https://www.abs-cbn.com/business/01/16/23/ai-do-my-homework-chatgpt-pits-teachers-against-tech.

Cooper, Amy. 2019. "Francis Bacon's Idols and the Reformed Science." *Studies in Philology* 116 (2): 328-350. https://www.jstor.org/stable/26695098

Crane, Tim. 1998. "Intentionality." In *Routledge Encyclopedia of Philosophy*. http://doi.org/ 10.4324/9780415249126-V019-1.

Descartes, Rene. 1996. *Meditations on First Philosophy*. Translated by John Cottingham. Cambridge University Press.

Devlin, Keith. 1996. "Good-bye Descartes?." *Mathematics Magazine* 69 (5): 344-349. https://doi.org/10.2307/2691278.

Esposito, Elena. 1993 "Der Computer als Medium und Maschine." *Zeitschrift für Soziologie* 22 (5): 338-354. https://www.jstor.org/stable/23848775.

Esposito, Elena. 2022. *Artificial Communication: How Algorithms Produce Social Intelligence.* MIT Press.

Heinaman, Robert. 1990. "Aristotle and the Mind-Body Problem." *Phronesis* 35 (1): 83–102. http://www.jstor.org/stable/4182348

Hopfield, J.J. 1982. "Neural networks and physical systems with emergent collective computational abilities." *Proceedings of the National Academy of Science* 2554-2558.

Joachimiak, Marcin P. et al. 2024. "The Artificial Intelligence Ontology: LLM-Assisted Construction of AI Concept Hierarchies." *Applied Ontology* 19 (4): 408–418. https://doi.org/10.1177/15705838241304103

Johnson, Matthew David. 2023. *Attributing Intentionality to Artificial Intelligence: An Investigation*. *ETD Collection for Fordham University*. AAI30489362. https://research.library.fordham.edu/dissertations/AAI30489362

Laird, John E., Christian Lebiere, and Paul S. Rosenbloom. 2017. "A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics." *AI Magazine* 38 (4): 13-26. https://doi.org/10.1609/aimag.v38i4.2744.

Lisska, Anthony J. 2016. *Aquinas's Theory of Perception: An Analytic Reconstruction.* Oxford University Press.

Low, Jun Jian. 2025. "Standard Model of the Mind: Toward an Integrative Framework." *SSRN*. August 9. http://dx.doi.org/10.2139/ssrn.5385063.

Luhmann, Niklas. 2012. *Theory of Society*. Translated by Rhodes Barett. Stanford University Press.

Mei, Qiaozhu et al. (2023). "A Turing Test: Are AI Chatbots Behaviorally Similar to Humans?." *SSRN*. http://dx.doi.org/10.2139/ssrn.4637354

Moorman, R. H. 1943. "The Influence of Mathematics on the Philosophy of Descartes." *National Mathematics Magazine* 17 (7): 296-307. https://doi.org/10.2307/3029936

Moura, Francisco Tigre. 2023. "Artificial Intelligence, Creativity, and Intentionality: The Need for a Paradigm Shift." *Journal of Creative Behavior* 57 (3): 336-338. https://doi.org/10.1002/jocb.585.

Ngaihlian, Dorothy. 2025. "Machine Learning Algorithms: Simulating Intentionality in Artificial Intelligence" *SSRN*, May 27. http://dx.doi.org/10.2139/ssrn.5271061.

Nietzsche, Friedrich. 1974. *The Gay Science: With a Prelude in Rhymes and an Appendix of Songs*. Translated by Walter Kaufmann. Vintage Books.

Olschki, Leonardo. 1943. "Galileo's Philosophy of Science." *The Philosophical Review* 52 (4)): 349-365. https://doi.org/10.2307/2180669

Redaelli, Roberto. 2025. "Intentionality gap and preter-intentionality in generative artificial intelligence." *AI & Society* 40, 2525–2532. https://doi.org/10.1007/s00146-024-02007-w.

Rees, Graham. 1986. "MATHEMATICS AND FRANCIS BACON'S NATURAL PHILOSOPHY." *Revue Internationale de Philosophie* 40 (159): 399-426. https://www.jstor.org/stable/23946643.

Robertson, Adi. 2024. "Google apologizes for 'missing the mark' after Gemini generated racially diverse Nazis." *The Verge*, February 22. https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical.

Robinson, H. M. 1978. "Mind and Body in Aristotle." *The Classical Quarterly* 28 (1): 105–24. http://www.jstor.org/stable/638714.

Ryle, Gilbert. 2009. *The Concept of Mind*. Routledge.

Salkeld, Brett. 2024. "Technology results in 'playing God' if it denies reality." *Our Sunday Visitor*. January 16. https://www.oursundayvisitor.com/technology-is-only-playing-god-if-it-leads-us-to-deny-reality/.

Sandberg, Peter Eckersley and Anders. 2013. "Is Brain Emulation Dangerous?" *Journal of Artificial Intelligence* 170-194. doi:10.2478/jagi-2013-0011.

Sawczyński, Piotr. 2022. "Listening to the Unsaid: Giorgio Agamben and the Politics

of the Human Voice." *The Polish Journal of Aesthetics* 64 (1): 57-67. https://doi.org/10.19205/64.22.3.
Schroeder, Ralph. 2018. "Big Data: Shaping Knowledge, Shaping Everyday Life." in *Social Theory after the Internet: Media, Technology, and Globalization* UCL Press. https://doi.org/10.2307/j.ctt20krxdr.9.
Schulman, Ari N. 2009. "Why Minds Are Not Like Computers." *The New Atlantis* 23 (1): 46-68. https://www.jstor.org/stable/43152940.
Searle, John R. 1990. "Is the Brain's Mind a Computer Program?." *Scientific American* 262 (1): 25-31. https://www.jstor.org/stable/24996641.
Searle, John R. 1980. "Minds, brains, and programs." *The Behavioral and Brain Sciences* 417-457.
Shagrir, Oron. 2006. "Why We View the Brain as a Computer." *Synthese* 153 (3): 393-416. https://www.jstor.org/stable/27653433.
Sheridan, Daniel P. 2006. "The Catholic Case: The Index of Prohibited Books." *Journal of Hindu-Christian Studies* 19 (1): 22–26, https://digitalcommons.butler.edu/cgi/viewcontent.cgi?article=1363&context=jhcs.
Sloterdijk, Peter. 2020. *After God*. Translated by Ian Alexander Moore. Polity Press.
Suter, Rufus. 1939. "Aristotle and the Scientific Method." *The Scientific Monthly* 49 (5): 468-472. https://www.jstor.org/stable/16892.