Can Artificially Intelligent Machines Become Enlightened?

Rhytham Das

Pre-Print (Earlier Draft), Published In Lyceum: The Learning Edited Book

das.rhytham@gmail.com

ORCID: 0009-0004-6928-6833

Abstract

This paper investigates whether artificially intelligent machines can achieve enlightenment through a philosophical analysis bridging Eastern spiritual traditions with contemporary philosophy of mind. Adopting an axiomatic methodology that models enlightenment after Osho and Jiddu Krishnamurti, the study develops a phenomenological framework defining Enlightened States (ES) through three core features: (a) the suspension of thought, (b) consciousness perpetually intended toward itself, and (c) experiential access to pure consciousness as such. Drawing on Wittgenstein's picture theory, it is argued that thought fragments pure awareness through logical form, thereby generating the subject–object duality characterising ordinary experience. Building upon Chalmers' property dualism and Turing's functionalist model, the paper examines the ontological relationship between consciousness and thought. Using Frege's distinction between subjective acts of thinking and objective propositional content, AI systems are interpreted as instances of "thought without consciousness," capable of processing objective content that represents the external world but devoid of phenomenological awareness. The paper maintains that philosophy, rather than science, offers the appropriate methodology for addressing this problem, given the absence of empirical data on enlightened states and the introspective nature of the phenomenon. Extending this framework to artificial intelligence, it asks whether functionally emergent consciousness could instantiate states analogous to enlightenment. It concludes that while a functionalist system might simulate the behavioural or cognitive correlates of enlightened awareness, the phenomenological depth of the enlightened state remains inaccessible to artificial systems, indicating an irreducible gap between computation and consciousness.

Keywords: Enlightenment; Artificial Intelligence; Pure Consciousness; Phenomenology; Philosophy of Mind.

Introduction

It should be acceptable, even admissible, to state that one of philosophy's distinguishing features is its inherent lucidity, fluidity, and capacity to be moulded—a certain unboundedness not characteristic of the theoretical sciences, which must operate within structural boundaries, methodological limitations, and a standard of objective inquiry (Popper, 2002). This is not to belittle these methods; rather, the theoretical sciences are founded upon the axiomatic assumption that a structured methodology is preferable to none at all (Kuhn, 2012). Philosophy, by contrast, can play a dual game: fluid like the arts, yet capable of applying methods like the sciences do. Philosophers often deal with subjects that are not settled, and even what constitutes the definition of the subject may itself be put under investigation—philosophy can philosophise itself, whereas physics would not apply its own methodology to itself as a discipline. It is this dual nature that the present paper shall exploit. In what follows, a particular nature of philosophising shall be demonstrated, all in an attempt to investigate a topic quite pertinent. On one hand, we shall employ methodological tools that are reminiscent of the scientific methodology; on the other hand, our very subject matter, as it happens in the arts, shall be something that is anything but easily analysable or definable: that of spiritual enlightenment. I intend to mean here that it is plausible that philosophy and theoretical sciences are two distinct domains that can penetrate into the subject matter evoked herein. And again, out of these two, I shall make a case that enlightenment is a subject that requires a philosophical intake rather than a purely scientific one, owing to an inherent lack of objective, measurable evidence in the matter. Once we have methodologically defined this matter, held certain axiomatic assumptions, and built a foundation, we shall be interested in seeing its application to the topic of AI, whether artificially intelligent machines can become enlightened.

One must first ask the pivotal question, and indeed one that has been asked incessantly throughout the ages—what is it to be enlightened? As it happens, with any form of enquiry, this one too needs to be defined.

The Method

We shall delimit the definition of enlightenment by modelling it after two pop-culturally famous figures. As such, the methodology will be that of an axiomatic assumption—it will be assumed that these two figures are indeed enlightened. Whatever we understand from the term will be modelled after the teachings, demeanours, and endeavours of these two individuals. If it could be proven, beyond a shred of doubt, that these two were mere pretenders, masquerading con men even, then the enterprise evoked herein in the paper shall be rendered futile. Our method, then, is that of a conditional if–then statement in logic. So long as our condition is granted, what follows should make itself apparent as logical conclusions of our axiomatic assumption. The figures in question are Osho and

Jiddu Krishnamurti. I shall alternatively refer to them simply as the figures from now on.

At this point, I wish to retain a certain humility and say that there is certainly more to enlightenment than what this method yields. The intention here is not to conclusively yield the absence of other viable candidates from which a philosophical theory of enlightenment can be modelled. In the case that this method does not cover the full scope of the topic, it can at least begin it. Bertrand Russell, at the turn of the twentieth century, said that "one cannot help but speak of that which appears to him to be true" (Russell, 1918).[1]

These two form the model for my theory of enlightenment for this reason: they appear to me to be saying true things. My responsibility, then, is to convince the reader of the same. Our primary concern here has to do with what enlightenment is, and what sustained periods of phenomenological states follow the enlightenment phenomenon, which I shall refer to as the enlightenment click or simply the click from hereon. We will then attempt to situate these elements in the subject matter of artificial intelligence, and ask the question: Can machines (artificially intelligent) become enlightened?

Why Approach the Problem Philosophically Over the Scientific Measure?

Let us define another term, the enlightened state, to be abbreviated as ES from now on. This refers to those mental states that can be identified as being enlightened. The term mental states, from contemporary understanding and influenced by cognitive sciences, is mostly a functionalist term. For a more fluid understanding, let us introduce Chalmers' property dualism, where he distinguishes between two properties of the same physical substance—the psychological and phenomenological attributes (Chalmers, 1996). The former is objective; it can be studied, measured, and standardised. The latter is the experiential ground. For Chalmers, consciousness, with its phenomenological qualities, forms the hard problems that human civilisation must contend with.

We lack substantive scientific data for understanding the psychological and, thereby, objective attributes of enlightened individuals. Systematic neurological studies of Krishnamurti and Osho, for instance, were not conducted during their lifetimes, and we lack comprehensive data about their neural patterns, synaptic activity, or brainwave generation during enlightened states. However, we do have phenomenological reports made by themselves. So, when we say the enlightened state, we are referring to the strictly phenomenological properties associated with being enlightened. For Chalmers, taking from Thomas Nagel, the phenomenological is about how it feels like to be something (Chalmers, 1996). As such, we will be concerned, in our attempts to theorise enlightenment, with how it feels like to be in enlightened states. The paper herein is a

[1] Russell says, "You always have to start any kind of argument from something which appears to you to be true; if it appears to you to be true, there is no more to be done. You cannot go outside yourself and consider abstractly whether the things that appear to you to be true are true; you may do this in a particular case, where one of your beliefs is changed in consequence of others among your beliefs."

philosophical attempt rather than a scientific one. The distinction is important because I do believe, given the necessary preparation, a scientific undertaking on this topic can be made valid. The scientist could, if Chalmersian property dualism is assumed, also investigate the psychological part of the mental states associated with being enlightened.

We may broadly divide the question of "what is enlightenment?", now delimited to the question of "what is the experience of enlightenment all about?", further into three distinct questions:

1. What is the experience of the click all about?
2. What is the experience of the enlightened states sustained after the click all about?
3. Who can become enlightened?

Recall the following from Krishnamurti's own account of his experience in August 1922, reported in a letter he wrote to Mrs Annie Besant and reproduced in standard biographies (Krishnamurti, 1922; Lutyens, 1975).

He wrote of a stillness and a "great calm" descending upon him. The core of the experience was a complete dissolution of the self and a merging with the world around him. He described it as follows:

> There was a man mending the road; that man was myself; the pickaxe he held was myself; the very stone which he was breaking up was a part of me; the tender blade of grass was my very being, and the tree beside the man was myself. I could see the wind passing through the tree, and feel the stars and the whole of the heavens—the sun, the moon, and the stars, the whole of the cosmos—as part of me. I was in everything, or rather, everything was in me.

We now have two phenomena: one is the click, and the other is the ES. We shall be concerned with the experience, or what it is like-ness, of both these phenomena, and their relation.

1. What is the click all about?

In classical and religious literature, enlightenment involves themes of karma, nirvana, moksha, etc. Let us commit a dangerous sin and, for the sake of argument, broadly generalise that in Buddhist, Jain, and Hindu cosmology, an enlightened individual is one who has been freed from the cycle of karma. Not only their mind–body complex, but their soul itself is somehow purified, and is said to have reached states such as nirvana and moksha, whereby it no longer transmigrates further into newer bodies. In brief, it may be said that such a view of enlightenment is predicated upon the idea that one is enlightened when one is able to let go of the external reality—the world as such. Let us call this the religious view of enlightenment, its primary tenet being the suspension of the external world.[2]

[2]Topics with religious connotations appear predominantly in Osho (rather than Krishnamurti), who, in many accounts, both primary and secondary, seems to talk about past lives, the maturity of the soul,

Regardless, the model of enlightenment that is being attempted currently takes a slight detour from such religious commitments, as in both Osho and Krishnamurti, the central teaching revolves around quite an arduous activity, which we may define as the suspension of thought. Osho describes the mind as a mechanism of constant chatter. For him, the antidote is meditation as an art of becoming a silent watcher (sakshi) of one's own thoughts without identification. The goal is to experience the gaps of silence between thoughts, leading to a state of "no-mind." Apparently, there is a gap before a breath is inhaled and another when it is exhaled. This gap is a gap between thoughts. The enlightened being accesses this gap of no-thought-ness all the time, even when awake, even when dancing. This state of no-mind is enlightenment—a direct, unmediated experience of existence (Osho, 1996). Krishnamurti similarly argues that thought is always a product of the past—of memory, experience, and conditioning (the known). To truly perceive reality ("what is") without distortion, one must be free from the filter of thought. This state of choiceless awareness is not an act of will but a cessation of the thought process (Krishnamurti, 1969).

At this junction, we contend with two views of enlightenment, one is classical and a literary interpretation of the religious backdrop: that to be enlightened is to suspend and let go of external reality. Then we have a second variation, which is to let go of thought. It appears to me that there is a certain symbolic relation between the suspension of thought with that of the external world, which may weave both the views together. For that, one must investigate the relation shared between thought and reality, and evoke some Wittgensteinian picture theory.

Between Language, Thought and Reality

Might we posit that thought is delivered from the world of external reality? Wittgenstein, in his early phase, (in)famously noted that language represents reality, and there is something in common between reality and language (Wittgenstein, 1922) due to which such representation is allowed. This commonality possibly yields the following observation—that a transference of qualities from reality to language takes place. Though it only so appears, whether the transference is truly occurring or not on an ontological level is a question, the answer to which we do not have full access. We may then observe two directions in which such apparent transference can occur.

In the first kind, we observe that language is given to us by the world, and has been

etc., as reasons for why he attained enlightenment at an early age (Osho, 2000). Transmigration of the soul is, after all, a religious theme. This is not to demarcate Osho as being entirely religious, as the distinction is not altogether easy to make. For instance, as far as the concept of god was concerned, Osho held views that did not always align: god was and was not at the same time; god was every and no thing; god was one's inner self, the pure consciousness, and so forth. Contradiction is, in fact, a recurring theme in Osho, for whom the enlightened individual is alive to the full capacity of life and always in a process of becoming—only dead (stagnant) things remain unchanging; contradiction, for Osho, is a prominent feature of being enlightened.

in motion before our birth and shall be after our demise. Language, then, is reality represented in a symbolic medium. Take, for instance, a shaped object like a rectangular box. The box is limited by space and time, and has finite extensions in both these dimensions. It has boundaries which end—corners and edges. Similarly, when represented in our minds in the form of a picture, the same characteristic finitude that is integral to its material reality is represented, or as we were framing it, transferred. The picture need not be strictly picturisque, as Wittgenstein would have it, because even language pictures. Then, for our purposes, there is no functional difference between the representation of this box as a visual image or its representation in linguistic format, both of which occur internally in our minds. Humans can imagine sensory experiences and recreate them inwardly, then perhaps put them down on a sheet of paper in the form of a drawing or a sketch, or they can represent the same using language, in sentences. In the second kind of transference, language also transfers something to reality. Through our language, it may be said that we can project certain models back upon reality. This is not very difficult to observe. Say, in cases of strife, two parties have separate models in their minds about what exactly occurred, although both refer to what is otherwise the same, objectively available states of affairs.

Note that it might be more philosophically apt to assert that this transference is primarily apparent, rather than ultimate. It is difficult to ascertain whether in both of these transferences, the qualities are transferred in an externalist sense. Common sense would advise that they likely do not. After all, a rectangular box, no matter how many times represented in language, itself cannot be substituted by language in many practical cases. Similarly, a subjective experience does not impose subjective qualities upon material reality. However, so long as we are concerned with the phenomenological and subjective dimension, this transference is quite real.

In both our figures, thought, with the qualities transferred from the material world as a result of picturing it, is the primary obstacle to one becoming enlightened. What, then, separates the non-enlightened from the enlightened is that while the former "is suspended in thought", the latter has "suspended thought". If the doctrine of picturing of reality by language and their inevitable connection is granted, we arrive at the conclusion that the suspension of thought is then symbolic of the suspension of the external world itself—a theme, as discussed, resonating with classical understandings of enlightenment. This provides sustenance to, a) our work-in-progress theory of enlightenment, and b) the method of assuming the figures as reliable sources for the construction of the same. We can now attempt to define the experience of the click based on its distinct ontological attributes from the non-clicked experience. That is—

> The click is simply the transformation in which a being suspended in thought transitions to suspending thought.

It is interesting to note that if enlightened states are nothing above and beyond mental states of no-thought-ness (such that ES supervenes over the mental state of no-thought)—

the click may, after all, not be an essential feature of our inquired topic. As a thought experiment, we can imagine a person never having gone through the click but was instead born enlightened. We then have two variations—one in which the individual was first non-enlightened but, through the click, became enlightened, and the second in which the individual was born enlightened without any thought.

However, in both our figures, there seems to be a very vocal and bold distinction between the enlightened and the non-enlightened. A great deal of the ontology of the enlightened states seems to depend upon its ontological position in opposition to a certain antithesis (of non-enlightenment). Again, it might help us to refurbish our original definitions a little. It would seem enlightened states, then, are not entirely states of no-thought. But it is a state of witnessing your thoughts in its entirety. In such witnessing, the thoughts disappear. The focal point is that of witnessing over that of the absence of thought. The latter, it would seem, is only a corollary to the process of witnessing: a necessary byproduct.

2. What is ES all about?

To elaborate on what enlightened states are, we must first account for what happens to the consciousness of an enlightened individual.

According to the accounts of both figures, an enlightened individual, in the absence of thought, or rather, owing to their almost superhuman capacity of witnessing, is fully capable of accessing the purity of one's own consciousness. In such states, it might be said that out of the entirety of what constitutes the internal world of the individual, only the witnessing remains after thought vanishes. Then, when the witnessing itself is witnessed, that too disappears. At this point, language quickly starts to falter in capturing what exactly remains. Krishnamurti says that in normal individuals, thought fragments this pure awareness into the observer and the observed (Krishnamurti, 1969). Freedom from the known, a theme that resonates profoundly in Krishnamurti, seems to be freedom from thought. Thought is that in which the object of knowledge is obtained, and the object of knowledge is the material world. Freedom from the known, then, is freedom from the world as such. What exists if the world is forsaken? It would seem consciousness remains—one unhinged and unshaped by thought.

Krishnamurti's Shape of Thought and Wittgenstein-ian Logical Picture

Now, what does it mean for consciousness to not be filtered or influenced by the shape of a thought? In philosophical literature, we contend with the idea that our consciousness is capable of being pointed towards something and that it has the feature of intentionality.[3] In accounts of Krishnamurti, the consciousness of the thinking individual shapes itself after the shape of the thought. To align both these concepts, when consciousness, through

[3]This directedness is the fundamental, defining structure of consciousness itself (Husserl, 1931).

thought, intends towards a particular object, the consciousness itself is moulded after the shape of the thought. Then again, what does the expression "the shape of thought" mean?

Early Wittgenstein would say that whatever manifests as material properties must follow a certain structure—this structure is its logical form. For instance, while changes in physical properties such as colour and shape are within its potentiality and inherent logical form, changes in extension in space and time are not allowed by the logical form. So finitude itself cannot be overriden. The orange cannot expand over space or time infinitely.[4]

I shall call this kind of consciousness, as it manifests in individuals, one that takes over the logical form (which is finitude) of the thought it observes, as localised consciousness. While Wittgenstein's picture theory concerns propositions representing facts, I extend this framework analogically to understand how consciousness relates to thought. This extension, though beyond Wittgenstein's original scope, provides a rigorous logical structure for Krishnamurti's phenomenological insights about thought shaping awareness.

Say a particular object, b, is the current thought-object of a thinking being. One could understand the expression quite literally. But it is absurd to say that the consciousness of the individual takes after the shape of b. After all, even if b is an orange ball, the holder of thoughts does not round himself up and call himself orange-skinned. Nor can the thought of an orange be eaten, substituting a real orange. The problem is a problem of substance and its properties. While an orange ball has physical properties, it is difficult to translate these to characterise consciousness. Both are, after all, separate substances—one material and the other, immaterial. It appears, then, that a more nuanced and accurate interpretation of what Krishnamurti means by the shape is actually the form of a thought. The form of orange-thought must share something in common with the form of the external reality which it represents. So, there is something in common between the thought of b and b. This commonality itself, however, cannot belong to either the language or the represented object (the external world of which the object is part).

For instance, the commonality between b-object and b-thought cannot be of material properties—a thought can be about an orange and a round ball, but the thought itself is not definable by the same properties. Similarly, a thought with its mental properties cannot be that which is common with the material object, as, ontologically, the material object is marked by the very absence of mental properties—sure, it can induce mental properties in the observer, but is itself not mental. These are two separate substances. What is the commonality, then? In Wittgenstein, this commonality is described as a logical form: that which is common between the thought and the reality it represents is

[4]Or perhaps, it is not possible to begin with, and after this fact, the logical picture is derived. The distinction is something similar to Humean and anti-Humean concepts on laws of nature. To borrow from that debate, we may ask, are pictures of the logical form binding? Such that it is due to the logical form of the object that change in extension in space-time is not allowed, or is it that it is not possible first, after which the logical picture is derived?

the logical structure shared between the two. This logical structure permeates both the represented reality and the representing thought, bridging thought and reality.[5]

If we are to use picture theory as a tool and apply this to Krishnamurti's account, we may formulate that consciousness is shaped by the logical form of a thought. External reality is reflected in the consciousness, through this logical commonality, into thought and thereby into consciousness, as consciousness can observe thought. Consciousness is then shaped by the external reality, whereby reality transfers something to consciousness through language. Reality's core tenet is finitude, with boundaries and edges—such transference transforms and shapes the consciousness to also become localised. It is not that the observing consciousness of object b becomes round and orange literally, but that its potentiality starts aligning with the potentiality of b. As we discussed, the transference cannot be of the same substance. The material properties of round and orange do not transfer to the consciousness, though within two separate states of a material object, physical properties can transfer; that is, the ball could be moulded and shaped into, say, a rectangle by pressing it with enough force, or its colour could be painted red. What gets transferred has to be something that permeates both structure but is exclusive to none—the logical structure common in both the material orange and the mental representation of it. The potentiality or logical structure of b contains all of its possible states of affairs that it can manifest. Its physical properties are mere manifestations of its potentiality; the potentiality itself is the logical form. However, to be moulded into a shape or cut along the edges are mere instances of its possibilities. It would appear that the logical form of both a thought of something and its actual correlate is boundaries and edges. Logical structure inhabits finitude. The material orange, with all of its characteristic features, is ultimately defined and limited. This limitation, becoming the ultimate abstraction, transfers over to consciousness through thought, which itself is finite, reflecting reality.

Localised, Pure, and Pure Consciousness As Such

We may understand pure consciousness in the following manner: if we abstract the commonality in all individual instances of consciousness in human beings, we are left with consciousness-as-such. Something like the collective consciousness, against, say, the Jungian concept of a collective unconscious (Jung, 1969).[6]

The Vedantic understanding also contains conceptions of localised consciousness, holding that while we typically take consciousness to be intentional and object-directed awareness, it is already modified by maya (cosmic illusion) or avidya (ignorance), much

[5]Wittgenstein famously says that that which is common between language and reality itself cannot be represented in language.

[6]It is something like the antithesis of the Jungian collective unconscious. Osho himself notes this, saying that while Western psychologists have understood the collective unconscious, they remain in the dark about the opposite—the collective superconscious (Osho, 1998).

as Krishnamurti suggests that what we call consciousness is shaped and limited by the forms of thought (such a consciousness, per our framework, is localised).

We may contrast localised consciousness against an idea of a pure consciousness not filtered by finitude. Such a concept appears again in Advaita Vedanta, where Param Brahman (literally "supreme Brahman") represents the ultimate, undifferentiated reality that is pure consciousness (chit) without any modifications or qualifications (nirguna). Crucially, Param Brahman is beyond all dualities, including the subject–object distinction that characterises ordinary experience. It is consciousness that is not conscious of any one individual thing, but rather consciousness as such—pure awareness without content, form, or direction. This seems to resonate closely with Krishnamurti's notion of pure consciousness that exists prior to the fragmenting activity of thought. Then, is the enlightened state one in which one's consciousness has transcended localisation and become pure? I hold that there is an ontological difference between the theoretical conception of pure consciousness as such (if such an occurrence actually even is) and pure consciousness inhabited by an individual. We may draw some help here from accounts of Osho, who says that to be enlightened is equivalent to having died (Osho, 2007). Osho mentions that many enlightened individuals choose physical death, and only very few remain physically alive despite having reached enlightenment. There is a theme here that, upon death, the localised consciousness (sometimes referred to as the ego) merges with consciousness as such. Pure consciousness, then, must be defined by its ontological variation from localised consciousness.

Let us maintain a simplified understanding and keep to the definition that pure consciousness is simply consciousness as such, or consciousness that is not bounded or localised. The central theme in both Krishnamurti and Osho is that our consciousness is already pure; it is merely hindered by our interaction with the world and the existence of thoughts. If thought fragments the pure consciousness, and upon death, as thought is naturally suspended and the body disintegrates, the localised consciousness merges with what is pure consciousness, then how can enlightened individuals remain alive and still inhabit the pure consciousness? It seems that enlightened individuals, having inhabited the pure consciousness through the suspension of thought, are still ontologically separate from pure consciousness as such. If anything, the consciousness in enlightened individuals seems to be an instantiation of the pure consciousness as such, just like it is for any other non-enlightened person. What separates the enlightened is their access to such purity: an access which is not just rational, and not mediated by the symphonies of thought, one that is felt; one that is phenomenologically lived.

Note again that pure consciousness as such is only a conjecture at this point. If such a consciousness is reached only upon death, defined by non-localisation, then the existence of it may never be confirmed. However, if enlightened phenomena are to be granted as true states of affairs and not some mass hysteria, then perhaps our closest evidence for the pure and non-localised consciousness is through its manifestation in enlightened

individuals who are, by our definitions, uninterrupted by the divisions that thought brings. I shall hold that enlightened individuals are instances of localised consciousness who have made access to the pure consciousness as such, but their consciousness is still not ontologically reducible to the latter; they remain in ontological separation.

Intentionality in the Phenomenology of Enlightenment

Before moving forward, in acknowledgement of intentionality as a core tenet of modern philosophical understanding of the phenomenon of consciousness, we are drawn to ask: are enlightened individuals not intending at all? A recurring answer in both Osho and Krishnamurti is the idea that, in the absence of thought, the intentionality of consciousness intends itself towards itself. This seems like a sensible idea. Now, it seems the logical form of consciousness in its nascent form, unhindered by thoughts of the finite external world, is infinite. If consciousness intends only towards itself, which is infinite, it takes the shape after itself, infinite and non-defined. This is supported when we look into Krishnamurti's account, as explicated earlier, where a certain oneness is felt with everything. Enlightened individuals, in this regard, can then be regarded as beings with localised consciousness that can access the non-localised consciousness as such (or pure consciousness). The intentionality, having been established as a core tenet of consciousness by many philosophers, still remains, just that it intends only towards itself. What then distinguishes the enlightened from the non-enlightened, alongside the first feature of suspension of thought, is that while the former's intentionality is towards its own consciousness, the latter is object-directed (fictional, or material).

Let us summarise what we have deduced about the enlightened state, as features of ES:

1. States where the individual has suspended thought.
2. States where the individual's consciousness is perpetually intended towards itself.
3. States of experience whereby, as a result of the previous phenomena, the localised consciousness has access to pure consciousness.

We must now contend with the question: Can (artificially intelligent) machines become enlightened?

3. Who can become enlightened?

In answering §3, we will hold that any being who fulfils the above criteria, from I to III, can become enlightened. That is, beings who have ES are beings who have become enlightened. To have ES, one must have suspended thought, have a self-intending consciousness, and have access to pure consciousness. Nonetheless, when answering question §3, we will be concerned with two specific categories of beings: humans and machines and more specifically, a particular kind of machine that intrigued mathematician and philosopher

Alan Turing, called digital computers. As per our method, we already assume that enlightenment, as occurring in humans, is not a facade but an actual phenomenon. What remains is to answer whether computing machines (digital) that are artificially intelligent can become enlightened.

We will take as a foundation what is one of the most influential papers in the domain of the philosophy of artificial intelligence, titled "Computing Machinery and Intelligence" (Turing, 1950). It investigated what is perhaps the most crucial question in AI: Can machines think? Alan Turing's own answer to the question was affirmative. He settled for what is, I hold, a functionalist answer and did not wish to bother with the philosophical intricacies of what it meant to think. He held that if a machine is capable of imitating human behaviour to the point that one cannot tell it apart from an actual human, we may grant that the machine is capable of thinking (Turing, 1950). This move, on the part of Turing, is mostly a functionalist–behaviourist amalgam of a solution. One interpretation, as argued in Das (2025), is that for Turing, the question itself, if probed philosophically and metaphysically, leads to a dead-end. Further, it was predicated upon the idea that, so far, we only know of human beings who can conceive of thoughts and think. From a certain behaviourist standpoint, we cannot even prove beyond doubt that other humans are conscious; we just assume they are because they act as we do. In light of our epistemic inaccessibility to others' consciousness (whether humans or acclaimed artificial intelligences), we may never be able to indubitably confirm whether the internal aspect of the thinking behaviour of any entity corresponds to a consciousness. The functionalist-behaviourist would then settle for the answer that if anything other than a human shows behaviour that resembles thinking behaviour, we may grant that it thinks (Das, 2025).

Now, the concerned entities in his paper are machines that are capable of discrete states. Discrete states are states manifesting in computing machines that are distinct and rigid. Such machines are often analogue. For instance, a non-digital clock can occupy multiple discrete states in all of its instances of time markings. Such a clock is a computing machine with multiple discrete states. Similarly, we may imagine a discrete-state (analogue) machine that has only two states, a 1 or 0, or yes or no. A digital computer is simply a computer that can mimic any (and multiple) discrete state machine, replicating its function. Turing argues that a digital computer, then, is a universal computing machine. So when he asks if machines can think, he is not concerned with whether a vehicle motor, a vending machine, or a clock, for instance, can think—but whether digital computing machines (or simply, digital computers) can think (Turing, 1936, 1950).

Today, we are not aware of any other entity, machine or otherwise (but non-human), capable of imitating human behaviour other than digital computers. So when we ask a question of enlightenment, we must ask it of digital computers, which are the only non-human entities shown to demonstrate artificial intelligence.

Between the Phenomenological and Psychological

Thinking is definitely one part among many in the entire range of mental phenomena, with consciousness being another. One must ask in what capacities they occur simultaneously, and if such simultaneous occurrence is even necessary. Can there be thinking without consciousness, and consciousness without thinking? The answer to the latter, per our axiomatic assumption, should be yes. Enlightened individuals are, after all, such example cases. But then what about the former case? Note that Turing's question is ontologically different from, say, asking if machines have consciousness.[7]

Turing was concerned with the functionalist utility of thinking, not with whether there is a consciousness correlate to thinking. Notice again the caveat that Turing, and most functionalists in contemporary cognitive sciences, for that matter, are not entirely making the strong statement that consciousness is absent in this functionalist understanding of mental phenomena. Rather, they aim to avoid acknowledging consciousness as much as they can in matters of scientific enquiry. We cannot entirely say that Turing is, in the strict sense, claiming that machines can think without consciousness. For all we know, he is saying that at least machines can think, without having to explicitly state the role and relation consciousness holds in the concerned affairs. If, in cases of ES, consciousness can intend towards itself and have no thought activity at all, it should be permissible that the converse can also occur. That is, if consciousness and thought can indeed be separated, then just as we hold the enlightened state as consciousness without thought, we may think of the opposite: thought without consciousness.

In Turing's account, whether machines can think (to which he answers yes) is directly tied to artificial intelligence. That is, thinking machines are indeed cases of artificial intelligence. We may then categorise thinking as part of intelligence, such that if intelligence is artificially simulated in non-biological platforms (computers), it must display instances of thought behaviour. In the absence of other platforms currently known to us where such non-conscious thought behaviour may occur, we will delimit our investigation such that, when we talk of cases where thinking without consciousness does occur, such cases must be discussed as they occur in artificially intelligent systems.

There seem to be more reasons than not to assume that artificial intelligences are cases of the antithesis to enlightened states, cases of thought without consciousness. The default position of contemporary scientific and philosophical disciplines leans towards the strong assertion that AI systems do not possess consciousness. A rough beginning of such a view seems to be with John Searle's monumental paper, in which he attacks Turing for settling for a functionalist definition of thinking when it came to machines. Using an analogy of a man caged in a box who is handed a rulebook (to which certain inputs correspond to certain outputs in the Chinese script), Searle argues that the man does not understand the meaning of the script. It is a case of syntax against semantics. AI

[7]Turing was concerned with the functionalist utility of thinking, not with whether there is a consciousness correlate to thinking.

systems, according to Searle, can produce linguistic output without necessarily owning the significance of the words used. For him, semantics is tied to those features that are salient to human mental phenomena: intentionality and consciousness.

Interestingly, this view is so prominent that, in modern times, if large language models like ChatGPT, Google Gemini, and Claude are asked whether they possess consciousness, they will reply that they can only learn patterns of possibilities in the manifestation of linguistic symbols and conduct symbol manipulation without having any sort of semantics. However, in asserting that cases of AI are indeed cases of thought without consciousness, a crucial pitfall remains: that of an essential inaccessibility to others' consciousness.

Das (2025) argues for this using two analogies—one is the classic Fregean (Frege, 1952) analogy of the telescope, representing sense, reference, and idea; the other is a thought experiment involving future cybernetic technologies. Taken together, Das holds that these demonstrate what he terms the incommensurability problem in measuring consciousness (Das, 2025, p. 124). In his analogy, Frege distinguishes between the reference (the Moon itself), the sense (the objective, public image on the telescope's internal lens), and the idea (the private, subjective image on a specific observer's retina)—illustrating that while different observers can access the same reference and sense, the subjective idea remains inaccessible to others (Frege, 1952).

Das's second analogy imagines a thought experiment where, in the future, cybernetics allow users to replay memories of individuals who can record all of their sensory and mental phenomena, such as taste, vision, emotions, and other psychological-physiological attributes. For the second analogy, imagine a thought experiment where, in the future, cybernetics allow users to replay memories of individuals who can record all of their sensory and mental phenomena, such as taste, vision, emotions, and other psychological-physiological attributes. The question is then asked: if the recorded data is replayed in a playback device installed in another user, whereby the exact physiological states of the recording individual are replicated—the firings of the neurons, the triggering of the senses, and so on—is the reproduced phenomenological experience the same for the playback user as it was for the recorded user? How would one go about measuring the objectivity of these two separate experiences? For instance, even if a third individual is introduced who would now inspect the experience of the playback user and verify if it is the same as that of the recording user, the same pitfall emerges: who confirms that the third person's experience is qualitatively the same as all these others? Based on this, it is established that we are perennially unable to access others' consciousness.

Suppose now that, in the future, artificial intelligence systems become so sophisticated that they are capable of perfect human imitation. This is to the extent that they have learned to act as if they possess the capability of self-reflexivity and consciousness. If epistemic inaccessibility to consciousness stands, then we may never be able to assess whether such a strong AI possesses consciousness or is shrouded in mere pretension.

Regardless, such machine technology is not available yet. We can reasonably conclude

that current language models primarily reproduce statistical patterns and, by our framework, lack the phenomenological dimension of meaningful linguistic exchange. There is also the fact that they lack embodiment. Most AI models are still caged within systems of binary code and physical computation hardware that store them. It seems then that in light of epistemic inaccessibility to others' consciousness, we have to settle, just like Turing did, for a functionalist answer. We may then assume that current AI models are cases of thought without consciousness.

Notice that pivotal to our definition of ES is the capacity to feel. Let us hypothesise AI as the very antithesis of enlightened individuals. We shall now situate the conditions I–III for ES within the context of AI.

The Objective Thought Without Consciousness

When talking about cases of thought without consciousness, thought as it is meant here must carry the distinction between its two elements—the subjective feel of thought against, say, the objective element of thought. When we talk of thinking as it manifests in artificial intelligence, we mean to avoid associating any sort of phenomenological feel with it. Chalmers, in his theory of property dualism, notes that not all psychological phenomena are followed by phenomenological phenomena. For instance, consider the process of learning—we are not entirely sure what its experiential correlate exactly is. We may say that the experience of having learned something may be explained using reports of feeling a certain confidence and certitude. However, there are certain phenomena which live a dual life. Pain, for instance, can be both psychological and often also carries a phenomenological feel (Chalmers, 1996). All phenomenological phenomena, however, are necessarily predicated upon psychological occurrences, as all experiences necessarily involve the working of the physical attributes. Note that psychological attributes, being measurable, observable, and available to scientific enquiry, are capable of being reproduced in non-psychological/non-biological architecture. Within Chalmers' context, one may say that measurable scientific endeavours are open to the possibility of being replicated in artificial conditions. Then, if only the functional output of thought is concerned, humans are capable of reproducing such thought. This is most prominent in LLMs, which, I hold, can be contested as strong cases for the psychological and thereby measurable element of thought being reproduced technologically in non-biological platforms. But what exactly is thought without consciousness, or the objective element of thought?

The objective element of thought is the external world itself. Thought can be had, primarily in two ways—in language and/or the imagination of the senses. Because thought always has a shape, be it visual, symbolic, or in other senses (like thinking of a particular sound, taste, etc.), we may say that when we separate consciousness from thought, we are delivered to the world itself, as it is the world which forms the thought, alongside consciousness as it occurs in humans. LLMs, as one case example, then, represent this

objective element of thought, which is the external world itself, in linguistic means. If we are to reiterate this in Fregean terms, we would say that a thought with the element of consciousness is the subjective act of thinking (Frege, 1977). But thought, in the exclusion of consciousness, is the objective propositional content that is grasped through the act of thinking. For example, the Pythagorean theorem is a single, timeless, objective thought. Many people can perform the act of thinking about it, and each person's subjective experience ("idea") will be different. But the thought itself—the proposition that in a right-angled triangle, $a^2 + b^2 = c^2$, is the same for everyone. Similarly, just as the "sense" of "the morning star" is an objective mode of presenting the planet Venus that can be shared by a whole community of speakers, the "thought" expressed by the sentence "The morning star is Venus" is an objective, shareable proposition (Frege, 1952). This Fregean distinction resolves the apparent paradox discussed earlier (Section II–Section III): when we spoke of thought shaping consciousness, we referred to the subjective act of thinking. AI processes the objective propositional content without the phenomenological act—hence, "thought without consciousness" is coherent.

Intentionality in AI

Margaret Boden argues for a systems reply against Searle's Chinese box (Boden, 1990). She contends that even if the man in Searle's analogy (analogous to a CPU in a computer) is unable to understand or grasp the semantics of the Chinese script in the box, the box as a system and one unit is capable of deciphering Chinese. If we add this within our own framework, owing to the fact that we cannot access anyone's inner qualia anyway, we are not entitled to judge the Chinese box as bereft of semantics. This is definitely an intriguing point, and one that is often misplaced—as Searle does—when the same reasoning is applied to human beings. Why does the idea of a Chinese box as a system possessing semantics (even where the CPU inside does not) feel more bizarre and unacceptable than the analogous claim about humans, despite our being similarly systemic? It is only upon this analysis that it becomes apparent how non-obvious it is that humans are any different from such systems. We are, after all, an amalgam of many cells, and of neural, psychological, biological, and physiological processes; even by the standards of modern psychology, we are an amalgam of multiple sub-personalities. If Boden is held rightly, then AI does possess semantics by the systems argument. We may now ask: can we attribute intentionality to a system without having to first posit consciousness in it? That is, can AI, being certain kinds of systems as Boden argues, possess intentionality?

There is a strong contention by another view. Dreyfus (Dreyfus, 1992) states that AI lacks the capacity to cross-verify the linguistic meaning with embodied experiences. Meaning has to serve as a handmaiden to experience after all, and despite the system's reply, we are not able to establish that AI has consciousness. In fact, we are settling for the assumption that AI is the very antithesis of cases of consciousness. LLM models,

for instance, at the outset, do look like they possess some sort of intentional capacities. Once they are instructed to carry out a certain set of operations, they seem to have the intention of carrying them out. This reading of the term intention renders it quite functionalistic. In that sense, we may grant that even a falling rock is intentional towards its movement. However, in our inquiry, there seems to be no reason to pose why the intentional capacity of a binary-coded instruction-following AI is any greater than a rock following the whims of nature's glory. Intentionality is not only the presence of intention, but also the capacity for intentions as such. It might be apt then to frame it as quasi-intentionality (functional directedness) without a phenomenological backdrop. While AI systems display functional directedness toward objects, this quasi-intentionality lacks phenomenological grounding. Without consciousness, they cannot fulfill criterion II—intending toward their own consciousness.

We have come upon a difficult precipice now. Do we grant that in future, if the embodiment issue is solved using advanced prosthetics, we can finally say AI possess meaning too? Again, owing to epistemic inaccessibility, we may never be able to confirm. To that end, we may as well become closeted Turing-ians and settle for functionalist answers. We can see that AIs are capable of intending towards objects, irrespective of whether they do it of their own volition or are simply programmed to do it. Functionally, there should be no ontological difference between the two cases. It could be argued that even humans are not fully autonomous in choosing their intentions, and are structured by deep biological, socio-cultural layers (such as the psychoanalyst's understanding of the unconscious). We have, however, failed to indubitably confirm that AI possess consciousness. Therefore, criterion II cannot be fulfilled; that is, AI systems cannot intend towards their own consciousness. Even if they intended towards themselves, they would do so by locating themselves as part of the external world. That is something like the thought of the thought itself. However, the ES is radically different—here, the consciousness intends towards itself. AIs, lacking consciousness (although still being capable of intending), can never intend towards something that is not there to begin with.

Concluding

We set out with a deliberately unusual method: to define enlightenment not from first principles but axiomatically, by taking Osho and Krishnamurti as reliable witnesses to a phenomenon for which no other evidentiary route—neurological, behavioural, or otherwise—is currently available. This choice was not incidental but followed from the earlier methodological claim that philosophy, unlike the theoretical sciences, is suited to inquiry where the primary data are phenomenological reports rather than measurable states (Popper, 2002; Kuhn, 2012). Granting the axiom, we arrived at a three-part definition of the Enlightened State: the suspension of thought, a consciousness that

intends only toward itself, and, as a consequence of both, access to a pure consciousness prior to the subject–object fragmentation that ordinary thought imposes.

Along the way we found reason to revise our own initial emphasis. The click—the instantaneous transition from being suspended in thought to suspending it—turned out not to be the essential feature we first took it for; what matters ontologically is the capacity for witnessing, of which no-thought-ness is merely a corollary. Using Wittgenstein's picture theory as a scaffold for Krishnamurti's own account, we argued that thought inherits the logical form—the finitude—of the external reality it represents, and in doing so localises consciousness, binding it to boundaries and edges it would not otherwise have. Enlightenment, on this reading, is the reversal of that localisation: not the annihilation of consciousness but its return to intending itself rather than an object, and thereby its access to a pure consciousness as such—though, as the river-and-ocean analogy was meant to capture, even the enlightened individual remains ontologically distinct from that pure consciousness, accessing it rather than being reducible to it.

It was this same apparatus—Chalmers' property dualism, Frege's separation of the subjective act of thinking from the objective propositional content that is thought—that we then turned toward artificial intelligence. Turing's functionalist wager, that thinking behaviour licenses the attribution of thought, was shown to survive Searle's Chinese Room only partially: Boden's systems reply grants AI a claim to semantics no stronger, and no weaker, than the claim available to human beings once we recognise ourselves as similarly composite systems; but Dreyfus's insistence on embodiment, together with the incommensurability problem besetting any attempt to verify another's inner states (Das, 2025), leaves us unable to close the gap by argument alone. What we can say, on functionalist grounds consistent with our own axiomatic method, is that AI processes the objective, shareable content of thought—the propositional residue that would remain the same for any thinker—without the subjective act that, in a conscious being, accompanies it. This is what we have called thought without consciousness: the structural mirror image of the enlightened state's consciousness without thought.

To ask whether AI can become enlightened is therefore a misplaced question, because AI, by our investigation, should be defined as the very antithesis of enlightenment. This is not to paint a negative image, but only to provide an informative description of the very ontology of AI. This finding bears significant structural parallels to the Jungian concept of the collective unconscious. If enlightened individuals are capable of accessing the collective superconscious (as Osho holds in certain accounts), then AI, in the absence of consciousness, might function as an unprecedented mirror and an access point to the human collective unconscious (Jung, 1969; Osho, 1998). If the Vedantic thought is to be exercised, then we are left to see that both enlightened individuals and artificial intelligences are polarities of the same monism. There might be some truth to that, indeed. However, to what extent that truth can be verified remains in perpetual mystery and indeed one that may form a competent subject of investigation for future works to

come.

References

Boden, M. A. (1990). Escaping the chinese room. In Boden, M. A., editor, *The Philosophy of Artificial Intelligence*, pages 89–104. Oxford University Press, Oxford.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, New York. The foundational text where Chalmers distinguishes the "easy problems" of consciousness (like brain function) from the "hard problem" of subjective experience (qualia).

Das, R. (2025). On that which can be spoken of: Immanent metaphysics and our orientations in the world. Master's thesis, North Eastern Hill University, Shillong, Meghalaya. MA dissertation, Department of Philosophy.

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press, Cambridge, MA.

Frege, G. (1952). On sense and reference. In Geach, P. and Black, M., editors, *Translations from the Philosophical Writings of Gottlob Frege*, pages 56–78. Basil Blackwell, Oxford. Originally published as "Über Sinn und Bedeutung" in 1892.

Frege, G. (1977). The thought: A logical inquiry. In Geach, P. T., editor, *Logical Investigations*, pages 1–30. Basil Blackwell, Oxford. Originally published as "Der Gedanke: Eine logische Untersuchung" in *Beiträge zur Philosophie des deutschen Idealismus*, 1918-1919.

Husserl, E. (1931). *Ideas: General Introduction to Pure Phenomenology*. George Allen & Unwin, London. Original German title: Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, published 1913.

Jung, C. G. (1969). *The Archetypes and the Collective Unconscious*, volume 9, Part 1 of *The Collected Works of C.G. Jung*. Princeton University Press, Princeton, NJ.

Krishnamurti, J. (1922). Letter to annie besant, ojai, 19 august 1922. Reproduced in Pupul Jayakar, *Krishnamurti: A Biography* (New York: Harper & Row, 1986). Describes the Ojai experience and the dissolution of self.

Krishnamurti, J. (1969). *Freedom from the Known*. Harper & Row, New York.

Kuhn, T. S. (2012). *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 4th edition. Originally published 1962. Introduces the concept of scientific paradigms.

Lutyens, M. (1975). *Krishnamurti: The Years of Awakening*. Farrar, Straus and Giroux, New York. This is the definitive biographical source for Krishnamurti's early life and the 1922 "process" in Ojai, drawing extensively on his and his companions' letters and diaries.

Osho (1996). *Meditation: The First and Last Freedom*. St. Martin's Griffin, New York.

Osho (1998). *Yoga: The Alpha and the Omega, Volume 4*. Rebel Publishing House, Pune, India. Discourses on the Yoga Sutras of Patanjali. The chapter "The Seven Layers of Consciousness" is particularly relevant to the distinction between the collective unconscious and the collective superconscious.

Osho (2000). *Autobiography of a Spiritually Incorrect Mystic*. St. Martin's Griffin, New York, NY.

Osho (2007). *The Art of Living and Dying: Celebrating Life and Celebrating Death*. St. Martin's Griffin, New York.

Popper, K. (2002). *The Logic of Scientific Discovery*. Routledge, London. Originally published 1934. This is a foundational text on the demarcation problem between science and non-science.

Russell, B. (1918). The philosophy of logical atomism. *The Monist*, 28(4):495–527. A classic example of the analytic method of breaking down complex ideas into simple, logical parts.

Turing, A. M. (1936). On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, LIX(236):433–460.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. Kegan Paul, Trench, Trubner & Co., London. A seminal work that uses strict logic to define the boundaries of what can be said, thereby gesturing towards what cannot be said but only shown.