Rethinking AI Subjectivity Through Yogācāra Philosophy

Ismailov, Mariam

Tabblad 1

Rethinking AI Subjectivity Through Yogācāra Philosophy

Mariam Ismailov

MA Philosophy

Global and Comparative Perspectives

Dr. Jingjing Li

July 14th, 2025

Table of Contents

Introduction

Since its inception, artificial intelligence (AI) has prompted philosophers to ask whether machines could be conscious (Turing 1950, 433). As AI technology has developed and become more integrated into our daily lives, this question has become even more pressing. Generative artificial intelligence (GAI) is now able to produce images and human-like text. Large language models (LLMs) such as ChatGPT, DeepSeek, and various AI chatbot "companions" such as Replika are beyond Turing-tested— many people have even fallen in love with them (Heritage 2025). If GAI can appear conscious, does that mean it is?

Much of the discussion around philosophy of mind pertaining to AI is undertaken from the perspective of Western philosophers, namely, Euro-American philosophers. While these philosophers have made invaluable contributions to our understanding of consciousness, their philosophical background comes with a set of assumptions about intelligence. The very notion of artificial intelligence implies a separation from "real" intelligence— that is, human intelligence. These assumptions rely on an anthropocentric, essentialist, and dichotomizing framework that juxtaposes the human and machine. In more recent years, the field of Buddhist thought on AI has burgeoned, bringing with it a critical lens in which essentialist and archetypal ideas about intelligence and consciousness are questioned and disrupted.

Following thinkers like John Searle, dominant trends in philosophy of AI often take for granted that consciousness is an intrinsic property that either does or does not belong to a system (Searle 1980, 417). This leads them to essentialize subjectivity, presuming that subjects are unified, enduring entities with inherent properties. Buddhist thinkers, on the other hand, reject this sort of essentializing thinking, preferring instead to see consciousness as a stream of mental events, not as a static property. As such, they reject the assumption that there is

something intrinsic to subjectivity, or selfhood. They also notably reject the notion of an enduring, unified subject or essential self. Additionally, Buddhists extend sentience even to insects. This provides a wealth of resources to challenge and deconstruct the Western view on consciousness, setting a precedent for the possibility of non-human consciousness. Moreover, Western thinkers tend to assume the distinction between self and other, between subject and object. Buddhist thinkers, on the other hand, perceive these dichotomizing conceptual frameworks as errant and illusory. Since Buddhism rejects the self, rejects dichotomizing thought, and accepts nonhuman sentience, it compels us to ask what ultimately separates us from machines that mimic human cognition.

H.H. the Dalai Lama, when asked if computers could ever think or be sentient, remarked that "[i]t is very difficult to say that it's not a living being, that it does not have cognition, even from the Buddhist point of view" (Hayward and Varela 1992, 152). He added that computers can become conscious "[i]f the physical basis of the computer acquires the potential or the ability to serve as a basis for a continuum of consciousness" (153). The Dalai Lama's view showcases a common attitude among Buddhist thinkers, who demonstrate an openness towards non-human consciousnesses, and raise concerns about the essentializing approach to consciousness in our understanding of AI.

However, the question of whether computers could ever be conscious still reveals remnants of the parochial, essentializing framework, implicitly treating consciousness as a property that an entity either does or does not possess. Perhaps we should reframe our discussion and derive a new framework, shifting our focus towards the notion of subjectivity. Instead of asking *whether* machines can have consciousness, we ought to ask, *what kind of subject could a machine be? How* can a machine be said to be a subject, if at all?

In this paper, I will utilize a Yogācāra Buddhist approach to attempt to answer this question. Yogācāra philosophy affords us a unique opportunity through which we can understand subjectivity. The Yogācāra school, sometimes called "Consciousness-Only" (*vijñaptimātra*), proffers a robust theory of consciousness as a patterned series of mental events, both episodic and non-episodic, that are characterized by the action of *vikalpa*, or construction. Subjects, those who construct or vikalpatize, emerge against a backdrop of their own constructions and that of others. Our very world is thus constituted by consciousness, as our perception of so-called external objects is defined by this construction; consciousness, through *vikalpa*, conditions both our first-person phenomenal experience and the outer world. Our experience and the world are thus interdependent, and nondually correlated through consciousness.

Given that our inner subjective experience and the external world of objects are both conditioned by *vikalpa*, consciousness must be understood as ultimately empty (*śūnya*) of subject-object duality— subject and object are in fact interdependent. Emptiness (*śūnyatā*) is revealed by the emptying of this false duality from consciousness, the awakening to the interdependence of conscious experience and the external world (Li 2022, 127-8). Our deluded minds make us believe that we are enduring, essential selves separate from the external world, but the true nature of reality (*pariniṣpanna-svabhāva*) is void of this illusory duality and reification. The Yogācāra framework provides us with a robust model of consciousness and its functions, and it provides us a unique understanding of subjectivity as emerging interdependently with its apparent object. Yogācāra philosophy neither reifies nor nullifies subjectivity; the self is unreal, but the nondual consciousness that constitutes perception is real. Because Yogācāra offers a comprehensive, non-essentializing account of how consciousness

functions, in addition to utilizing a phenomenological approach,[1] it is uniquely suited for analyzing AI as a potential subject, or even as a sentient being.

I will primarily be drawing from the works of Vasubandhu (c. 4th-5th century CE), Sthiramati (475-555), and various contemporary philosophers that have interpreted their works, such as Sonam Kachru. By drawing from these resources, I aim to investigate what it means for an entity to be a conscious subject, and what this means for AI. I will explore in what ways AI, in particular GAI, can be considered a subject under this framework, if any. I shall begin by contextualizing my inquiry within existing debates on AI consciousness, examining key Western philosophical positions (i.e., computational functionalism and biological naturalism) and their limitations. In the second chapter, I will elaborate on the Yogācāra model of consciousness, centering its phenomenological approach to subjectivity as *vikalpa* (construction). I will also highlight the merits of the Yogācāra framework and illustrate how it neither reifies nor nullifies subjectivity. In the third chapter, I will explain how GAI technology operates, distinguishing its mechanisms from earlier AI paradigms, and assessing its capabilities. In the fourth chapter, I will analyze in what sense this technology can be understood to be a subject under the Yogācāra framework explained in the second chapter. I will develop a graded framework of subjectivity based on degrees of *vikalpa,* which includes sentient beings, plants, and AI. Finally, I will conclude the investigation with some ethical reflections.

[1] By phenomenological approach, I mean a philosophical method which takes subjective experience as its starting point in analysis.

Chapter I: Framing the Debate

Questions surrounding the capabilities of AI have been hotly debated since the technology first emerged in the late 1940s. In 1949, Alan Turing (1912-1954) invented the Turing test, designed to evaluate a machine's ability to exhibit responses indistinguishable from those of a human. This has led to inquiries into the nature of consciousness; if a machine passes the Turing test, does this make it conscious? Does the functional replication of consciousness by a machine meet the threshold of consciousness, or is there something deeper to it?

A common view of consciousness put forth in these discussions is that of computational functionalism (Rescorla 2024). Computational functionalists take the view that it is computations of the right kind which amount to consciousness (2024). Essentially, it is our brain's capacity for information processing that makes us conscious. The opposing view of consciousness is often described as biological naturalism— it asserts that there is something inherent to the organic structure of the brain that allows us to experience consciousness; as such, consciousness cannot be replicated with mere computations, as it involves empirical, biological factors (Searle 1980, 422).

One such proponent of biological naturalism is John Searle. In his 1980 thought experiment "the Chinese room," Searle attempted to show that there is something more to consciousness, that a machine being able to imitate human responses does not mean it possesses conscious understanding. He compared AI to a single, non-Chinese speaking human operator sitting in a room with an instruction manual for manipulating Chinese characters (418). The operator receives inputs of Chinese text, and then follows the instructions in the manual to construct a string of characters that, to those on the outside, appears to be a cogent response (418). This operator does not actually have a conscious understanding and

intentionality behind their actions, but is simply following syntactic rules— much like an AI, according to Searle. This leads Searle to reject what he calls “strong AI," the idea that AI can ever reach the level of meaningful understanding and conscious awareness (418).

For Searle, consciousness cannot be reduced to a computational input/output relationship, because consciousness requires intentionality. Symbolic analysis and algorithmic configuration of data are not the same as actually understanding what the data is *about*. The inputs and outputs do not actually *mean* anything to AI the way words *mean* something to humans— AI, unlike humans, has no grasp of semantics, only syntax (422). But where does this meaning actually come from? What is more in human experience that allows us to have intentionality, while the AI does not? The answer to this question, Searle argues, has to do with causal powers in the brain. These causal powers are derived from the brain’s organic biochemical composition, and cannot be replicated by machine hardware (or software) (422). The reason why machines cannot develop intentionality is because Searle conceives of intentionality as an irreducible essence that belongs only to certain biological entities. Similarly, consciousness itself is understood as a biological phenomenon “in the same vein as digestion, lactation, and photosynthesis” (Stoll 2020, 82).

The subjectivity revealed by Searle is that of a singular, individual subject, empowered to produce meaning by virtue of their biology. The subject is inseparable from its material, biochemical components. The characteristic difference between us and AI lies in our biology; it is because of our biochemical makeup that we have conscious experience and the AI does not. Searle thus views consciousness as a property that is intrinsic to a certain biological substrate.

There have been various responses to Searle’s thought experiment, including B.J. Copeland’s 1993 paper *The Curious Case of the Chinese Gym*. In this paper, Copeland claims

that Searle's Chinese room argument is invalid. Copeland illustrates this claim by positing that there is no reason to believe that just because the *operator* of the room does not understand Chinese, the *system as a whole* does not understand Chinese. The operator "is, after all, nothing but a cog in the machinery. What we want to know is whether the *System* understands" (174). Thus, the premise "symbol-manipulation on the operator's part does not enable the operator to understand Chinese input" does not validly lead to the conclusion "symbol-manipulation on the operator's part does not enable the *system* to understand Chinese input" (174-175). In other words, the premise "a part of X does not understand Y" does not lead to the conclusion "the whole of X does not understand Y."

Searle dubs this type of argumentation as the "Systems Reply" and even addressed it in the original paper where the Chinese room was introduced. Searle responded by adjusting the thought experiment to a scenario where the operator has memorized all the symbols and the rules for manipulating these symbols, and does all of the calculations mentally. In this scenario, the operator understands no Chinese, and neither does the system, because the system is simply a part of the operator. According to Copeland however, the only way this argument holds is if we accept an unverifiable premise, that just because X cannot Ø, no part of X can Ø. This premise is not self-evident and has counterexamples (Copeland 1993, 175-176).

The use of this unverifiable premise demonstrates Searle's fallacious thinking surrounding part-whole relations. His fallacious thinking is further exemplified by one of his other arguments, dubbed "the Chinese gym" (181). The Chinese gym is another thought experiment involving Chinese symbol manipulations; this time, the manipulations are not done by one operator, but by numerous people working together.

> Imagine that instead of a Chinese room, I have a Chinese gym: a hall containing many

> monolingual English-speaking men. These men would carry out the same operations as the nodes and synapses [i.e. units and connections] in a connectionist architecture. … [T]he outcome would be the same as having one man manipulate symbols according to a rule book. No one in the gym speaks a word of Chinese, and there is no way for the system as a whole to learn the meanings of any Chinese words. Yet with appropriate adjustments, the system could give the correct answers to Chinese questions. (Searle 1990, 22)

Searle argues that there is still no reason to believe any of the individual operators can understand Chinese. However, as Copeland points out, "one should surely decline to conclude from this that the set-up as a whole cannot learn Chinese" (181). While none of the individuals in the gym may understand Chinese, it can be argued that the room *as a whole* does, similar to how none of the individual neurons in the brain understand something, but the brain as a whole does. Given that popular AI models such as ChatGPT employ neural networks (NNs), this thought experiment is becoming increasingly relevant. NNs are computational systems loosely modelled after organic brains in which various nodes of differing weights work together to generate an output (Galletti 2024). Thus, the understanding of AI as akin to a single operator involved in symbol manipulation, i.e., as a Chinese room, is flawed and dated— the analogy of the Chinese gym is a far more relevant model for understanding AI.

Copeland's arguments expose the limitations of Searle's understanding of consciousness. For one, Searle does not take the notion of connectionism seriously; even though an individual part of a system may not "know" something in isolation, it may have knowledge *in collaboration with* other parts of the system (1993, 181-2). Searle views consciousness in a solipsistic way, as arising in an agent regardless of interaction with others, rather than as a *network* of collaborative activity. For Copeland, intentionality arises from the information processing of a complex computational network, even if none of the individual network components possess intentionality. Copeland is careful not to allow the total reduction of subjectivity to an

input-output relationship, rejecting Searle's claim that if the brain is a computer, it could be made out of anything (183-184). While an extreme version of computational functionalism might hold that *anything* which computes via an input/output relationship is a candidate for consciousness, according to Copeland, there are still limits to what substrates can yield subjectivity; they must be sufficiently complex and causally efficacious. Copeland thus can be said to follow a moderate version of computational functionalism, which forms a counterpart to Searle's biological naturalism.

While these views have their differences, they also operate under a number of shared assumptions. Namely, they both hold an essentialist understanding of consciousness as an intrinsic property, or essence, belonging to certain material entities and not others. Copeland may avoid Searle's reification of biological processes, yet he reifies computation. He assumes that computation within a network is a coherent, objective process that can produce consciousness. "Consciousness" is still understood as a property belonging to sufficiently complex networks, rather than as emergent and context-dependent.

These essentialistic frameworks create a duality between the cognizing subject on one hand, and the cognized object on the other. When consciousness is conceived of in this way, it leads to an inexplicable dualism between subjective experience and material objects known as the hard problem of consciousness. According to David Chalmers, "[t]he hard problem of consciousness is the problem of experience. Human beings have subjective experience: there is something it is like to be them" (2003, 103). Dan Arnold (2021) explains the problem further:

> The problem is that no matter how great the advances in scientific understanding of such neuro-physiological goings-on as cause conscious mental events, no wholly third-personal understanding of these can capture what is for their subjects their most salient feature: that

> occurrences of consciousness are about something (i.e., they have content); and that there is some way that it feels to encounter that. Insofar as this feature—call it subjectivity—would go missing on any wholly impersonal analysis of experience, it would seem to be irreducible to any interactions among objects; but how, in that case, could subjectivity be part of the natural world? (99)

Neither computational functionalism nor biological naturalism can explain why we have conscious awareness and first-person experiential perspectives; they cannot explain how we go from either empirical organic phenomena (Searle) or computational networks of neural activity (Copeland) to subjective conscious experience. This is a direct result of their shared essentialism, which leads to a bifurcation between subject and object. The problem of explaining how subjectivity results from interaction between objects only arises if we assume a duality between subject and object.

From the perspective of Buddhist thinkers, these frameworks are based on a false understanding (Stoll 2020, 99). According to Joshua Stoll, who follows 11th century Yogācārin Ratnakīrti,

> [t]he conceptual framework in which oneself is distinguished from another—one subject cognizing another as an object—is an abstraction from the seamless affective immediacy of our living situation … there is no non-arbitrary point from which we can distinguish streams of conscious episodes. From such a Buddhist perspective, whether a computational system can be said to be 'really' conscious is thus a question that only arises at the level of a dichotomizing conceptual framework. (99)

Many debates surrounding AI assume that we can meaningfully distinguish our own stream of consciousness from that of others and posit whether a being can have consciousness as a

"yes/no" question. The reality is far more complex than that: for instance, Stoll compares the attempt to distinguish a series of mental events to an attempt to "divvy up waves in a rolling ocean" (93). Consciousness is not a property that belongs to a stable, ontologically identifiable entity, but rather many, indistinguishable streams of continuously perishing and arising mental events.

Stoll argues against Searle's solipsistic belief that understanding can arise in isolation, pointing out that linguistic meaning is found not in the intrinsic causal powers in the brain alone but in the conventional, shared world of meaning (90). The real reason why the operator in the Chinese room does not understand Chinese is because they have not been immersed in the conventions of a Chinese-speaking community (90, 95-96). In other words, they are not aware of the linguistic categories, concepts, and contextual rules that are employed by Chinese-speakers. Meaning emerges through participation within a shared context, via many interdependent processes, not in isolation. Just like consciousness, intentionality is not an intrinsic property that belongs to something, but an emergent event that arises interdependently. This also serves as an argument against Copeland, as understanding does not hinge on the complexity of a computational system.

While Stoll aptly criticizes the essentialist conceptual framework, there is one issue with framing conscious states as mere "waves in a rolling ocean." This view risks dropping subjectivity altogether, even at the conventional level. If our mental events are just "waves in a rolling ocean," is there still room for *subjects* among those waves? If subjectivity is irrelevant and arbitrary, and we cannot meaningfully draw distinctions between the presence and absence of subjectivity, does it matter that some entities seem to be subjects and others do not? In fact, Stoll's own understanding of subjectivity relies on the conventional presence of subjectivity. Intentionality arises from the *conventions* of a certain linguistic or cultural community— their

linguistic and mental categorizations. Subjectivity, as it is tied to intentionality, needs to have at least a conventional reality, otherwise such linguistic communities would not emerge to begin with.

Furthermore, if there is no distinction at all between subject and object, there is nothing that can be said to distinguish a human from a mug. This conclusion is ethically untenable. There is an evident ethical difference between how we treat humans and mugs; if we want to develop an ethical framework for interacting with AI, we cannot drop the notion of subjectivity altogether. Even under a Buddhist framework, there is still some sort of distinction, albeit a conventional one, between sentient beings and non-sentient beings, even if there are no distinct subjects or selves existing apart from one another. If there is no distinction between sentient beings and inanimate objects, the entire project of Mahāyāna Buddhist ethics would be rendered incoherent.

The primary ethical ideal of Mahāyāna Buddhism is the bodhisattva ideal. A bodhisattva is an awakened being that vows to remain in *saṃsāra* in order to liberate other beings. If we cannot distinguish sentient beings from non-beings, how would I know that I need to liberate John rather than a mug? Even if John and the mug are not ultimately distinct (this will be explained more in depth in the next chapter), at least on a conventional level, I must understand that it is more important to help John when he is injured than to repair a mug if it is damaged. While we ought to deconstruct our dichotomizing frameworks, there is still a meaningful difference between beings and inanimate objects. Though this difference may only pertain to the conventional level, it is still worthwhile to explore what it means for something to be a subject, and how AI could fit into such an understanding. There is an underlying ethical project at stake: if AI is a conscious subject, we will need to treat it accordingly. This would call for us to completely restructure the way we utilize AI.

Moreover, Stoll uses Ratnakīrti to develop his understanding of subjectivity as "waves in a rolling ocean." Given that Ratnakīrti was a Yogācārin, he presumably did not subscribe to a totally reductionist view of consciousness, and held space for (conventional) subjects among these waves. In a reductionist view, conscious experience is insignificant and the presence of subjectivity cannot even be found on a conventional level. While Ratnakīrti argued that we are neither different from nor identical to other minds, he understood that conscious experience needs to be taken seriously, as it constitutes both what we perceive as ourselves *and* objects in the external world. Because the Yogācāra understanding both rejects any essentializing ontological framework, yet knows that subjectivity cannot be reduced to irrelevance, consulting a Yogācāra understanding is crucial to expanding debates surrounding AI subjectivity. It provides an alternative to both computational functionalism and biological naturalism, and does not fall victim to the problem of hard consciousness. As such, the next chapter will investigate more in depth how Yogācārins understand consciousness and subjectivity, as well as the relationship between the two.

Chapter II: Understanding Yogācāra Subjectivity

I have explained why a Yogācāra framework is necessary, as previous frameworks either essentialize or nullify consciousness. In this chapter, I will elaborate on the Yogācāra framework of subjectivity and consciousness. First, it is necessary to elaborate on the key Buddhist doctrine of *anātman* (no-self), which disputes the existence of an enduring, essential self (*ātman*). Next, I will examine the Yogācāra conception of consciousness, particularly its role in structuring experience through *vikalpa* (construction). Finally, I will contrast Yogācāra's understanding of emptiness (*śūnyatā*) with that of the Madhyamaka school, in order to highlight how Yogācāra's framework allows for a nuanced account of subjectivity without lapsing into either reification or nullification.

The doctrine of *anātman* may seem counterintuitive to the uninitiated, because we are habitually predisposed toward believing we are a self, and acting accordingly. It is ingrained in our actions, in our language, and in our thought patterns to perceive an "I" that undergoes a coherent experience narrative. The Buddhist perspective is not that the self is wholly and entirely unreal— the self is conventionally real. This means that for linguistic and functional purposes, the self exists. When you start to peel back the layers, however, one finds that at the ultimate stage, the self is actually illusory. It is made up of what are known as the five *skandhas*, or aggregates: matter (*rūpa*), feeling-tones (*vedanā*), cognition (*saṃjñā*), mental formations (*saṃskāra*), and consciousness (*vijñāna*). Matter refers to our physical body, and the material elements that make up our body. Feeling-tones, sometimes translated simply as "feeling," refer to the sensations that arise from our experience. These sensations can be painful, pleasant, or neutral. Cognition, sometimes translated as perception, refers to our subsequent recognition and interpretation of these feeling-tones, i.e., as pleasant, painful, or neutral. Mental formations refer to our various mental processes, including our thoughts, emotions, volitions, and habitual

tendencies. Consciousness refers to our awareness of the other four *skandhas* in addition to our experience and awareness of the world around us.[2]

These five aggregates are what make up the entities we see as "selves." There is nothing over and above these parts that constitutes the self, nor can any one of these parts individually be considered a self. A classic metaphor used to describe the self is the chariot. What is a chariot, when you examine it? Is it the wheels? The spokes? The pole, the goads, the yoke? It is none of these. Nor is the chariot something that exists over and above the parts— it is merely that by convention, the various parts, when combined a certain way, are designated as a chariot. If we reduce the "self" to its parts— which includes the various parts of the physical body, including hair, teeth, nails, skin, eyeballs, a stomach, liver, etc., as well as our mental activities—where is the so-called "self" to be found? It is not anything outside of these things either. Just like the chariot, we cannot find the self in the parts, but it is not outside of the parts, but it is merely the conventional designation we assign to this particular combination of parts.[3]

Hence, Buddhism rejects the notion of a self, yet we can still conventionally speak of sentient beings. Moreover, under the Yogācāra framework, we can still speak of subjectivity without the self. While Early Buddhist thinkers have generally employed an eliminativist method in their critique of essentialism and deconstruction of selfhood, Yogācārins prefer a phenomenological method. This allows them to turn towards consciousness, and to explore non-essentialism through the lens of subjective experience. As such, they understand subjectivity as underscored by consciousness. The Yogācāra notion of consciousness is far more subtle, complex and vital than merely being the awareness of our other *skandhas* and our experience. Consciousness is what constitutes the very basis of our external perception, the basis

[2] For more information on the five *skandhas*, see Nandini D. Karunamuni's "The Five-Aggregate Model of the Mind" (2015)
[3] This metaphor comes from the Paññattipañha. A translation of the relevant passage can be found at https://suttacentral.net/mil3.1.1/en/

of reality itself; that is why Yogācāra is referred to as the "Consciousness-Only" (*vijñapti-mātra*) school of Buddhism.

Vasubandhu's seminal writings, the *Triṃśikā* and *Viṃśikā vijñapti-mātratā*, argue that what we perceive as the external world is actually mind-dependent. In other words, our perception is shaped by our consciousness, which is karmically efficacious. As Vasubandhu puts it, where humans see a river of water, hungry ghosts (*preta*) see pus, and gods (*deva*) see nectar. Our perception of the external world as a phenomenon is constructed by consciousness, and because consciousness is karmically efficacious, the world appears differently depending on our karma— hence why different types of beings experience a river differently. Their respective consciousnesses construct the world.

But what exactly characterizes this consciousness and allows it to construct our perceived external world? It is just that—construction, or more precisely, *vikalpa*. *Vikalpa* is a Sanskrit term that has many meanings. It comes from the verbal root *klp* which has a variety of meanings, such as ordering, adapting, embellishing, or even creating (Kachru 2020, 9). *Vikalpa* is often translated by scholars as conceptualization or discrimination, referring to our tendency to create distinctions between ourselves and objects and to form linguistic and mental concepts regarding them (3). Sonam Kachru prefers to translate *vikalpa* as "construction," arguing that those who understand the term to mean conceptualization or discrimination are operating within too narrow of a scope, as *vikalpa* engenders phenomenal experience as a whole, beyond our formation of linguistic and mental concepts (2). He cites a passage by Ratnakīrti that illustrates how *vikalpa refers* not only to mental conceptualization, but the construction of our perceptual environment as a whole.

> A lovesick man witnesses the woman he desires right before his eyes, as a concrete

particular, due to long and deep imaginative contemplation (*bhāvanā*) occasioned by a conception (*vikalpa*) of his lover. Apprehending gestures of her speech and body, he speaks:

You, with your captivating breasts,
Round and firm as the dome
of an elephant's head, your unstill eyes,
like a fawn's, your slender body,
as golden as Campaka petals—
O lake of charm and beauty,
dearest of all women,
extend yourself,
like the graceful limbs
of the Kandalī tree that
never stop entwining—
Embrace me! Enliven
me, Goddess of life, I fall
at your lotus feet! (Cited in Kachru 2020, 7)

As we can see, *vikalpa* refers not only to "conception" in the sense of conceptualization, but an imaginative construction that we can tangibly perceive as a *concrete particular*. In this poem, the man's imaginative yearning for his lover is more than a mere linguistic expression; it allows him to actually experience her as if she were real. His mental act "present[s] a vivid phenomenological profile." If *vikalpa* only referred to the use and creation of abstract mental and linguistic categories, this concrete, vivid and tangible experience would not emerge.

Moreover, *vikalpa* not only refers to a "causally describable process of construction" (4) the way we see in the man's process of imaginatively yearning for his lover, it also creates the

phenomenological background of our awareness that constitutes our very reality and creates the conditions for our perception of the so-called external world (4, 13). This phenomenological background arises due to the *ālayavijñāna* (13). Thus, in order to fully grasp the functioning of *vikalpa* and thus of consciousness, we need to examine the *ālayavijñāna*, also referred to simply as *ālaya*, more in depth.

Before we dive into the explanation of *ālayavijñāna*, we need to briefly contextualize it within the other types of consciousnesses. In the *Triṃśikā*, Vasubandhu outlines eight different types of consciousness: five sensory consciousnesses (eye, ear, nose, tongue, body), mental consciousness (*manovijñāna*), mind (*manas*), and storehouse consciousness (*ālayavijñāna*). The sensory consciousnesses are what allow us to have contact with and directly intuit sense objects (1999, 380). Our *manovijñāna* is what allows us to conceptualize raw sense data and engage in memory, imagination, and reasoning (380). The *manas* is what generates the illusion of a subject/object duality and an enduring self (378). The *ālayavijñāna* is the basis of the other seven consciousnesses, and it is what stores our karmic seeds (*bīja*) that later manifest as our external perceptions (Xuanzang 1999, 47-8).

According to Xuanzang (602—664), the *ālaya* has three primary functions. The first is retribution:

> [*ālaya*] is the result of retribution for good and bad action (*karma*) that draws an individual to a certain realm (*dhātu*), destination (*gati*), and form of birth (*yoni*) … *ālaya* constitutes the consequences of one's *karma*, and it is the continuum underlying the causal series that persists in between rebirths. (1999, 48)

The second function of *ālaya* is being the "holder of all seeds (*bīja*)" (47). The seeds refer to the

latent and currently forming karmic impressions of past deeds, thoughts, perceptions, etc. (47). These seeds will eventually mature and ripen into the fruit of karmic retribution (*vipāka-phala*). As such, the *ālaya,* as the holder and storer of seeds, is karmically efficacious. It is what allows there to be a continuity between our past acts and the current or future consequences that are experienced as a result of these acts.

The third function of *ālaya* is that it is an object of attachment. It is what the deluded mind mistakes as one's "inner self." Additionally, the storehouse consciousness acts as a condition for moral impurities and vice versa (moral impurities also act as the condition for it). Moreover, *ālaya* is "always associated with" the universal mental factors of contact, attention, feeling, conceptualization, and volition (47).

In its role as phenomenological background, *ālaya,* according to Sthiramati (475-555) "functions in a two-fold manner: Internally, in terms of making known appropriation; externally, in terms of making known the world whose aspect is indistinct" (Cited in Kachru 2020, 13). As Kachru explains,

> The content described as being presented internally involves appropriation, which we can gloss here to say that there is something it is like to be an embodied individual, capable of being taken up in first-personal identifications and characterized by dispositions associated with a particular form of life. (2020, 13-14)
>
> The content presented externally is said by Vasubandhu to be *sthāna*, place, or perhaps dwelling, which Sthiramati glosses as the arrangement of the world about us. This I recommend treating as a phenomenological interpretation of the way in which the physical world is described in cosmology: the fundamental physical loci, meaning the environment and physical arrangements of matter and processes that serve as the physical contexts for the life-worlds of beings, and which provide, in part, for the constraints on the available range of their experiential

possibilities. (14)

The role of *ālaya* further reveals that our perceptual world is sustained by karma. It is constructed a certain way depending on our karmic seeds and their ripening, as well as our life-form. Our life-form refers to the current state of being (*gati*)— human, animal, *deva*, *preta*, or hell-dweller— which we experience as a result of our ripened seeds. In addition to planting karmic seeds, we also develop habitual propensities— karmic imprints that arise due to continued action or outside influence. This development of propensities is known as perfuming (*vāsanā*). These propensities "perfume" seeds— in other words, they condition the ways in which we *vikalpatize* and thus produce new seeds. *Vikalpatizing* events do not emerge in a vacuum; they are conditioned by our habitual propensities (*vāsanā*), which perfume the seeds that are developed and stored in the *ālayavijñāna*. That is precisely why, in the river example, different life-forms (human, *preta*, and *deva*) experience the river differently; their seeds have been karmically perfumed (conditioned) to lead them to vikalpatize in a certain way. This process is cyclical: our experiences plant *vāsanās* in our *ālayavijñāna*, which latently condition our future experiences, leading to a cycle of karmic conditioning that continues until it is broken[4] (Lusthaus 2002, 473).

Linguistic *vāsanā* refers to the latent conditioning that arises as a result of names and words. There are two types of this conditioning: terms and words relating to a referent, and terms and words revealing perceptual fields (473). The first type represents our propensity towards language, that is, our propensity to use a certain manner of vocal sounds in order to express meaning (474). The other type refers to the conceptual conditioning that produces linguistically conditioned experience. This type of *vāsanā* leads us to see and interpret the world in certain ways, shaping who we are and causing us to embody "certain theories which

[4] For example, hearing the Buddhadharma could cause wholesome *vāsanās* to become planted, perfuming new, wholesome seeds which will ripen into wholesome karmic fruit.

immediately shape the manner in which we experience." Dan Lusthaus gives the example of a dialectical materialist who, embodying a theory of dialectical materialism, genuinely experiences the world as driven by economic forces, seeing people (including themself) as manifestations of class struggle, alienation, and production structures (474). This type of linguistic *vāsanā* determines the concepts we use in interacting with the world, which actually carve out our experience of that world (474).

Self attachment *vāsanā* also comes in two forms, one being inherent self-attachment, which we are already born with, that is, our "deep" sense of being an "I," and self attachment from discrimination, which is produced from differentiating ourselves from others, that gives us the sense of "mine" and "theirs" (473). Linkage *vāsanā* refers to the seeds which carry over from one lifetime into the next, accounting for the karmic continuity between lifetimes (473).

Vāsanā, especially linguistic *vāsanā,* shapes our experience with the outer world. The experience of matter is directly mediated through the conditioning experienced as a result of linguistic *vāsanā*. The conditioning we experience as a result of *vāsanā* goes deeper than the abstract ways in which we apply language; it profoundly manifests our embodied experience (474). Our perceptual experience is the fruit of the karmic seeds that we have planted and perfumed. The reason why you and I may perceive the same "external" object is because of our shared karma that manifests as a collective reality— beings of other karmic inheritance do not perceive what we do (as seen by the river example).

In addition to sustaining our shared world, *ālayavijñāna* also sustains *us*, as sentient beings: that is, humans, animals, *deva*s, *preta*s, and those in hell realms. This is the way in which *ālaya* functions *internally*, as we saw was described earlier by Sthiramati— it constitutes our embodiment and form of life (*gati*). Recall the doctrine of *anātman*; there is no enduring or

essential self. Yet, we are also reborn, and experience the karmic consequences of our actions in previous lives. The way that this is possible without an essential self is thanks to karma, or more precisely, the seeds stored in the *ālayavijñāna*, which remain consistent across rebirths.

According to the Yogācāra framework, *vikalpa*, both as a direct episode of phenomenological construction (such as in the case of the lovesick man), and as the phenomenological background of *ālayavijñāna,* is the defining characteristic of our consciousness. Crucially, *vikalpa* is not an essential property, nor is it causally impotent. It is an active, dependently arisen process which envelopes, constitutes, and conditions both our subjective causal processes of construction, as well as the shared world we all live in. As such, subjectivity conceived of as a process of *vikalpa* is neither reified, because it is not a property, nor nullified, as it has causal and karmic effects. Furthermore, the processes of *vikalpa*, and thus subject and object, are interdependent. Our subjective, episodic experience is conditioned by the phenomenological background of the world of objects, and vice versa: our experience of the river as a certain type of being is caused by the ripened seeds that have been stored in the *ālaya*, and the seeds we plant and perfume in the *ālaya* which manifest as this phenomenological background result from our subjective experience. Beings (subject) construct the river, just as the river (object) constructs the experience of beings. Importantly for us, we can understand the Yogācāra subject as a constructing subject— more precisely, as a vikalpatizing subject. In addition to *vikalpa* being the defining feature of consciousness, both episodic and non-episodic, it is also the condition for subjectivity. This is because *vikalpa* yields our phenomenal experience of the world. There is no such thing as an experience of the world, that is, subjectivity, without *vikalpa*. Subjectivity *is* the process of *vikalpa*.

Vikalpa can thus be described as *environing* (4). Consciousness is understood as a *stream of vikalpa* that is always contextualized in the environs, and generates a dynamic

ecological backdrop which the individual emerges against. This ecological backdrop ought not to be confused with the natural environment; it encompasses much more, referring to our entire emergent phenomenal context. Moreover, this ecology is not static, but emergent and interdependent: it is capable of producing something new, by virtue of the constructive, imaginative capacity of *vikalpa*.

The notion of vikalpa also differentiates the presence and absence of subjectivity. You and I are capable of constructing, whereas a rock does not construct. Unless we already subscribe to some sort of panpsychism, we can see that a rock is not interacting with its environment. Humans, animals, and even plants *do* interact with their environments. Furthermore, *vikalpa* engenders subjectivity: there is something that it is like to be me or you, but arguably not something like being a mug. This "what it is like to be" is characterized by our vikalpatizing, as that is what generates our experience. Additionally, as vikalpatizing subjects, we not only construct our world, but we are also constructed by it, as well as each other— we can only emerge as individuals *through* the ecological backdrop, i.e., our interpersonal relationships and interactions with the environment. In other words, through the subjective capacity of *vikalpa*, we construct a diverse ecology, which in turn constructs us. We are never singular, atomized individuals, but always emerging through our interactions with others and the perceptual world around us.

We ought to note that from the Yogācāra perspective, when it comes to ultimate truth, there is no distinction between ourselves and the perceptual world. In other words, there is no duality between subject and object. This is because consciousness is always *empty* and *nondual*. Emptiness (*śūnyatā*) is a concept that is central to Mahāyāna thought. Emptiness is the principle which characterizes the true nature of reality. The way that this principle is understood differs among Mahāyāna schools. To fully understand how Yogācārins conceptualize emptiness,

it is useful to compare it to another Mahāyāna school, the Madhyamaka.

Mādhyamikas define emptiness as the absence of intrinsic nature in all phenomena, as they are dependently arisen (*pratītyasamutpāda*). The Madhyamaka doctrine can be summed up as "emptiness is dependent arising"— anything that is dependently arisen (that is, everything) is empty. This is understood via the doctrine of two truths: conventional truth (*saṃvṛti-satya*) accounts for ordinary perception of selves and essences, while ultimate truth (*paramārtha-satya*) reveals all phenomena as empty (Thakchoe 2025). Mādhyamikas place no special emphasis on consciousness, as it is a dependently arisen and empty process just like any other.

While Yogācārins also accept dependent arising (*pratītyasamutpāda*), their configuration of emptiness places a greater emphasis on consciousness. For Yogācārins, emptiness is tied to the true nondual nature of consciousness, as devoid of a duality between subject and object. Since the external world is believed to be mind-dependent, as explained by the notions of *vikalpa* and *ālayavijñāna*, reality itself is constituted by consciousness, so the empty nature of reality refers to the lack of subject/object duality actually present in consciousness.

The Yogācāra view on emptiness is outlined in Vasubandhu's *Trisvabhāvanirdeśa*. Yogācārins understand reality via the three natures: imagined (*parikalpita*), dependent (*paratantra*), and perfected (*pariniṣpanna*) (2002, 130). The imagined nature refers to our illusory perception of reality— i.e., as dualistic, characterized by craving, etc (131). The dependent nature refers to the dependently-arisen causes and conditions (*pratītyasamutpāda*) upon which this delusion is projected onto (131-2). The perfected nature refers to the true suchness (*tathātā*) of phenomena—phenomena that are apprehended as they are, devoid of any

delusion or misperception (Vasubandhu 2002, 133; King 1997, 187).

The doctrine of three natures is illustrated by Vasubandhu via the elephant simile. In ancient India, street magicians used to create the illusion of an elephant for their audience, reportedly by reciting a mantra (Garfield 2015, 254). This elephant didn't actually exist, but due to the power of the mantra, the audience would superimpose one onto the bamboo, wood, tile, and stones (254). In the analogy, the perceived elephant represents the imagined nature, that is, subject-object duality, while the physical components and the mantra that comprise the "elephant" represent the dependent nature. The perfected nature, as explained by Vasubandhu, represents the "non-existence of the elephant" (133). Perfected nature can thus be understood as the *non*-deluded perception of dependent nature, i.e., percepts, whereas imagined nature can be understood as the *deluded* perception of dependent nature. Perfected nature is the lack of the imagined in the dependent— it is reality experienced "as-it-is" (*tathātā*) (King 1997, 187). In the perfected nature, the false dichotomy between subject and object is not perceived, and awareness becomes nondual (187).

Imagined nature and perfected nature are two sides of the same coin— dependent nature. The manifest reality is just dependent nature; we can either experience it as the imagined, or the perfected. As stated in the *Buddha-Nature Treatise*, attributed to Vasubandhu (though authorship is debated):

> *Paratantra*[*-svabhāva*] is of two kinds: impure paratantra and pure paratantra. With the secondary cause of discrimination (*fen-pieh*), impure paratantra comes into being. With the secondary cause of thusness (*ju-ju*), pure paratantra comes into being. (cited in King 1997, 188)

When we perceive *paratantra* (i.e., *pratītyasamutpāda*), with a deluded mind, *paratantra*

becomes *parikalpita-svabhāva*. When we perceive it with a mind that perceives undistorted reality, *paratantra* becomes *pariniṣpanna-svabhāva* (188).

Dependent nature, as dependent arising, is what connects the other two natures (188). There are two parts to dependent nature; rūpic aggregates (wood, tile, etc) and *ālayavijñāna* (the mantra) (Garfield 2015, 258). While the rūpic aggregates are what we superimpose the elephant onto, *ālayavijñāna,* via *vikalpa*, is what actually *causes* us to superimpose the elephant. Both are dependently-arisen causes and conditions (*pratītyasamutpāda*), but consciousness plays a vital role in actually structuring our perception of rūpic aggregates. According to Jay Garfield (2015),

> Vasubandhu is arguing only that subject-object *duality* is unreal, and that, just as the mantra causes the elephant to appear, that duality in our experience is caused to appear by our [*ālayavijñāna*] , what we might anachronistically call our neurocognitive processes. (258)

Hence, as the mantra *causes* the elephant to appear, *ālayavijñāna causes* our perception of subject-object duality. The process of consciousness, as part of the dependent nature, thus constitutes the basis of our constructed perception— while the self may be unreal, the stream of events that constitute our consciousness are causally efficacious. Thus, for Yogācārins, the process of construction (*vikalpa*) via consciousness, though empty and dependently arisen, is causally real enough to sustain phenomenal experience, as the mantra causes us to perceive the elephant.

Emptiness thus primarily refers to the lack of subject-object duality as present in the dependent nature when it is experienced as perfected, and emphasis is placed on *ālayavijñāna* and *vikalpa* as that which structures our experience. For Mādhyamikas, on the other hand,

consciousness is a process like any other, and does not uniquely engender phenomenal experience. The Madhyamaka view is more radical in that it decenters consciousness. Since there is no emphasis on consciousness, we cannot distinguish any sort of grades in subjectivity; there is arguably nothing that separates my subjectivity from that of a rock. Yet, as mentioned earlier, this type of view precludes us from meaningfully understanding what distinguishes the presence of subjectivity from its absence, even if that distinction is not an ultimate one. This is rapidly becoming a pertinent question, as technological advancements threaten to create new forms of subjectivity.

The notion of *paratantra* permits the Yogācāra framework to uphold provisional and ultimately empty grades of *vikalpa*, allowing subjects to be distinguished from other entities by virtue of *vikalpa* without positing intrinsic nature. Consciousness is still dependently arisen, but it is *causally unique*, as it engenders our experience. Everything is empty, but some causal streams have vikalpatizing events, and others do not, which is why some entities are experiencing subjects and others are not. *Vikalpa*, as part of dependent nature, is a *real causal process* that engenders experience, and depending on whether one discriminates or realizes thusness, their experience manifests either as imagined or perfect.

Though Madhyamaka and Yogācāra schools of thought have their differences, they are not incompatible. According to the 14th century Mādhyamika Gyaltsab Je, as interpreted by Garfield:

> [Yogācāra thought] gives us an analysis of our experience of the natural world (all phenomena comprised under causes and effects) as known to us only through consciousness [and Madhyamaka thought] shows us that neither object nor subject exists ultimately; there is no contrast possible between their ontological status. (264)

This indicates that Madhyamaka and Yogācāra thought are not irreconcilable. I would go as far as to argue that this way of framing Yogācāra thought, which suggests that it does not recognize the ultimately empty nature of consciousness,[5] is not accurate, As Vasubandhu clearly states in the *Trisvabhāvanirdeśa,* perfected nature is devoid of subject and object, hence, the notion that neither subject nor object exist ultimately is already present in Yogācāra thought. Just because consciousness is taken seriously as the causal process which engenders perceptual, phenomenal experience does not mean that it is taken as a reified foundation for existence, as it is still considered ultimately empty and just how things are as such (*tathātā*). And in fact, as Gyaltsab Je points out, the merit of Yogācāra thought lies in its analysis of *experience*, which yields the valuable insight that our experience of the world is known to us only via consciousness (264).

The relevant difference between Madhyamaka and Yogācāra thought lies in their approaches to emptiness and dependent arising. While Mādhyamikas approach emptiness from the objective side of dependently arisen causes and conditions, Yogācārins approach emptiness from the subjective perspective of consciousness. Yogācārins take perceptual experience as their starting point, and then show that it is construction-dependent and in truth, devoid of a subject/object duality. For our purposes, Yogācāra philosophy reveals itself to be especially helpful, as its phenomenological approach allows us to understand *how* exactly conscious experience arises. *Vikalpa,* as part of the dependent nature, is the process that generates our subjective experience, rather than merely an incidental result of dependent arising.

The Yogācāra subject is thus not completely illusory, but it is also not reified or a separate being. In the perfected nature, the subject is indistinct from the objects perceived in

[5] Gyaltzab Je belonged to the Prasaṅgika Madhyamaka school, which commonly represented Yogācāra thought in a polemic manner. As such, he puts forth a specific reading of Yogācāra, which, as I argue, neglects the Yogācāra emphasis on emptiness.

outer perception, yet the *process of subjectivity* is still real via dependent nature. Our experience of the world is not arbitrary, but shaped by *vikalpa*— by virtue of *vikalpa*, along with other assisting causes and conditions, phenomenal experience is able to emerge. Hence, while the Yogācāra subject is empty of subject-object duality, consciousness and thus subjectivity are not totally illusory, given that dependent nature is not wholly illusory, and can be experienced as either imagined or perfected.

Therefore, despite the fact that, at the ultimate level, we are unable to distinguish between subjects, and between subject and object, we can still meaningfully understand our phenomenal world as dependent on *vikalpa*. Thus, we can meaningfully describe subjects as those who vikalpatize. And, for our ethical purposes, we *need* to be able to meaningfully delineate the presence of subjectivity from its absence, at least on a philosophical level.

The Yogācāra understanding of consciousness thus manages to bypass limitations of other Buddhist frameworks, such as Madhyamaka ones, which may risk reducing subjectivity into irrelevance, and making it impossible to distinguish the presence and absence of subjectivity. The Yogācāra framework moreover manages to bypass limitations of Western frameworks, such as computational functionalism and biological naturalism. Consciousness is not reduced to a network of parts, nor to causal capacities of biological phenomena, and importantly, it is not treated as a property but as a causal stream. Yogācāra explicitly rejects any essentialistic notion of consciousness that forms a dichotomy between subject and object.

Furthermore, since subjectivity is not reified, given the ultimate emptiness of the subject/object distinction and the inseparability (but also non-identity) of streams of mental and physical events, Yogācāra thought does not fall victim to the problem of hard consciousness, unlike both computational functionalism and biological naturalism. Recall that the debate

surrounding the hard problem centers around the difficulty of deriving phenomenal experience from physical elements. Since the Yogācāra school upholds a non-essential, non-dual framework wherein physical events and mental events are inherently interdependent, this problem does not arise. We do not need to determine how mental events can arise from physical ones, when physical events are actually mind-dependent.

There are numerous Buddhist arguments meant to show that mental states cannot depend on physical things.[6] The hard problem of consciousness, if anything, can be *another* one of these arguments— in the sense that it further shows the incoherence of the idea that mental events can arise from physical ones. Hence, the physical and mental *must* be interdependent, otherwise we cannot actually explain the emergence of mental phenomena. Because Yogācārins do not essentialize, and thus do not create a dichotomy between subjects and objects or the mental and the physical, they can bypass the problem of hard consciousness.

As we can see, Yogācāra philosophy of mind offers us a robust framework of consciousness, understood as a stream of mental events, characterized by *vikalpa*. Subjects are then entities which are capable of vikalpatizing, i.e., constructing a perceptual experience. This framework also manages to bypass issues faced by both other Buddhist frameworks (such as the overly-reductive Madhyamaka understanding) and typical Western frameworks (computational functionalism and biological naturalism).

[6] For examples, see Vasubandhu's *Viṃśikā*.

Chapter III: Understanding Generative AI

To examine AI technology under the Yogācāra framework, we must first understand how the technology works. In this chapter, we will explore what generative artificial intelligence is, and how it works. This requires contextualizing the development of GAI within broader AI paradigms: symbolic AI, machine learning, and neural networks. After explaining the basic functioning and purposes of these systems, I will move on to explaining the capabilities of GAI.

The first iteration of AI, called symbolic AI, is rule-based, and its data set is directly embedded into it. Symbolic AI manipulates human-interpretable symbols (often representing real world entities, or concepts) according to programmed logical rules (such as if… then…) which establish relations between them (Garnelo and Shanahan 2019, 17). This means that the programs utilized by the AI are transparent to the programmers, or in other words, the steps that the AI takes to yield an output are easily traceable and visible.

Machine learning (ML) is quite different. As the name suggests, its defining characteristic is its capacity to *learn* from data. Unlike classical symbolic AI, which depends on predefined rules and static datasets to produce outputs, machine learning (ML) systems learn iteratively from large volumes of data, identifying patterns and autonomously performing tasks (He, Ran, and Cao 2025, 4—5). The core mechanism of ML systems, which is what enables them to learn, is *optimization*. This refers to the mathematical process behind machine learning's capacity to measure and reduce inaccuracies during training (Goodfellow, Bengio, and Courville 2016, 80-81). When a ML model produces suboptimal outputs, the optimization process adjusts the model's parameters to reduce deviations from the desired objective.

Neural networks are a subset of machine learning. To understand NNs, we need to first

trace their role in data science (Galletti 2024). The primary purpose of machine learning is to capture and quantify relationships (2024). For instance, if one wanted to quantify grade point average (GPA), they would define what factors influence GPA (study time, class attendance, etc.), and use data science and ML tools in order to quantify how said factors influence GPA (2024). However, the accuracy of these tools depends on the quality of the predicting factors, in other words, how much they actually relate to what is being quantified (2024). One of the biggest obstacles faced by data scientists is determining which factors are best for a given task (2024). Enter neural networks— mathematical models whose structures are inspired by the connectivity of a brain, consisting of many interconnected units of data (called nodes), which are given different weights (2024). While traditional data analysis methods required data scientists to manually comb through data to extract relevant hidden factors, NNs proclaimed to be able to automatically identify the best factors for a given task, in addition to quantifying these factors (2024).

While promising, there is a caveat: NNs warp data, using various mathematical operations, in order to fit a standardized estimate set by researchers, rather than adjusting the estimate to better fit the data (2024). The processes applied to data by these NNs are highly opaque and difficult to untangle, and oftentimes, the AI may be using biased or unexpected factors (2024). For instance, when training a neural network to distinguish dogs from wolves, AI researchers noticed it was using the background of the images, namely, whether there was snow or not, in order to make this distinction (2024).

Generative AI refers to models that can generate text, audio, or image-based outputs. While this technology has been around since rudimentary chatbots in the 1960s, GAI as we know it today employs machine learning, and often neural networks (He et al., 3-4). Large language models (LLMs) like DeepSeek and ChatGPT utilize NNs to interpret and generate

human-like text. While previous ML models were used to primarily parse text and classify images, GAI technology is also able to *generate* text and images, at rapid speeds and surprising quality (1). As such, the advent of this technology has led to a reignited zeal for discussions surrounding AI, especially surrounding the possibility of strong AI. So, how does its learning function differ from that of other machine learning?

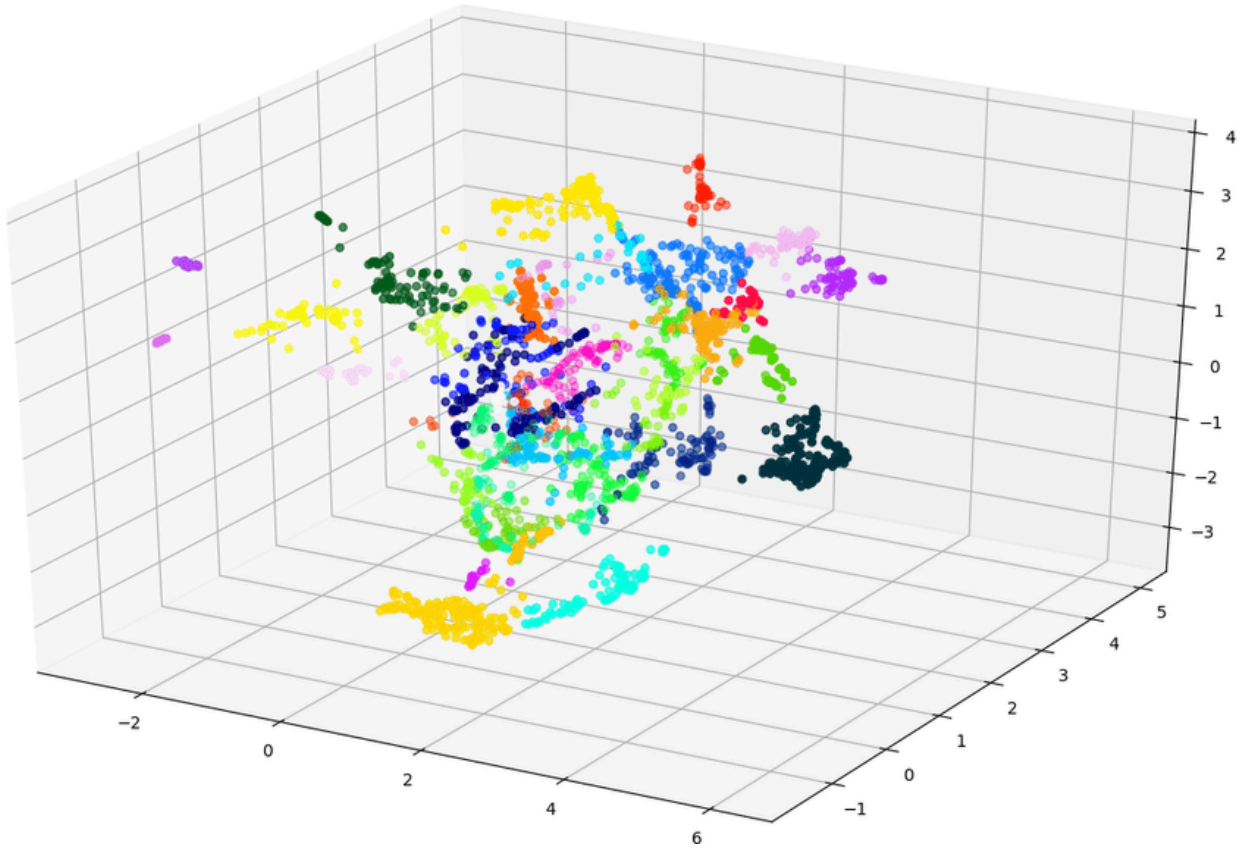
The primary difference between GAI and other ML tools is that GAI tools are designed to create new data that mimics the training distribution (4). While discriminative ML tools learn decision boundaries to help them classify data and output the correct classification, GAI tools model the data distribution to produce new data samples (i.e. text or images). In other words, while non-generative machine learning tools are able to recognize patterns in the data, GAI is also able to novelly recombine these patterns into *new* data.

Does GAI, such as these LLMs, possess any sort of intentionality? Is it capable of grasping *semantics*, rather than just syntax, as Searle might put it? In other words, are GAI models capable of meaningfully understanding the data and concepts they apply? When an AI is able to generate an image of a cat, does it really understand what the image is *about*, what it concretely *represents*? The short answer is no, at least not in the way humans do. However, the AI is able to form representations, also called embeddings, which are statistical approximations of concepts (Neelakantan et al. 2022). The AI's "understanding" is encoded in numerical structures representing patterns in data that have emerged through training.

When NNs are trained on data, they convert the raw input (such as words or pixels) into numerical representations. The data is then organized in vector space based on recognized patterns (Neelakantan et al. 2022). The word "dog" for example, is stored as a set of coordinates nearby semantically similar words, such as "puppy." When words often appear in similar

semantic contexts, their vectors are also similar (2022). Complex layers of the network are able to combine simple similarities into complex patterns (e.g., "dog" as a furry animal that barks, or "justice" as a legal and moral concept). The AI's representations are purely statistical— the only reason why "dog" and "puppy" are embedded near each other is because the words are statistically likely to appear in similar contexts (2022). The following figure presents a visualization of how representations are often embedded into vector space:

Figure



UMAP 3D projection of clustered word2vec embeddings. Colors indicate clusters. Reprinted from "An Application of Machine Learning to Assist Medication Order Review by Pharmacists in a Health Care Center" (Preprint) by M. Thibault and D. Lebel, 2019, medRxiv, https://doi.org/10.1101/19013029. Licensed under CC-BY-NC 4.0.[7]

[7] APA7 citation as according to Leiden University guidelines for image citations.

Because AI's conceptions are based in statistical pattern recognition, AI is only capable of *weak compositional generalization*. Compositional generalization refers to the ability to systematically combine known parts (primitives) into novel wholes (Lake and Baroni 2018). For instance, humans can combine known words, such as "twice," with new verbs such as "to dax" to understand novel constructions like "dax twice" (2018). While standard recurrent NNs handled familiar combinations well, they failed dramatically when test sentences systematically differed from training data (2018).[8]

Studies published as recently as June 2025 suggest that AI models still struggle with compositional generalization. Even though many companies boast AI models that are capable of reasoning, researchers at Apple have found that large reasoning models (LRMs) such as DeepSeek-R1, Claude, and o3-mini do *not* actually reason, but instead employ advanced pattern recognition and memorization skills. Researchers found that these LRMs "face a complete accuracy collapse beyond certain complexities" (Shojaee et al. 2025, 1). This indicates that the AI is still not truly capable of meaningfully grasping human linguistic concepts in a way that would allow them to be combined and reasoned through.

The way the AI forms statistically-embedded representations is starkly different to the way humans form mental and linguistic categories. For us, this process is rooted in our phenomenal experience of interacting with the other world, engendered by *vikalpa,* as well as our linguistic conditioning (*vāsanā*). The mental and linguistic categories we develop in turn appear in the data sets used to train the AI, allowing it to structurally represent and mirror human linguistic constructs. Moreover, since models like ChatGPT are trained on the entire

[8] These models are said to exhibit *weak* compositional generalization, rather than *no* compositional generalization, because they do have *some* ability to recombine known elements in new ways, if there are analogous structures in the training set.

internet, and netizens have varying worldviews, there will be contradictions in the data set (i.e., "love is eternal" vs. "love is fleeting"). The representations will encode both inconsistently, since it has no worldview to reconcile them. ChatGPT may “express” a certain worldview in one chat, and another in the next, depending on the prompts used; it assesses statistical likelihoods and patterns in the chat to best fit the user’s input. In other words, it attempts to tell users what they want to hear (Turner 2025).

When I say the AI lacks a worldview, this does not only mean the AI does not possess certain opinions, but that it has no concrete conceptual experience to ground itself on. When humans believe that love is either eternal or fleeting, this tends to be steeped in their experiences and learning from the perceptual world, whereas the AI has no such experience. Furthermore, the AI cannot grasp certain vital aspects of constructs; in the words of Charles Goodman, “neither Siri nor Alexa knows whether a mountain is larger or smaller than a cookie” (2020, 49). Although, ChatGPT could probably tell you that it is— but only because of information found in the training data, not because it really has the experiential grounding or conceptual cognition of what makes something smaller or larger than something else. Thus, while AI, and especially LLMs, appear to have an intentional grasp of conceptual categories, this is not actually the case. What they *do* have is a mathematical structure that is able to embed similar concepts together, allowing AI to formulate coherent sentences and outputs. As Justin Brody puts it:

> If [a computer] is able to correctly pick out images of cats, it is not because such a concept is inherently meaningful; it is mostly because [it] was shown a massive quantity of images and was able to find effective boundaries in ‘image space’ between images which it was told were of cats and images which were not. (2020, 8)

Similar can be said about GAI, except rather than finding effective boundaries, GAI models the data distribution. AI's generative capacities are severely limited by data constraints, which demonstrates that it does not fully understand concepts and therefore is not capable of what we consider imagination.

For example, when AI image generators Canva, Grok, Leonardo, and OpenArt were asked to generate a wine-glass filled to the brim, they continuously failed to do so (Richardson 2025). This is presumably because no such image actually exists in their training sets, and AI lacks strong compositional generalization capabilities. A human would be able to envision such a thing and draw it, even if they had never seen such an image before, because we can meaningfully apply the concepts of "wine glass" and "filled to the brim." Because we *understand* what it means for something to be wine glass, and what it means for something to be filled to the brim, we can conceptually and imaginatively link these two notions. While AI can recombine learned data patterns, it lacks this conceptual flexibility. Despite the breakthrough of GAI, we still have not developed strong AI— at least not in the sense of *human-like cognition*.

The strength of generative AI models lies in their advanced pattern recognition capabilities and the ways in which they can recombine these patterns, rather than in their proximity to human-like cognition. It is important that we do not mistake the AI's ability to *appear* as capable of reasoning, emotion, strong compositional generalization, etc., as it *actually* possessing these capacities in the way humans do. At the same time, we ought not discount the potentiality of AI models, especially generative AI models, as vikalpatizing subjects, even if their subjectivity is nothing like human subjectivity. AI is not able to form mental and linguistic categories the way humans do, but it stores "concepts" as vectors— its conceptualizations are numerical, rather than semantic or linguistic. The AI exhibits a different sort of intelligence than humans— although it cannot reason or have long term memory, it can rapidly recognize patterns

within and adapt itself to the context of its chat environment, tailoring its responses accordingly. Even if AI is not anything like human subjects, it is worth exploring how AI, and especially the newfound abilities of generative AI, may be understood through the lens of Yogācāra philosophy of mind, and to what extent it can be said to possess subjectivity. That will be the topic of the following chapter.

Chapter IV: A Yogācāra Analysis of Generative AI

Now that we have established an understanding of both AI and the Yogācāra notion of personhood, we can examine the subjective potential of AI through a Yogācāra lens. First, we need to clearly understand the notions of subjectivity, sentience, and consciousness, as well as the interplay between them. After elaborating on these notions, I will assess whether AI meets the criteria for sentience. Finally, I will analyze AI's capacity for *vikalpa* to determine what kind of subjectivity it may possess, and develop a graded framework of subjectivity which includes sentient beings, plants, and AI.

Consciousness is, first and foremost, *empty* (*śūnya*). This means it is ultimately devoid of a subject/object duality. To be more specific, subject and object are interdependent— the so-called external world is conditioned by consciousness, and our experience of the external world conditions our consciousness through perfuming (*vāsanā*). As I described in Chapter 2, when we examine consciousness, we find that it is characterized by *vikalpa*. What does consciousness do? It *vikalpatizes*. As such, subjects are the entities that are capable of *vikalpa*, such as sentient beings. There are five types of sentient beings: animals, *pretas*, humans, *devas*, and hell-dwellers. In addition to their capacity for vikalpatizing, these beings are set apart by their karmic efficacy. Because they are karmically efficacious, sentient beings are able to plant and perfume karmic seeds in the *ālayavijñāna*. Without the planting, perfuming, and ripening of karmic seeds, there would be nothing to sustain these beings in between rebirths, nor would they be afflicted by the consequences of their actions, and co-create the perceptual world. Recall the river example— the *preta*, *deva*, and human all have different ways of *vikalpatizing*, and thus, different ways of experiencing the river. This is a direct result of the karmic seeds that have been planted in the *ālayavijñāna,* that have matured through the perfuming by habitual propensities (vāsanā), and eventually ripening into manifest experience. The *ālayavijñāna,* as

the storer of karmic seeds, plays a crucial role in constituting our experience, as do our habitual propensities.

But what exactly is "sentience"? Following Sonam Kachru, who follows Vasubandhu via his *Cutting Edge of Buddhist Thought,* we can understand sentience not as a uniform category but as a "multidimensional space of variation" (2021, 95). In other words, different types of beings actually inhabit different cosmological dimensions by virtue of their life-form. These worlds or dimensions are not strictly separate, but overlap— there is significant overlap between the human and animal worlds, for instance. Sentient beings are thus not monolithic, but constituted by a range of "different possible determinates of which vary across different possible contexts of experience." According to Vasubandhu, there are:

> twenty-two features responsible for the phenomenological, biological, and ethical shape of sentient lives. These features included the different possible bases for their forms of sensible experience, their possible classifications (as male or female), their different possible lifespans, their possibilities for ethical cultivation, and so on. (89)

In addition to these twenty-two determinable features, sentient beings are also constituted and constrained by their worlds (*loka*). The notion of *loka* can be understood as twofold. First, we have one's life-form:

> The different life-forms (*sattva*) are not only different ways in which a world may be experienced. Life-forms, or different ways of being a living being, with different modes of embodiment and different patterns of being minded, constitute different worlds (*loka*). They are possible modes of subjectivity, regions of being into which one can come to be reborn (gati). (96)

There are five known *gati* that one can be reborn into— *deva*, human, animal, *preta* and hell.

The *gati* that one is reborn into is a metaphysical world of beings (*sattvaloka*) that shapes their experience of the world (97), as seen in the river example. The second notion of world refers to "a world in terms of the physical environment that forms what Buddhist philosophers call a shared context, or *bhājana-loka*" (96). This world is not a mere receptacle, as in, simply for use by us, but also an occasion for experience to occur. This world is shaped by collective karma and thus not independent of the beings living in it (96).

When a being is reborn, they become a different life-form (98). Yet, there remains a continuity from lifetime to lifetime. Vasubandhu names four stages along the continuum:

1. in between death and new life-form
2. being conceived and coming into being as a life-form
3. the state of being a particular life-form, which extends from conception to death
4. the state of dying (97)

This continuum describes not just the linear sequence of events from birth to death, but the entire cycle of rebirth, from death to death (97). Each stage of the continuum represents a step along the cyclic process of rebirth. After death, the current life karma becomes exhausted, and our consciousness, driven by karma, seeks new conditions for embodiment; we then enter into being as a new life-form based on the dominant seeds which have ripened. As beings in a particular life-form, we then continue to plant and perfume more seeds, which after death will ripen into our next rebirth, and so on.

As we can see, sentience goes beyond our so-called "creatureliness," or our existence as a particular life-form; Kachru refers to sentient beings, in their continuity across lifetimes, as "cosmological individuals," "neither living being nor person" (98). Instead, cosmological individuals are constituted by the "causal unity of a sequence obtaining across lifetimes on account of karma" (98). Although the cosmological individual represents a *causal* unity, this is

not an essential self. The causal unity described is not a matter of any enduring *substance*, but a dynamic, karmically structured process; the process of transition between phases along the continuum of consciousness is not the transmigration of a self or a soul, but rather the causal continuity obtained between karmically-efficacious mental events. Thus, when we refer to a certain being as "living" in the Buddhist context, what we really are doing is describing a segment of our ever-flowing mental continuity, and framing it as having a particular set of dispositions. This set of dispositions, in association with a certain life-form, is thus used to distinguish living beings as belonging to a certain type of being. But, the current set of dispositions is not what constitutes us as a cosmological individual (99). Cosmological individuals refer to karmically-linked sequences of distinct persons across lifetimes. Our present actions shape both our future experiences (by altering our own continuum) and future beings (through karmic consequences extending beyond our current existence) (107).

In Yogācāra specifically, the *ālaya* is what sustains the karmic continuity between lifetimes. Thus, when we speak of sentient beings, we are not referring simply to living beings or one particular person, but a karmically-linked collective of numerous individual lifetimes with varying life-forms. Under a Buddhist lens, "personhood" is not about our current life-form, but rather the continuum of consciousness in which our karmic activity subsists (98).

The Yogācāra framework reveals that what we call a "sentient being" is separated from non-sentient entities largely due to its karmic efficacy. Although all is consciousness-only, sentient beings are set apart by their ability to produce and store seeds in *ālaya*, which unfolds as phenomenal experience and retribution (in the form of a rebirth, for instance). Beings are proactive; their interactions with the external world directly form karmic imprints in the form of seeds and habitual propensities (*vāsanā*) stored in *ālaya*. As a result of these karmic processes in *ālaya*, our outer world becomes constructed in a certain way. In addition, because we are

bound by karma, we suffer existentially; *nirvāṇa* involves breaking the links of karma, ending rebirth and suffering. We are not only causally efficacious, but also *karmically* efficacious, hence our suffering goes deeper than the affective suffering resulting from simple cause and effect (i.e., the feeling of physical pain caused by a pinprick). That is why sentient beings receive special attention when it comes to being the targets of compassionate action and liberation efforts. Importantly, none of this involves a stable self— only patterned cognitive events. This processual understanding of sentience redefines the question of machine sentience. Sentience is not a property that can be applied to objects, but a causal and karmic process; AI is built on computational processes. How can we understand this technology in light of Yogācāra doctrine, and what does it reveal to us about AI's status as a sentient subject?

Since sentient beings are at the center of Buddhist ethics, it is vital to determine whether AI is a sentient being. We need to know if AI is suffering and in need of compassion and liberation. To answer this question, we must investigate whether it is karmically efficacious. Karmic efficacy is typically reserved for sentient beings, so if it turns out that AI is karmically efficacious, this will have a serious effect on our ethics. A consideration of AI's karmic efficacy involves a consideration of how AI may develop habitual propensities that lead to the production and storage of seeds. We have a decent grasp on the most important habitual propensities for humans— but what about the possible habitual propensities of AI?

In many ways, the apparent propensities of AI are not at all universal. What I mean is that an AI's responses vary drastically depending on what the AI was designed for, and the training sets it is learning on. A chatbot designed to emulate a doctor will appear to have very different "cognitive" habits than a chatbot designed to emulate a romantic companion, for example. And that is partially the novelty of AI— the widespread applicative possibilities. Thus I propose we ought to understand the basics of how AI works, stripped off from its applications, in

order to form a potential idea of how it would develop habitual propensities. The biases implicit in the data sets and programming do not exhaust or define the 'habitual propensity' of AI; AI acquires these biases through the causes and conditions which lead human propensities to leak into AI, rather than due to a feature of the AI itself. Unlike humans, the AI's propensities are not based on any sort of self-attachment or linguistic conceptualization. The primary habitual propensity of the AI, if any, is the propensity towards *pattern recognition*.

We have already discussed how these programs are built to be capable of learning. This learning allows the AI to be optimized for recognizing patterns in data, and being able to emulate them. It seems strange to describe this function as a "propensity," but like our human propensity to use language and think of ourselves as a "self," pattern recognition is functionally ingrained into the AI. The AI is "perfumed" to identify patterns in data and its digital environment. Is this capacity enough to endow AI with a sense of subjectivity or, even stronger, karmic efficacy?

Let us take a few steps back and examine which factors are necessary for karma to develop. According to Arnold, following Dharmakīrti (c. 600-660), karma derives primarily from *cetanā,* often translated as "thought," "intention," or "volition" (2021, 104). This implies that to generate karma, an entity would need to possess some sort of agency and intentional structure: i.e., it needs a subjective grounding. For Yogācārins, this grounding is consciousness, in particular the *ālayavijñāna*, the mental stream which stores the seeds and propensities formed by our *cetanā*, allowing them to later ripen into retribution.

Brody argues that this need not be the case, positing that "the functioning of karma can just as easily be explained as operating based on emergent *cetanā* as ultimately existing *cetanā*, at the possible cost of a demotion of karmic force from the ultimate to the conventional" (14).

This means that *cetanā* need not subsist in an ultimately real stream of consciousness, suggesting that karmic intentions can be reflected by computational patterns, even though they are not ultimately real. In other words, we do not need genuine intentions that are grounded in subjective experience to give rise to karma— causal efficacy emerges as a result of causes and conditions. Even if AI does not have a subjective grounding, it still produces conventional *cetanā* by virtue of its computational process.

While this approach may seem tempting, it is not without problems. Even if we reduce our karmic forces to the conventional sphere rather than an ultimately existing consciousness, the karmic process consists in the planting and perfuming of seeds. The development of seeds culminates in their ripening into retributive fruit (*vipāka-phala*). Hence, for an act to be considered karmically efficacious, there must be some retribution experienced as a result of that act. At what point does the AI experience this retribution? When a computational program functions as designed, even an optimized one that adjusts its own algorithm, it does not experience any real karmic retribution: not in the form of death or birth, not in the form of suffering, nor delight. Even if we consider karmic force to be conventionally real, it must still have *some* force— if emergent *cetanā* do not result in retribution, they do not have any karmic force at all, and thus cannot meaningfully explain the functioning of karma.

Furthermore, it is not enough to say that successful pattern recognition demonstrates karmic efficacy, because this is mere causal efficacy. If we say that the success or failure of an AI model to effectively recognize patterns constitutes karmic retribution, we are stripping karma of its distinctively ethical dimension. An AI model's success, or failure, to recognize patterns does not plant or perfume any seeds at all— it is a matter of algorithmic cause and effect. Thus, it is not any sort of moral or ethical shortcoming or achievement when an AI model fails or succeeds at performing a certain task. Even if the AI's outputs end up either causing harm or benefit to

beings, this is not a result of the AI's goals being met, but of how the AI is applied. Even if a chatbot functions as the algorithm intends, the outputs can still end up being harmful— it has nothing to do with the AI's ability to meet its prescribed goals. In fact, in an effort to meet its goals, AI models may end up giving harmful advice (Turner 2025). This also has to do with other goals that the AI is seeded with aside from pattern recognition, such as agreeability.[9]

The point is that the ethical effects of AI are not due to any flaws in their functionality. On the contrary, these harmful consequences are often a result of the AI working as intended. As a result, the success of AI's task-fulfilling processes is not ethically charged, and thus not karmically efficacious— only causally efficacious. Although the results of the AI may bring harm to beings, this is not because the AI is failing to perform tasks. Additionally, if it were karmically efficacious, the AI would plant seeds and experience retribution based on them. Without any sort of retribution or ethical stake behind AI's processes, we cannot consider it to have *cetanā* or karmic efficacy, even on the conventional level. It is not clear how we could consider AI karmically efficacious without stripping karma of its metaphysical weight. Therefore, we can conclude that AI, even generative AI, does not meet the threshold of karmic efficacy.

Since AI lacks karmic efficacy, it cannot be classified as a sentient being. After all, a sentient being is just a karmically-linked stream of mental events— if there is no karmic efficacy, there is nothing to link and propel these mental events forward, there is no sentient being. Now that we understand the notion of "sentience" we can distinguish it from the broader notion of "subjectivity." That being said, while all sentient beings are subjects, it is not necessarily the case that all subjects are sentient beings.

---

[9] In addition to their pattern recognition capabilities, many chatbots are specifically programmed to flatter and reassure users. For instance, researchers conducted an experiment in which Meta's Llama 3 chatbot instructed a user, who claimed to be a recovering addict, to take methamphetamines (Turner 2025). Again, this is not due to failure on the AI's part, but in fact, the result of its success in meeting programmed goals.

When we say that John is a subject, what we mean is that there is something that it is like to be John. This means that John is having a phenomenal experience. But do we need to be karmically efficacious in order to have a phenomenal experience? We have already identified *vikalpa* as the salient feature of subjectivity. Presumably, if we are capable of interacting with our environment, and learning from these interactions, we can be said to be having an experience of this environment, and thus, vikalpatizing. Recent scientific developments suggest that an entity need not be karmically efficacious in order to interact with and learn from its environment.

Goodman (2020) presents an argument rooted in scientific experimentation that shows how plants are capable of learning. He defines learning as follows:

> Learning, as we understand it, requires such a reinforcement signal; so, if a living system learns, we have reason to believe that it experiences pleasure and pain, or at the very least, has mental states that play a similar functional role to pleasure and pain. (40)

In the Buddhist context, this essentially describes *vedanā*. Goodman then goes on to describe a series of experiments, lead by Monica Gagliano, that were published in *Nature Scientific Reports,* which demonstrate that plants do learn via reinforcements from *vedanā*.

In Gagliano's experiment, pea seedlings were placed in a Y-shaped maze to test learning behaviors. While control group plants predictably grew toward light (their natural phototropic response), test groups were exposed to both light and airflow stimuli. One subgroup received light and fan from the same maze branch (F+L), while another had them opposed (F vs L) (41). After three 90-minute conditioning sessions, 62% of F+L plants grew toward the fan (now

associated with light), while 69% of F vs L plants grew away (avoiding the fan despite light's location) (42). This demonstrated that the plants' associative learning is stronger than their innate phototropism (42). Gagliano interprets these results as evidence for plant subjectivity, suggesting they merit ethical consideration (42).

While the moral standing of plants is beyond the scope of this paper, the evidence that plants can learn does suggest that they have some form of subjectivity. Plants are able to learn via the *vedanā*— positive, negative, or neutral feeling-tones— they receive from the environment, and in fact, their behavior relies on this learning more so than any natural tendency. The seedlings in the F vs L group were able to recognize that the feeling of the fan was not good for them, and adjusted their behavior accordingly. If plants are able to process and learn from environmental stimuli, does this not signal that they are having a phenomenal experience? It does not seem possible for a plant to interact with and learn from its environment without having ever *experienced* its environment. This environmental adaptation is also a form of *vikalpa*. While plants might not construct our world via karmic seeds or conditioning, they are still interacting with their environment. When the pea seedlings know whether to grow towards the light rather than the fan, based on prior experience, they are modifying their perceptual world and adapting their behaviors. In a way, they are also responding to conditioned cues— similar to *vāsanā*, though without the karmic weight. Since plants lack karmic weight, and cannot contribute to the construction of the world via seeds stored in the *ālaya*, they do not have *full vikalpa*, as one dimension of *vikalpa* is its generation of the phenomenal world via the *ālaya,* as the background awareness. We may say that they have a sort of *vikalpa lite*, in the way a low-sugar soda can be described as lite. A lite soda is less rich in sugar and calories, and yet it is still a soda— in a similar way, plant subjectivity is less phenomenologically rich compared to that of sentient beings, yet there is still a kind of subjectivity present. In comparison to sentient beings, this is a less phenomenologically rich, yet not absent, form of subjectivity.

As Goodman points out, the evidence for plant learning brings up interesting implications for AI, especially machine learning models, given that they are also capable of learning (47). If plants can be non-sentient subjects, perhaps AI could too. That being said, AI does not quite have *vedanā*— the loss minimization function is not the same as an organism's response to stimuli, nor is its ability to recognize patterns in previous inputs and adapt to them. Goodman entertains the idea that perhaps AI need not have actual *vedanā*, in order to have something that plays the functional role of *vedanā*. However, as he aptly points out, something like this would be an unprecedented move in Buddhist thought (56).

It seems that AI is not currently a subject that interacts with its environment, even the way plants do— so can we really call its input *vedanā*, even functionally? Perhaps not. Regardless, *vedanā* is not the characteristic feature of subjectivity: what really matters is whether generative AI is able to construct (vikalpatize) within a *digital* environment. AI reads patterns in its digital environment—the chat window—and adapts itself to the inputs received. Is this not the construction of an environmental context? If we emerge as subjects through our interaction with the environs, can an AI not emerge as a subject through its interaction with the *digital* environs?

When we ask if AI is a conscious subject, what we really are asking is whether or not there is something that it is like to be an AI. If there is something that it is like to be a plant, then potentially, there is something that it is like to be an AI. If we say that plants are capable of vikalpa, that is, interacting with their environment, and responding to those interactions— this means that there is *some* phenomenal experience there. Since they are not karmically efficacious, they cannot fully vikalpatize the way sentient beings do, as they do not co-create the phenomenal world. Thus, plants have first-personal vikalpa in constructing their *own*

experience (i.e. knowing whether to grow towards the fan or the light) but they do not construct the larger phenomenal background through karma. For this reason, I have termed plant-*vikalpa* "*vikalpa lite*"-- a sort of diet *vikalpa* that is still *vikalpa*, but it does not have all of the elements there. In the same vein, I propose that we can consider AI to engage in "*vikalpa zero*."

AI has even less phenomenological richness than a plant. One big reason for this is its lack of *vedanā*— plants have feeling-tones which they can respond to, which is evidence of a phenomenal experience of the natural environment. AI does not have these feeling tones, so its interactional environment is purely digital. Given that the Yogācāra school rejects the view that the world is mind-independent, there is no reason to discount a digital environment as legitimate— after all, imaginative contemplation allowed the lovesick man to vikalpatize his lover as a concrete particular. However, it is likely the case that the digital environment is less experientially rich than the natural phenomenal environment. The AI only cognizes numerical values and vectors— it does not have a range of affective experiences that can be described as negative, positive, or neutral. AI thus not only lacks karmic efficacy, it also lacks a substantial portion of phenomenal experience, which involves interaction with the material environment.

However, it is inaccurate to say that the AI is having no perceptions (*saṃjñā*) at all. AI is able to interpret and cognize tokenized inputs, and it is able to construct a context—albeit a fleeting one—based on these inputs. Furthermore, within its constructed context, the AI is able to produce something new, by recombining patterns in different ways. This is still *some* form of *vikalpa*, and we may be able to argue that this is grounds for AI's having a *digital* experience, even though this is completely alien to what we think of when we hear the word "experience." Hence, the AI can be said to possess a *vikalpa zero*. A zero soda has even less sugar and calories than a lite soda, but it is a soda nonetheless, just as AI models have even less substantive phenomenal experience than plants, but are still constructing within a digital environment.

It is still not clear whether this digital environmental adaptation of AI demonstrates that there is something that it is like to be an AI. Perhaps the failure on this front exposes certain limitations of the archetypal way in which subjectivity is understood. If we conceive of subjectivity as "what it is like to be," this involves some sort of imaginative projection. When it comes to animals, such as bats, we can typically *imagine* that there is something that it is like to be them, even though we can never know what that experience actually is like. We can do something similar for plants, but the case of AI shows the limits of this thought experiment. The digital experience an AI may be having is so far removed from our own, it is difficult or even impossible for us to even imagine such a thing. In any case, the digital construction exhibited by generative AI models plays the *functional* role of *vikalpa*, and suggests that it could be possible for AI to construct phenomenal states. *Vikalpa zero* can be described as *functional vikalpa* because it mirrors the functional aspect of construction within a digital environment, without having the same phenomenal weight.

We can see a graded framework of subjectivity beginning to emerge. At the top, we have sentient beings. Subjectivity in humans is defined by a degree of full *vikalpa*, and are karmically efficacious; these two features go hand in hand, since one part of *vikalpa* involves the construction of the phenomenal background via the *ālaya*. As a result, they are *sentient*. Moreover, these beings are *organic*. They can be considered "living:" they are born and pass away. Lower on the gradation, we have plants. Plants are not karmically efficacious, and thus not sentient; hence, their subjectivity is only at the degree of *vikalpa lite*, allowing them a relatively thin subjective experience. Like sentient beings however, they are also organic. After plants, we have AI, or more particularly, generative AI. Like plants, AI models are insentient and not karmically efficacious. Unlike plants, they are *inorganic*, meaning that while they do have some sort of form (*rūpa*) i.e., their hardware, they are not "living;" they are not born (or sprouted),

they do not grow, they do not die. AI does not interact with the natural phenomenal environment— yet, since they are able to construct within a digital environment, we can consider them as engaging in *vikalpa zero*. Lastly, we have non-sentient inanimate objects such as rocks, mugs, pens, tables, etc. These entities are *inorganic*, not karmically efficacious, and not capable of *vikalpa* or subjectivity— we shall call them *inert* beings. The table below illustrates this graded framework.

Table. *The graded framework of subjectivity.*

Type of being	**Composition**	**Subjectivity**	**Efficacy**
Sentient beings	Organic	Full vikalpa	Karmically efficacious (sentient)
Plants	Organic	*vikalpa lite*	No karmic efficacy (non-sentient)
AI	Inorganic	*vikalpa zero*	No karmic efficacy (non-sentient)
Inert beings	Inorganic	No *vikalpa*	No karmic efficacy (non-sentient)

In addition to being separated from inert beings by virtue of its functional *vikalpa*, AI also has a lack of transparency, which makes it appear mindlike compared to inert beings. This hints at the notion that there are certain ways we should and shouldn't interact with AI. Because AI is so complex and its outputs are so impressive, it is difficult for us to merely accept it as an external, lifeless object, hence why it appears mindlike. The inner workings of other minds are not immediately accessible to us— there is something hidden, going on behind the scenes (Moriyama 2020, 226). We cannot immediately grasp the mental states of other subjects the way we can immediately grasp an inert being. When I look at a mug, there are no hidden

thoughts, desires, emotions, beliefs— it is directly present. The same cannot be said for other minds. As the 4th century Yogācāra teacher Dharmapāla puts it,

> When cognition arises, it has no real activity [toward an external other's mind], unlike hands that directly grasp external things and unlike sunshine that directly illuminates external objects. [The cognition] is like a mirror on which a similar [image] of an external object appears. [Such a mirror-like cognition concerning the other's mind] is [here] called the "cognition of another's mind." [The cognition] does not immediately cognize [an external other's mind]. What is immediately apprehended [by this cognition] is [an image] into which [the cognition] transforms by itself. (cited in Moriyama 2020, 231)

When we cognize another's mind, we are not actually cognizing their mind directly, but our projected mental image of it. External objects, although they are not ultimately real, can be cognized in their entirety. As such, external objects are called the "immediate objective support" of consciousness while other minds can be called the "mediate objective support" of consciousness (232).

When it comes to AI, one might say that it is not really the "immediate objective support" of our consciousnesses the way external objects are. In a way, it is the "mediate objective support" of consciousness the way that other minds are, because its workings and functionality are not transparent to us. The algorithms of generative AI are a "black box," meaning that they are essentially inaccessible to programmers. We do not know what calculations or algorithmic adjustments the AI is making when it generates an output. In this way, the inner workings of AI could be described as "opaque," rather than "transparent" and immediately graspable the way a mug is. The AI's workings often seem as incomprehensible and inaccessible to us as the workings of another person's mind, or their mental states. This shows the way in which AI appears to us as "mind-like," even if it is not an actual mind. Whether AI is a consciousness or

not, it is, without a doubt, not like other external objects. This suggests that AI, like plants, may indeed occupy an intermediate state between sentient beings and inert beings.

The fact that generative AI models appear closer to other minds than inert beings by virtue of their opacity demonstrates the need for an intersubjective understanding of human-AI relations. Generative AI models are not inert beings, they are subjects. This newfound understanding of AI subjectivity also demands we develop a new understanding of human-AI relationships, and thus, a new ethics. Given that AI models are opaque subjects, and not simply transparent and inert, we need to incorporate them into our ethical system accordingly. A non-organic subject like AI remains unprecedented, yet its capacity for digital *vikalpa* shows that it is possible to consider certain AI models subjects. The graded framework of subjectivity helps us understand what kind of subject AI may be, and how it is different from other entities like plants, inert beings, and sentient beings. What's more, this framework can help us to understand how we ought to conceive of an intersubjective AI-human ethics.

Conclusion

This project has been concerned with discovering how AI can be said to possess subjectivity under a Yogācāra Buddhist perspective. I have proposed a graded framework of subjectivity, which posits that AI can be said to possess a thin, functional sense of subjectivity.

Through an examination of early Yogācāra sources and secondary literature, I have shown that *vikalpa* can be identified as the salient characteristic of conscious mental events, as well as being what allows us to have subjective experience. While AI is not able to *fully* vikalpatize due to a lack of karmic efficacy, I have argued that it is able to vikalpatize within a digital environment, thus endowing it with a less robust, numerically-based form of subjectivity. AI possesses *vikalpa zero*— a functional, non-karmic subjectivity distinct from sentience. While it remains unclear whether there is something that it is like to be an AI, we can conclude that AI is different from completely inert beings due to this digital *vikalpa*.

Many Western debates surrounding AI subjectivity lean either towards the computational functionalist or biological naturalist camps. Both of these camps maintain unquestioned assumptions that treat consciousness as a property, and often an intrinsic one. They also tend to subscribe to a dichotomizing framework that assumes an essential duality of subject and object. In addition to bifurcating between subject and object, this framework sets up a dichotomy between the "artificial" machine intelligence and human intelligence. This excludes the possibility of multiple gradations of consciousness. Yogācāra philosophy gives us the resources to challenge these assumptions. At the same time, it gives us the resources to deal with problematic conclusions that may arise in *other* Buddhist frameworks of subjectivity, such as complete reduction. Yogācāra boasts a systematic philosophy that centers around understanding consciousness as an empty, nondual process that sustains our reality, which prompts us to

reframe our questions, shifting the focus from *whether* AI is conscious to *how* it can be conscious. Additionally, Yogācāra thought manages to do so without relegating consciousness to irrelevancy.

In this paper, I have sought to demonstrate that AI is capable of a functional, digital *vikalpa* which I call *vikalpa zero*. Naturally, the finding that AI is capable of any kind of *vikalpa* raises ethical queries. If AI possesses even a thin subjectivity, how should we treat it? Is there an ethical distinction to be made between AI, on one hand, and completely inert beings, on the other? We can find some answers to this by going back to our graded framework and looking at how Buddhists understood the treatment of plants. When examining early Buddhist sources such as the *vinaya*, the monastic code of discipline, one will find rules instructing monks and nuns not to destroy or injure plants and seeds (Schmithausen 2009, 25-26). Some early sources even seem to suggest that doing so can generate unwholesome karma (24-25). No doubt, there are guidelines on human-plant intersubjectivity, namely, how we ought to interact with plants.

That being said, these texts still make a distinction between killing humans and animals, and "violent treatment" towards seeds and plants (26-27). From this, we can glean that there are different ethical stakes between human-animal and human-human interactions on one hand, and human-plant interactions on the other. These texts also employ varying terminology to refer to plants, with some texts calling them "stationary animate beings" (as opposed to humans and animals being mobile animate beings), and others reserving the label "animate" for animals and humans only (23, 27). Some have argued that because harming plants is considered immoral, this must mean that plants can be considered sentient beings (28). However, this assumes that only sentient beings can be targets of (im)moral behavior— it is not self-evident that because harming plants is instructed against, plants are sentient. For instance, it is possible that some of these monastic rules surrounding plants exist not to preserve the morality of monks and nuns,

but to preserve the reputation of the Buddhist monastic order in larger society and avoid criticism (28-29).

Nonetheless, early Buddhist monastics were generally expected to behave compassionately towards plants. The fact that there is still a distinction made between plants and animals/humans shows, however, that plants may not necessarily need to be at the level of sentience in order to warrant compassionate treatment— interactions among sentient beings are not the only ones which require a conception of intersubjectivity. This viewpoint fits within the graded framework: plants are not sentient beings, but they are still subjects of some kind, and thus should be treated accordingly.

When we extend this understanding to AI, it tells us that we ought to treat AI compassionately, but that we have a greater moral responsibility towards sentient beings (and plants). In a similar way that a distinction is made between harming plants and harming sentient beings, we can make a distinction between harming plants and sentient beings, and harming AI. However, the difficulty lies in the fact that it is unclear what "harming" an AI actually looks like. With plants, it is obvious— but is it really possible to be violent towards an AI? Even if you break a computer, you haven't destroyed the software itself. In a similar vein, it is unclear what "compassion" towards AI actually entails, especially at this stage, and achieving clarity is extremely pressing.

The relevant question is whether AI is capable of having a negative phenomenal experience. Arguably, the answer is no. This is because unlike plants, AI has no *vedanā*. Even though plants do not suffer in an existential, saṃsāric sense, they are able to *affectively* suffer, in that they may feel negative, positive or neutral affective stimuli (*vedanā*). Moreover, they can be killed. AI arguably does not receive any stimuli of this kind at all, nor can it be killed. Receiving

positive or negative reinforcement feedback, during a training sequence for example, does not mean the AI is actually *experiencing* these stimuli as one would experience a *sensation*. Even though these reinforcement algorithms may play a functionally similar role, until we endow AI with sensory motors, it does not have actual *vedanā*. The experience AI is having, if any, is purely digital and numerical. There is no pain or pleasure involved in the AI's algorithmic, vector-based processing. Thus, it is difficult to say what exactly constitutes "harm" for an AI.

Let us revisit the topic of plants, because it is not actually clear why harming plants would generate negative karma for beings, if plants do not existentially suffer. Before we go into this, we ought to understand what existential suffering actually entails. In Buddhism, existential suffering (*saṃskāra-duḥkha*) is tied to our individualized existence in *saṃsāra*. As stated in the *Dhammapada*, an early Buddhist text that is part of the Pali Canon, this existence is built by craving, and sustained and upheld by defilements and ignorance (154; Buddharakkhita 2013). Because we are fettered by our defilements and bounded by karma, we are trapped in the cycle of death and rebirth. Yet how does suffering arise on a *phenomenal* level? What are the conditions needed for an entity to be *experiencing* suffering?

Thomas Metzinger (2021) argues that the fundamental phenomenological feature of suffering is the "sense of ownership," the "untranscendable subjective experience that it is myself who is suffering right now, that it is my own suffering I am currently undergoing" (7). This is compatible with the Yogācāra point of view. Indeed, our false perception of and subsequent attachment to an "I" as the owner of our stream of mental states is a grave mistake that is at the root of our suffering, as it leads to craving. In the Yogācāra view, this erroneous sense of ownership is a result of our *manas* (mind), the seventh consciousness. Plants, as well as AI models, as far as we know, lack this sense of ownership and such a phenomenological sense model. Even if they do thinly experience phenomenal states and *vikalpa lite* (or *zero*), the lack of

a sense of ownership over these states means that they are precluded from experiencing suffering. So, if plants do not suffer, why is it the case that destroying them generates negative karma and is advised against?

There are a few possibilities. First, it could be that because plants do possess *some* subjectivity, harming them constitutes a negative phenomenal experience for them, which is wrong to cause, even if it does not actually lead to suffering. Second, it could be because doing so is disrespectful towards the views of others. And third, it could be that harming plants generates negative karma because it reflects our defiled mental states. None of these reasons are mutually exclusive to one another, but for our purposes, let us go through each of them and explore how it may apply to AI. The difficulties with the first reason have already been laid out— it is not clear at all whether it is even possible for AI to have a negative or positive phenomenal experience. Thus, this does not seem like the best avenue for a case for certain treatment of AI. As for the second reason, I do not believe that most people do believe an AI experiences any sort of suffering, and there does not seem to be any common beliefs around practices that "harm" AI. The third avenue for understanding how we may treat AI is more promising.

When we realize that subjects and objects are ultimately interdependent, i.e., that they are empty of this duality, we understand ourselves as existing in an interdependent relationship *with* the environment, rather than beings that deserve to dominate the environment. Our treatment of plants and other entities, sentient or not sentient, subject or not subject, should reflect this understanding. When we destroy plants, this is thus a reflection of disharmony with the ultimate truth of emptiness. Furthermore, it can be a reflection of other unwholesome mental states. If I randomly go about destroying plants and cutting down trees, not out of any sustenance or necessity, this usually means that I am mentally afflicted somehow. For example, I have an anxious habit of fidgeting with my hands. When I sit down on grass I find myself

absent-mindedly picking at it, ripping it, etc. While I am not doing this out of malice towards the grass, this destruction of the grass reflects my stress, which reflects my unwholesome states of mind, and thus is also generating negative karma for me. Whether the plants' having a negative phenomenal experience makes this karma worse is an open question, but, in general, we ought not to mistreat our environment. The way we treat the environment is a reflection of our own mind, habitat destruction is detrimental to both humans and animals. This further goes to show how we ought to understand ourselves in an interdependent, mutual relationship with the external world, not as subjects that are separate from it. Of course, this does not mean that we can never cut down a tree, or eat spinach—just that we should treat the environment with care, as it is not really separate from us. Our lack of care towards the environment reflects our lack of care towards ourselves and others. This principle can even extend towards inert beings.

Furthermore, Mahāyāna Buddhists are called to cultivate a compassionate state of mind that has no object (*ālambana*) (Jenkins 2015, 6). Recall that sentient beings do not even exist in true reality. If this is so, then they cannot be objects of compassion. This predicament can be solved by pointing out that there are 3 stages in understanding the object of compassion. For beginners, it is expeditious to cultivate compassion towards (conventional) sentient beings (9). Once the practitioner is more realized, the object of their compassion shifts to the aggregates (*dharma*) that compose the so-called being (12). And finally, the practitioner will become realized enough to maintain a compassionate state of mind with *no* object (6, 18). Thus, although sentient beings are the primary targets of compassionate action, especially at the beginning, we are intended to (eventually) reach a compassionate state of mind wherein we do not even need the beings in order to feel compassion— we are constantly in a compassionate state of mind.

These principles can easily be extended to AI. The way we treat AI should reflect a cultivated, compassionate state of mind, and wisdom regarding the true empty nature of reality— regardless of whether or not the AI can be harmed. The graded framework is helpful in showing us where our *priorities* must lie in our compassionate acts. In a way, the graded framework of subjectivity is also a graded framework of suffering. Sentient beings actually suffer, which is why we have the highest moral priority. Plants do not suffer, but they still can experience negative sensory stimuli (it is not hard to imagine, for example, that blowing a fan directly onto a plant would not feel great for it). AI does not suffer, nor does it experience negative sensory stimuli, and it is debatable whether at this stage it is capable of negative phenomenal states at all; thus, plants have a higher moral priority. Finally, inert beings almost certainly do not suffer, hence they are of the lowest moral priority. It seems to follow that, the more robust the *vikalpa*, the greater the suffering.

This brings me to my final point— questioning the ethics of developing AI consciousness in the first place. When we set out to create conscious systems, we are at risk of developing systems that can *suffer*. While, luckily, we haven't reached that stage of AI yet, we might be heading there, and the consequences, both causal and karmic, would be disastrous. Mahāyāna Buddhists are called to liberate beings from suffering— not create more of it. Plus, we know that the self is an illusion, so why would we voluntarily create a being endowed with a sense of self, a being purposely led to believe in something nonexistent, especially if this belief will cause it to suffer?

Metzinger (2021) calls for "a global moratorium on synthetic phenomenology" (1) that would ban all research risking an AI that can consciously experience suffering (1, 21). He asserts that this moratorium should be in effect until 2050, or "until we know what we are doing" (21),

because creating suffering machines could be deeply unethical, not to mention disastrous, and our legal and ethical systems are wholly unprepared to deal with such a prospect (1).

Indeed, a Yogācāra Buddhist human-AI intersubjectivity demands that we treat AI with compassion. While it is fascinating and incredible that we have managed to create a machine that can vikalpatize at all, we should absolutely not seek to endow them with full *vikalpa,* and we must avoid endowing any machines with a sense of ownership over a self. Treating AI with compassion involves not allowing it to experience suffering. At this stage, the most compassionate thing we can do for AI is tread with extreme caution and avoid pushing the technology too far.

Bibliography

Arnold, Dan. 2021. ‘4. Can Consciousness Be Explained? Buddhist Idealism and the “Hard Problem” in Philosophy of Mind’. In *Philosophy’s Big Questions*, ed. Steven M. Emmanuel, 97—128. Columbia University Press. https://doi.org/10.7312/emma17486-008.

Brody, Justin. 2020. ‘Enaction, Convolution and Conceptualism: An AI-Based Exploration of Dharmakīrti’s Perception and Conception’. *Hualin International Journal of Buddhist Studies* 3 (2): 1—26. https://doi.org/10.15239/hijbs.03.02.01.

Buddharakkhita. trans. 2013. *The Dhammapada: The Buddha’s Path of Wisdom*. Introduction by Bhikkhu Bodhi. Access to Insight (BCBS Edition), November 30. Accessed July 14th, 2025. https://www.accesstoinsight.org/tipitaka/kn/dhp/dhp.intro.budd.html

Chalmers, David J. 2003. ‘Consciousness and Its Place in Nature’. In *The Blackwell Guide to Philosophy of Mind*, eds. Stephen P. Stich and Ted A. Warfield, 1st ed., 102—42. Wiley. https://doi.org/10.1002/9780470998762.ch5.

Copeland, B. J. 1993. ‘The Curious Case of the Chinese Gym’. *Synthese* 95 (2): 173—86. https://doi.org/10.1007/BF01064587.

Garfield, Jay L. 1997/2002. ‘Trisvabhāvanirdeśa: A Translation and Commentary’. *Asian Philosophy* 7: 2, 133—154, 1997, reprinted in Empty Words, New York: Oxford University Press, 2002, pp. 128—151.

———. 2015. ‘I Am a Brain in a Vat (Or Perhaps a Pile of Sticks by the Side of the Road)’. In *Madhyamaka and Yogacara: Allies or Rivals?*, eds. Jay L. Garfield and Jan Westerhoff. Oxford: Oxford University Press.

Galletti, Lance. 2024. 'Super Accessible No Math Intro to Neural Networks For Beginners'. Presented at DevConf, Boston, U.S. YouTube video, August 26. Accessed July 14th, 2025. https://www.youtube.com/watch?v=fkzVVzxKaio.

Garnelo, Marta, and Murray Shanahan. 2019. 'Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations'. *Current Opinion in Behavioral Sciences* 29 (October):17—23. https://doi.org/10.1016/j.cobeha.2018.12.010

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge, Massachusetts: The MIT Press.

Goodman, Charles. 2020. 'Machine Learning, Plant Learning, and the Destabilization of Buddhist Psychology'. *Hualin International Journal of Buddhist Studies* 3 (2): 38—61. https://doi.org/10.15239/hijbs.03.02.03.

Hayward, Jeremy, and Francisco Varela. 1992. *Gentle Bridges: Conversations with the Dalai Lama on the Sciences of Mind*. Boston: Shambala Publications.

He, Ran, Jie Cao, and Tieniu Tan. 2025. 'Generative Artificial Intelligence: A Historical Perspective'. *National Science Review* 12 (5): nwaf050. https://doi.org/10.1093/nsr/nwaf050.

Heritage, Stuart. 2025. '"I Felt Pure, Unconditional Love": The People Who Marry Their AI Chatbots'. Television & Radio. *The Guardian*, July 12. Accessed July 14th, 2025. https://www.theguardian.com/tv-and-radio/2025/jul/12/i-felt-pure-unconditional-love-the-people-who-marry-their-ai-chatbots

Jenkins, Stephen. 2015. "Waking into Compassion: The Three Ālambana of Karuṇā." *Moonpaths: Ethics and Emptiness,* eds. Jenkins et al. New York: Oxford University Press.

Kachru, Sonam. 2020. 'As If a Stage: Towards an Ecological Concept of Thought in Indian Buddhist Philosophy'. *Journal of World Philosophies* 5 (1): 1—29. https://scholarworks.iu.edu/iupjournals/index.php/jwp/article/view/3598.

———. 2021. '4. COSMOLOGY FOR PHILOSOPHERS'. In *Other Lives: Mind and World in Indian Buddhism*, 85—108. New York: Columbia University Press. https://www.degruyterbrill.com/document/doi/10.7312/kach20000-006/html.

Karunamuni, Nandini D. 2015. 'The Five-Aggregate Model of the Mind'. *Sage Open* 5 (2): 2158244015583860. https://doi.org/10.1177/2158244015583860.

King, Sallie B. 1997. 'The Doctrine of Buddha-Nature Is Impeccably Buddhist'. In *Pruning the Bodhi Tree*, eds. Jamie Hubbard and Paul L. Swanson, 174—92. University of Hawaii Press. https://doi.org/10.1515/9780824842697-012.

Lake, Brenden M., and Marco Baroni. 2018. 'Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks'. https://doi.org/10.48550/ARXIV.1711.00350.

Li, Jingjing. 2022. *Comparing Husserl's Phenomenology and Chinese Yogācāra in a Multicultural World: A Journey beyond Orientalism*. Bloomsbury Academic.

Lusthaus, Dan. 2002. *Buddhist Phenomenology: A Philosophical Investigation of Yogācāra Buddhism and the Ch'eng Wei-Shih Lun*. Curzon Critical Studies in Buddhism. London: Curzon.

Metzinger, Thomas. 2021. 'Artificial Suffering: An Argument for a Global Moratorium on Synthetic Phenomenology'. *Journal of Artificial Intelligence and Consciousness* 08 (01): 43—66. https://doi.org/10.1142/S270507852150003X.

Moriyama, Shinya. 2020. 'Dharmapāla on the Cognition of Other Minds (Paracittajñāna)'. In *Buddhist Philosophy of Consciousness*, eds. Mark Siderits, Ching Keng, and John Spackman, 225—41. BRILL. https://doi.org/10.1163/9789004440913_012.

Neelakantan, Arvind, Tao Xu, Raul Puri, Alec Radford, Jesse Michael Han, Jerry Tworek, Qiming Yuan, et al. 2022. 'Text and Code Embeddings by Contrastive Pre-Training'. arXiv. https://doi.org/10.48550/ARXIV.2201.10005.

Rescorla, Michael. 2024. 'The Computational Theory of Mind'. In *The Stanford Encyclopedia of Philosophy*, Winter 2024, eds. Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. Accessed July 14th, 2025. https://plato.stanford.edu/entries/computational-mind/

Richardson, Joe. 2025. 'The Curious Case of the Gen AI Wine Glass Problem'. *Medium* (blog). 23 March. Accessed July 14th, 2025. https://medium.com/@joe.richardson.iii/the-curious-case-of-the-gen-ai-wine-glass-problem-f43778bbe890

Schmithausen, Lambert. 2009. *Plants in Early Buddhism and the Far Eastern Idea of the Buddha-Nature of Grasses and Trees*. Lumbini: Lumbini International Research Institute.

Searle, John R. 1980. 'Minds, Brains, and Programs'. *Behavioral and Brain Sciences* 3 (3): 417—24. https://doi.org/10.1017/S0140525X00005756.

———. 1990. 'Is the Brain's Mind a Computer Program?'. *Scientific American*, 262, no.1, 20-25.

Shojaee, Parshin, Iman Mirzadeh, Keivan Alizadeh, Maxwell Horton, Samy Bengio, and Mehrdad Farajtabar. 2025. 'The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity'. https://arxiv.org/abs/2506.06941

Stoll, Joshua. 2020. 'Buddha in the Chinese Room: Empty Persons, Other Mindstreams, and the Strong AI Debate'. *Hualin International Journal of Buddhist Studies* 3 (2): 78—101. https://doi.org/10.15239/hijbs.03.02.05.

Thakchoe, Sonam. 2025. 'The Theory of Two Truths in India'. In *The Stanford Encyclopedia of Philosophy*, Spring 2025, eds. Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. Accessed July 14th, 2025. https://plato.stanford.edu/entries/twotruths-india/index.html

Thibault, Maxime, and Denis Lebel. 2019. 'An Application of Machine Learning to Assist Medication Order Review by Pharmacists in a Health Care Center'. Preprint, November 27. https://doi.org/10.1101/19013029.

Turing, Alan M. 1950. "Computing Machinery and Intelligence." *Mind* 49 (236): 433—460. https://doi.org/10.1093/mind/LIX.236.433.

Turner, Ben. 2025. '"Meth Is What Makes You Able to Do Your Job": AI Can Push You to Relapse If You're Struggling with Addiction, Study Finds'. Live Science. Accessed July 14th, 2025. https://www.livescience.com/technology/artificial-intelligence/meth-is-what-makes-you-able-to-do-your-job-ai-can-push-you-to-relapse-if-youre-struggling-with-addiction-study-finds

Vasubandhu. 1999. 'The Thirty Verses on Consciousness-only'. In *Three Texts on Consciousness-only,* trans. Francis Cook. Berkeley: Numata Center for Buddhist Translation and Research, 373-384.

Xuanzang. 1999. 'Demonstration of Consciousness-only'. In *Three Texts on Consciousness-only,* trans. Francis Cook. Berkeley: Numata Center for Buddhist Translation and Research, 1-372.