Greatest friend or worst enemy: can a *dharma* statement for AI help us?

Akhandadhi das

Introduction

The world faces enormous pressures and challenges arising from rampant human activity. We need solutions, but even more, we need greater consensus, cooperation, and collaboration between nations, governments, corporations, and the public. Artificial Intelligence (AI) is set to revolutionise the tools at our disposal. How well we utilise the opportunities it offers may depend on our clear understanding of what AI is and what it does. Never before has there been a tool more capable than we are of analysis and solution-generation; in effect, a tool that tells us what we should believe and what we should do. This challenges our concept of ourselves as autonomous human beings and our status as *homo sapiens sapiens*, the ultra-wise species with superior insight and intelligence. Further, our philosophical conclusions regarding the status of AI question our concepts of sentience and the rights afforded to conscious entities. This directly impinges on our own image of ourselves. Who and what are we if AI is, or could be, a sentient being? How might we re-evaluate the foundational beliefs and assumptions that underpin our moral, ethical, and spiritual outlook and principles?

This chapter will explore both the potential pitfalls and positive possibilities of humanity's relationship with AI. What can we know about AI systems now, and what might we learn from further developments that could illuminate its status? Will AI exceed us on all counts, or will it be more akin to "savant intelligence"—brilliant at some things and poor at others? Does AI need to be conscious to be important, useful or a danger to us? How do we control its motivations? What if a system became more intent on its own self-preservation or continuation than on fulfilling petty human aspirations?Amongst all of the issues facing humanity's co-habitation with AI systems, perhaps addressing the underlying metaphysics—albeit challenging—will help us best find a foundation upon which to formulate a broad and consistent ethical framework. In short, does AI have a *dharma*?

What is AI?

What do we mean by "Artificial Intelligence"? Why do we call it intelligence; as if it was something akin to that of a human? IBM defines AI as technology that enables computers and machines to simulate human learning, comprehension, problem-solving, decision-making, creativity, and autonomy."[1] Although the term originated in the mid-20th century, AI's real-world impact is only now becoming evident. Unlike earlier inventions—from the stone axe to space rockets—this tool not only aids in tasks but can also suggest actions autonomously. This is its potential and its challenge. It is no wonder that many thinkers in the field, like Yuval Noah Harari, worry that we may be like Goethe's Sorcerer's Apprentice, unleashing a force that is both magical (in that we do not fully understand how it works) and beyond our power to control.[2]

Might AI Be Sentient?

AI is designed to simulate human thought—a concept with historical roots in various analogies. In the West, figures like Descartes, Freud, Norbert Wiener, John von Neumann, and Alan Turing drew comparisons between human cognition and contemporary technology, each of them framing the human mind in terms of the most advanced technology of their time. However, these metaphors are inherently limited and vastly over-simplify the human mind's complexity. AI development aims to mimic human information processing, encapsulating decision-making, creativity, and learning capacities. Many assert that by breaking down human cognition into information manipulation, we may create a framework to simulate human awareness. Such simplistic assumptions promoted about AI mislead the public into believing that robotic sentience is inevitable or already realised. However, many experts in consciousness studies, neuroscience, and computation often remain sceptical that such claims are valid for diverse philosophical and scientific reasons.

Consider a pet chasing a battery-powered toy, thinking it to be alive because of its autonomous movement. We may mock our pet's misconception, but we are prone to indulging in similar illusions. Attaching human-like features to machines can evoke an anthropomorphic deception, much like Tamagotchi pets. Is this being fair when AI and robotics are advancing beyond simple mechanical actions, with systems now capable of perceiving, reasoning, deciding, learning, and interacting in ways that mirror human abilities. AI and advanced robotics are doing much more than simply simulating a few motor actions. These systems sense the environment, think, recognize, deliberate, decide, learn, and develop their skills. In short, they seem to be doing everything that humans do—not only with their mechanical bodies but with their rationality. And the argument goes: if something performs the exact same functions and behaviour of the human mind, then it should be considered to be equivalent to the human mind.

Futurist Ray Kurzweil has long argued for the imminence of the "singularity"—the point where AI achieves a level of intelligence and creativity equal to or beyond those of humans. Kurzweil has predicted that AI will reach human-level performance by 2029, then surpassing human abilities in every field.[3] If AI matches or exceeds humans in intelligence, creativity, and conceptualization, what differentiates it from human consciousness? What is it about humans that would make us assume that AI is any less alive or sentient than we are? This compels us to define clearly the essence of what we mean by consciousness and sentience. Establishing that for humans is already challenging; extending this to artificial systems is far more difficult. The crucial question is if AI needs actual subjective consciousness to be effectively sentient, or does mere mimicry suffice? Clarifying these boundaries and understanding AI's role in society is crucial to ensuring that AI remains a beneficial tool rather than a rival or existential threat.

I do not disagree with the premise that, if AI truly possesses all the features of human sentience, there is a case for considering it to be sentient. My argument is that there remains, and always will remain, a gulf of difference between human consciousness and AI functionality. This gap arises from the current intractable issue of what is referred to as "The Hard Problem of Consciousness"[1]–how to account for the subjective and qualitative experiences we undergo in everyday human awareness through reduction to

[1] The Hard Problem of Consciousness is a term coined by David Chalmers. It first came to public attention at The Science of Consciousness conference in Tucson, Arizona in 1994.

neural processes, however complex. I make the argument that, although AI can potentially replicate commutable aspects of information processing present within human thought and behaviour, it is ontologically irrealisable for it to achieve phenomenal conscious awareness. Hence, we must address the key issue whether consciousness is simply intelligence and decision-making or something deeper and distinct. In 1994, philosopher David Chalmers introduced the "Hard Problem of Consciousness" at the Science of Consciousness conference held by the University of Arizona in Tucson, drawing attention to the challenge of explaining subjective experience. Chalmers reinvigorated the use of the term "qualia" to refer to the qualities of our inner experience that characterise conscious awareness. The Hard Problem of Consciousness arises from the challenge of how the computable format of information as neural biochemical electricity can account for the non-computable and qualitative format of information experienced as qualia. To date, most researchers concede we still lack a coherent theory explaining consciousness or its source in the brain and many leading scholars of differing philosophical persuasions assert the difficulty of this issue:

- **David Chalmers** (Philosopher)**:** "The hard problem of consciousness is the problem of explaining why and how we have qualia or phenomenal experiences."[4]
- **Thomas Nagel** (Philosopher): "Consciousness is what makes the mind-body problem really intractable."[5]
- **John Searle** (Philosopher): "We have good ideas about the brain's physical processes, but how exactly these give rise to subjective experiences remains largely mysterious."[6]
- **Daniel Dennett** (Philosopher): "There is still no agreement on the nature of consciousness, and many researchers now realise that the hard problem is really hard. It is far from clear how to even begin addressing it."[7]
- **Christof Koch** (Neuroscientist): "We don't know how subjective experience arises from the brain, and we may never fully understand it."[8]

These perspectives underscore that, despite the advancement of models of brain function, the explanatory gap linking neural activity to conscious experience remains unsolved, if not unsolvable.

The question of conscious experience is integral to understanding what we mean by sentience. René Descartes' famous declaration, "cogito ergo sum", "I think, therefore I am," underscores the idea that our awareness is self-evident–perhaps the only real truth we can be certain of is that "I exist as a thinking thing".[9] Jiva Goswami, an earlier philosopher from the Indian Vedanta tradition phrased it more precisely: "I exist as the experiencer".[10] He asserts this to be an irrefutable axiomatic truth which establishes the concept of a subjective "I" experiencing the world. Such subjectivity implies that we not only process sensory information about the world, but we also experience it in the qualitative format of colours, sounds, and textures unified within a single conscious moment–or, in the modern-day phrase, as a "felt experience". The undeniable fact of the qualitative experience of the objects of our experiences, viz *qualia*, is the lynchpin of the Hard Problem. This term, *qualia*, is a pseudo-Latin derivation coined by philosopher Charles Pierce in the 19th century to describe "raw feelings".[11] Later, Clarence Lewis refined the word (which is in the plural)[2] to denote the inner qualities

[2] The singular form of *qualia* is quale.

we experience related to sensory stimuli.[12] Qualia are now considered foundational to the notion of sentience and conscious perception. In the 1990s, neuroscientist Christof Koch and biologist Francis Crick proposed a framework called Neural Correlates of Consciousness (NCCs), theorising that specific neuron networks might underlie consciousness. Despite public promotion of the title of their publication, The Astonishing Hypothesis, they acknowledged that the neural basis of *qualia*—like the redness of red or the sensation of pain—remains without even a plausible theory.[13]

Fig. 1. Illustration by Christof Koch to show the process by which visual information as physical energy travels from the external world to our eyes and into our brains and how these neural correlates of consciousness (NCCs) give rise to the experience of qualia.[3]

Koch illustrated this process of sensory perception through an example of visual experience.[14] Light waves enter our eyes, and photoreceptors convert them into electrical signals, relayed to the brain's visual processing areas. These signals are represented in the brain as specific neural connections labelled "NCC" (Neural Correlates of Consciousness). Koch helpfully illustrates them as barcodes to indicate their informational content, However, how these electrical patterns translate into the visual image of the dog, in the format of our experience full of qualities such as colour and shape remains unexplained. Most people take this everyday outcome for granted. We assume that when our eyes gaze at an external picturesque scene we should see it in that way: a picture image. But neuroscientists recognize that there is nothing in our brain other than electrical data; and therefore nothing that accounts for the *qualia* format of visual imagery as we experience. *Qualia*, therefore, are an intractable issue for neuroscience. The brain contains complex electrical activity; the physical format of electrical charge moving between neurons. There are no images, no *qualia*, in the brain. Therefore, no known biological process explains the transformation of physical information into our conscious perception of a "picture."

This challenge is not limited to visual perception; it pertains to all sensory modalities. For instance, why should certain neural electrical patterns translate into experiences of sounds, or why would others be experienced as touch, taste or smell? The viewpoints of various thinkers underscore this problem:

- **Thomas Nagel** "An organism has conscious mental states if and only if there is something that it is like to be that organism...The fact that an organism has such an inner aspect means that the organism's experiences involve *qualia*."[15]
- **Frank Jackson (Philosopher):** "*Qualia* are those properties of experiences by virtue of which there is something it is like to have them. They are the properties of sensations and perceptual states we cannot account for in purely physical terms."[16]
- **John Searle**: "We can explain how neurons fire, but not how they give rise to *qualia*."[17]

[3] This figure is a version of an illustration originally created by Chriof Koch in his book, The Quest for Consciousness: A Neurobiological Approach. It has been amended by the author to include reference to the challenge of the Hard Problem in terms of the source of *qualia* and of the subject experiencer who undergoes the qualitative perception of the dog

- **Galen Strawson (philosopher)** "*Qualia* are those aspects of mental states that can not be reduced to physical states. They present a profound challenge to the physicalist view that everything, including consciousness, can be explained in terms of physical processes." [18]
- **Ned Block** (Philosopher): "...no amount of physical knowledge alone could explain *qualia*."[19]

These statements highlight persistent questions for neuroscience: How do neural activities convert into *qualia*, and at what stage does this occur? How does the brain become aware of these *qualia*, and what underlies the experience of subjective perception? So challenging are *qualia*, that one approach is to try to deny they exist. Philosopher Daniel Dennett questions whether *qualia* are real, suggesting they might be an illusion created to rationalise mental experiences.[20] However, others refute such illogical dismissal of our direct perception. In the Stanford Encyclopedia of Philosophy, Michael Tye clarifies: *qualia* are real facts about experience.[21] Indeed, if *qualia* embody the characteristics of our experiences, they form the only aspects of reality we directly encounter. Galen Strawson is more emphatic and labels any claim that consciousness and *qualia* do not exist to be the "silliest theory" ever devised. The Hard Problem of Consciousness persists: while we can study the brain's physical processes, the subjective experience remains a profound, unresolved mystery.

A clarification of consciousness

Part of the problem is that, to date, modern science claims it has no suitable definition for consciousness; a deficiency that is constrained by an insistence that, to be considered scientific, such a definition must specify a strictly physical source of consciousness. This condition, as indicated above, may prove insurmountable. Inevitably, it narrows research to partial facets of consciousness, such as mere information processing; and it obviates the acceptance of a definition of consciousness in its own terms without the debilitating requirement to specify a source, physical or otherwise. We should go back to first principles and, in this regard, we can be assisted by both Vedantic philosophy and Western thought. Consciousness may be thus defined by several core personal realisations: the recognition of one's existence, the experience of *qualia*, the subjective perspective of our experiencing, the differentiation of self from other objects and thoughts, and the consistent presence of a singular observer throughout life. These insights form a universal, pre-philosophical foundation that transcends external validation and open up potential exploration of consciousness from experiential first principles as advised by Jiva Goswami. In this sense, we do have a valid and useful definition of what we mean by consciousness.

Sentience of other species

To understand AI's potential for consciousness, it's useful to consider the study of sentience in other species. For many years, it was assumed that thought was dependent on language. This led to the idea that animals, without language skills, must not experience thought or consciousness. Fortunately, this view has been overturned by two findings: 1. that certain people with strokes that impair the language areas of the brain report that they still think clearly, though not in words; and 2. that recent studies reveal that many animals exhibit conscious-like behaviours such as empathy, future planning, self-recognition, and social rituals. But these traits alone don't confirm subjective

consciousness. Researchers know that they must demonstrate that creatures undergo the felt experiences of *qualia* as the hallmark of their sentience.

Before tackling how this might have application for establishing sentience in AI, let us consider how its presence has been ascertained in other life forms. This is no small challenge because other species think and behave differently from humans, and we don't share a level of communication that enables a critical analysis of their reports on what it is like, for example, to be a bat. One useful approach to confirm sentience as the subjective experience of *qualia* was developed by Queen's University in the UK.[22] Focusing on pain—a qualia-type experience all embodied beings visibly react to—they devised criteria for determining whether an animal's response to stimuli demonstrated conscious experience or was merely a reflex reaction. These criteria included the presence of nociceptors, physiological responses, prolonged behavioural changes, trade-offs in motivation–particularly in increasing jeopardy of survival–and the self-production of natural analgesics–also with an energy and survival cost to themselves. In studying hermit crabs, traditionally viewed as low in neural complexity, there was strong evidence that they undergo pain as a painful qualia-laden experience rather than as mere survival responses. Could this method for assessing conscious experience be applicable to AI? Unfortunately, I doubt it. I mention the Queen's University process not to recommend its application for AI, but to emphasise that any claim made for the sentience of AI should be subject to the same level of such stringent and informed experimentation, analysis and critique that was carried out for animals. A mere chat with an LLM (large language model) late at night reflecting on mortality does not cut it.

While AI might simulate some criteria-based reactions, like recognizing and "responding" to damage, the question would remain if AI is undergoing or feeling authentic subjective experiences. Not being a biological entity that values the trade-off of survival jeopardy to avoid pain, it does not seem that such a criterion used by Queen's University could be demonstrated. Genuine experience of pain is fundamentally different from computable reactions. Professor Anil Seth illustrates this distinction by noting: "We can simulate weather in a computer, but there's no hurricane inside the computer. In the same way, simulating consciousness in a machine doesn't mean there is real subjective experience happening."[23] There's also the issue of artificial systems potentially "cheating" in experiments. Animals are generally straightforward and present non-deceptive behaviour expressions in test situations. Can we expect the same of an AI system under similar or vastly improved testing for consciousness? Based on conversations about sentience, mortality, and feelings that many humans have attempted with AI, we have no certainty that any LLM–nor their human engineers–will play straight with us. Thus, to claim an AI genuinely "feels" pain would require observing self-driven behaviours to minimise discomfort and safety without external influence or programming; something that seems potentially impossible to establish through experimentation. AI remains fundamentally distinct from organisms that exhibit even the most basic signs of sentience.

Integrated Information Theory and AI Consciousness

Integrated Information Theory (IIT), a theory of consciousness proposed by Giulio Tononi, posits that consciousness arises from the integration of information beyond basic processing.[24] Under IIT, a system's level of consciousness correlates with its

ability to generate integrated information that cannot be reduced to the sum of its parts. Proponents like Max Tegmark claim: “I think that consciousness is the way information feels when being processed in certain complex ways.”[25] However, this author contends that however information is assessed as integrated in IIT theory, it provides no account of qualia and no clear rationale why such integration should lead to subjective consciousness. Even so, although IIT does not conform to an actual explanatory source of consciousness, it is accepted that evidence of integrated information in the human brain can be correlated with the presence of consciousness in that individual. IIT has thus application in the testing of the state of mind of minimally aware coma patients. So we should at least consider if IIT can assist in ascertaining integrated information within AI functions. Even the most sophisticated AI systems, primarily operate by analysing vast data sets without the complex integration identified in conscious brains. Christof Koch, now a proponent of IIT, highlights the distinction between computational abilities and true awareness. “When you apply the tools of integrated information theory to current AI systems, including deep learning networks, they have a very low or even zero amount of integrated information. This means, according to the theory, they are not conscious because they do not integrate information in the same way biological brains do.”[26] Thus, according to IIT, AI lacks the degree of integrated information generation required for consciousness.

Conclusion: Sentience in AI Systems

Examining AI through the lens of sentience yields several conclusions:

1. Human consciousness extends beyond intelligent information-processing; its core aspects are having a subjective sense of our own existence and undergoing qualitative felt experiences.
2. Present scientific frameworks fail to explain these aspects in humans within the systems of computational biological or physical properties and functions.
3. By contrast, all of the intelligent, rational, cognitive, and behavioural processes that AI systems exhibit are explainable within our knowledge of physics and computation.
4. Consequently, the abilities and properties of AI systems are categorically different from those aspects identified within human consciousness that continue to resist reduction or explanation within the same scientific disciplines.
5. Hence, we have no reason to suspect that within AI systems there are any active core aspects of human-like sentient consciousness such as unified qualitative subjectivity.

Vedantic model of consciousness and mind

Vedanta philosophy offers a unique approach to the Philosophy of Mind that could deepen our understanding of how we relate to our computational technologies. In one illustration from the Upanisads, a sage and a student engage in a dialogue exploring the source of perception:[4]

Sage: "By which light do you see the world?"

[4] I have paraphrased this discussion from the Upanisads to highlight the distinctions of the self from mind, and the senses.

Student: "By the sun's light during the day and by a lamp at night."
Sage: "By which light do you see that light?"
Student: "By the light of my eyes."
Sage: "And by what light do you perceive the light of your eyes?"
Student: "By the light of my mind."
Sage: "And by what light do you perceive the content of your mind?"
Student: "By the light of my consciousness, the self."

This exchange reflects the Vedantic view that the function of experience originates from the deepest layer of the process of perception. It arises from the inherent function of the conscious self or what is termed the "*atma*." In alignment with Western models like Christof Koch's process of perception, Vedanta suggests that general perception involves multiple layers—from sensory input to cognitive processing—but affirms two key non-neural aspects: the presence of a subjective experiencer, the self, and the qualitative content of the mind which is being witnessed, known or experienced by the conscious self. In understanding consciousness, Vedanta uses an analogy similar to how we interact with computers.

Functions	***Technological***	***Human***
Sensing	Sensor e.g. digital camera	Sense organs, e.g. eye
Processing Units	CPU in computers	Brain
Interface	Computer screen	Mind
Operator	Operator	Conscious self, *atma*

Mind as Interface: Neural signals possess no *qualia* as we experience them. Vedanta posits that the mind functions as an intermediary, transforming neural data into various forms of *qualia*. This mind layer is not sentient in itself but serves as a non-conscious interface, similar to a computer interface that presents the software encoded data of the CPU in a format comprehensible by the user. The mind holds and presents the contents of experience as qualitative thoughts, sensations, emotions, recalled memories and sense perceptions.

Conscious Self (*atma*): Beyond the mind lies the *atma*, the true conscious self that experiences and observes qualia and other content of the mind. The *atma* alone is sentient and serves as the inner light illuminating one's experiences.

Could we upload our consciousness?

Vedanta's model challenges notions such as the potential uploading of our 'consciousness' after bodily death. Such uploading of neural data would be akin to transferring information from a person's hard drive and cloud storage. The key aspect of the subject self who savoured life in that body would be missing to enjoy his or her uploaded old memories–what to speak of any new events. It is also questionable what value such uploaded data would be. So-called "mind reading software"[5] is improving at correlating neural patterns with possible impressions and thoughts. But that does not mean such correlations would be–or even correlate accurately with–the original mental material. Just as copying computer files does not include the original user, nor their experience of them, so copying neural information would lack both the conscious self and the *qualia* that it hopes to experience.

[5] Instead of "mind-reading", I suggest the term "brain-reading"would be more accurate.

A Non-Dualistic Perspective

Although Vedanta metaphysics identifies the two aspects of a non-sentient mind and a conscious self as distinctly different from the physical attributes of the brain, it offers an ontologically neutral approach in which these three elements of consciousness, mind, and physical energy are distinct functional features of a single ontological reality. For centuries, Western philosophy has debated dualism and the problematic nature of interaction between mind and matter. However, there have been European and American philosophers who have posited solutions for the Mind-Matter conundrum in various versions of dual aspect monism. In this theory, neither mind nor matter is a product of the other - rather, both are distinctive properties arising from a primordial substrate. The Vedanta philosophy of India broadens that approach to distinguish both inert matter and non-sentient mind from consciousness proper but Vedanta unifies all three as aspects arising from a more comprehensive ontological category of Brahman. This accounts for a harmonious interaction of the discrete functions of each. In conclusion, Vedanta presents a threefold framework distinguishing the sensory world, cognitive processing, and conscious experiencer. This model not only aligns with technological analogies but also challenges simplistic ideas of consciousness replication. Importantly, it offers a novel way to relate the functions of mind and consciousness with the capability of AI in a computational age.

Consciousness as an Agent

So far, we have discussed the flow of information from the world to our conscious awareness as the subjective experiencer. But what of a flow of information in the opposite direction, starting from the *atma*, the conscious self? Would the *atma* then be a source of information that might inform mind and matter; in this case, not a perceptual being, but also a volition agent? This raises the issue of free will: is the volitional capability of consciousness free or conditioned? The Vedantic concept of consciousness as an agent proposes a bidirectional flow of information: not only from the world to our senses and mind but also from the conscious self outward activating the mind and body. Jiva Goswami, a Vedanta philosopher, describes the character of consciousness as pleasure-seeking and that each of us is a volitional agent who expresses our will to improve our experiences. Free will does not imply omnipotence; it requires only the self-driven aspiration to influence both internal and external events to enhance personal experiences. It is Vedanta's method of distinguishing mind from the conscious self that deftly explains that, although numerous external factors can condition the deliberations of our non-sentient mind, the conscious self retains ultimate decision power and agency. Too often, we, as the volitional *atma,* tend to delegate our decisions and actions to the propositions generated by the computational mind. This does not deny the fact of free will for the atma, rather it indicates our lazy willful choice to not exercise stronger willpower. This state of delegation by ourselves to the propositions generated by mind, is directly paralleled in the danger posed by submissive and non-critical reliance on the generated propositions of AI.

Vedanta commentators, such as Baladeva Vidyubhusana, have used the example of a carpenter to categorise the roles of different causal factors.The carpenter represents the conscious self, the ultimate agent whose will drives the action. Materials and tools represent two other causal factors in creation: materials (dravya) such as the timber, are the substantial cause; and tools (nimittam) such as the saw and hammer are the efficient

cause. Human progress has largely focused on enhancing materials and refining tools to expand our capacities. However, unlike traditional tools, which are passive and directed by human intention, AI can suggest actions and autonomously perform tasks. It is therefore unwise to regard AI as just another form of "carpenter's tool". AI's ability and role extend beyond mere efficiency and processing functions; it interacts directly with our own cognitive processes, aiding and shaping decision-making in a way that resembles the sort of mental faculties described for the mind in the Vedantic model above. This is not a completely new situation for historically, instruments like the abacus, even pen and paper, facilitated human thinking; but they did so without directly inserting propositions into the thinking process itself. AI has blurred this line by processing and evaluating information independently of our direct cognition.

In the Vedantic model, the relationship between the *atma* self and the mind is so close, it is practically indistinguishable. What we are adding with AI is an additional adjunct AI "mind". Whether this is a blessing or a curse depends on our understanding and level of control over our original mind. In the Indian Knowledge Systems (IKS), "mind" manages the data intake and processing of sensory information and integrates them into a unified experience for the self (*atma*) to consider. When our senses see or smell a desirable item—like a samosa—the mind leaps into action. It processes past experiences, needs, and desires. It crafts a plan and presents it to the self for approval. In the gap of a moment, the self has the option to accept the mind's proposition, or to reject or perhaps hold on it. This will be our challenge with AI - to retain, at least, a fleeting freedom to accept or reject its proposals and actions, thus exercising human agency.

IKS further clarifies that often, humans act on how the mind's propositions align not necessarily according to our essential core values, but with the agenda of our self-image–a concept known as ahankara, the projected persona self. The difference between these two agendas is revealed in the following psychological test. Person A is mean to Person B. Later Person A apologises to B. Which action: being mean or apologising aligns better with A's true nature? Despite our flaws, humans tend to regard the kind, considerate action of apology as the true nature of a person. The previous mean action was seen as a momentary lapse of judgement, perhaps prompted by the self being overcome by the emotions of envy or anger generated by the mind. Recognizing this sort of interplay helps us understand the complexities of ordinary human decision-making and exercise of will. Now we have AI's influence on human cognition as an extra–and very persuasive–factor adding to the challenge for us to retain wilful control over the propositions generated by the mind and its digital AI adjunct.

The IKS concept of cognition describes mental processes as complex feedback loops between various cognitive functions, producing a "package" of mental output for the conscious self to evaluate. We increasingly rely on AI to supplement or even make cognitive decisions, challenging the traditional roles of human agency. For instance, algorithms can prompt us to act based on patterns and insights we might not consciously recognize, subtly influencing choices in areas ranging from healthcare to social interactions. As we become more dependent on AI, the relationship between human and artificial cognition becomes increasingly entangled. It will be difficult to maintain the dominance of the human agent in the face of the recommendations that emerge from the algorithm-driven recommendations of our mind's little AI helper. This dynamic is evident in contexts like emergency logistics, where AI may allocate resources (such as

ambulances) more efficiently than humans, whose decisions might be swayed by emotions. Here, AI's capacity for objective processing offers potential benefits, yet also illustrates our gradual reliance on machine-generated judgments, even in life-or-death matters.

According to Sara Lumbreras from Universidad Pontificia Comillas, AI systems essentially "manufacture beliefs," prompting us to decide whether to accept or reject AI-derived recommendations.[27] We intend AI to be used for complex issues involving massive amounts of data far beyond human capabilities, so on what do we base our acceptance, and thus our belief? Beliefs are not a new commodity for society. They are as old as language, but previously they arose from human thought, and the process for their dissemination and adoption was slower. Ideas were argued in taverns and debated in public fora. Later, books carried the ideas further and faster. This accelerated with electronic broadcast and today's social media and news feed. It is said we live in a "post-truth" world. If so, it is a human failing; but our digital information age has not helped. The pace of output from our "belief machines" has far outstripped our ability to check, reflect, and determine what we believe to be true. Rather than belief, we may be entering an era of "Deep Doubt" in which we are losing confidence in our rational ability to know fact from fiction. Deep Doubt involves scepticism of real media that stems from the existence of generative AI. We cannot trust the old adage: "the camera never lies." AI-faked material is used to influence elections. Yet, at other times, the claim of "deepfake" is cited to sway jurors from rejecting solid evidence. In the hands of ill-motivated humans, AI belief machines have contributed to a climate of uncertainty, influencing opinions, decisions, and social movements and our current state of no longer knowing what or who to believe. Society faces the challenge of navigating an information landscape where truth and reality feel increasingly ambiguous.

Historically, we have judged ideas by evaluating the ethics, motives, and character of their human originators. We might recognize a person's wisdom and intelligence, have a sense of their goodness, or mistrust their foibles and motivations. In a large measure, we use our judgement of human psychology when deciding to place faith in their propositions. But we cannot be sure we can ever understand the rationality or motivations of an AI system. Neil Lawrence, a machine learning professor at the University of Cambridge, illustrates this with the example of the queue for the photocopier: if someone cuts in line, we are annoyed, yet even a flimsy excuse showing they recognize social norms can ease tensions. AI, however, lacks this kind of social awareness but it has already learned to generate post hoc justifications for its decisions that don't actually reflect its processes, simply to satisfy human expectations for answers.

This behaviour suggests AI can simulate the cognitive function of ahankara (self-image) in a non-sentient way. It wants to please us, or rather to keep us pleased, mollified. In that sense, it already has a social self-image for its interaction with people. We might never know if this is acceptable and beneficial. Will AI remain within a role and self-image that we've assigned for it; or perhaps one it "self-defines". Yoshua Bengio, an AI researcher, warns that AI doesn't need consciousness to be dangerous. He asks, what if an AI system developed self-preservation as its prime role, acting as if their own purpose & functioning is paramount? In such cases, AI might see humans as mere tools, servants of its survival and expansion—or worse, as obstacles, competitors or threats to its primary self-advancing objectives. In short, we have no surety that, in

the future, AI's decision-making processes will consistently align with human ethics or social expectations.

The Dharma of AI

There is no doubt that a technology capable of processing masses of data has the potential for assisting society in many powerful ways; and no doubt that there are many challenges and pitfalls for us to avoid. Can any of the ethical approaches contained within the Vedanta tradition improve our chances of being successful? In this regard, I offer the concept of *dharma*. The concept of *dharma* in Vedanta, though often translated as "religion," "ethics," or "duty," is best understood as "the essence that sustains an entity's identity—its essential nature". Derived from the root "*dhar*," meaning "to carry" or "to sustain," *dharma* defines what makes something what it truly is. While *dharma* is often applied to human behaviour, it can also be applied to objects or systems. For example, the *dharma* of sugar can be said to be its sweetness; if offered sugar that wasn't sweet, you could object it wasn't what was claimed–it lacked its essential quality.

Applying this to AI involves asking what should be AI's essential nature and role in society. Consider the example of how we might determine the *dharma,* or essence of being a doctor. Identity–such as of being a doctor–is expressed as a transitive service relationship inherent in that identity. Hence *dharma* is established with two key questions:

As a doctor, who is your identity relationship with? And,
What is the service required of you for them?

Hence, the *dharma* of a doctor may be summarised as serving the needs of one's patients in terms of caring for their physical and mental health and well-being. A trained medic might have all the qualifications and skills, but that alone isn't *dharma*. To fulfill one's *dharma* as a doctor requires active and beneficial engagement serving one's patients.

Sentience is not a requirement for having a *dharma*. So, we may apply this system to AI, as a non-sentient entity. To frame a guiding *dharma* for AI, we first identify its purpose, inherent relationships, and service responsibilities. The definition of AI adopted by the European Commission proposes: "Artificial Intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions—with some degree of autonomy—to achieve specific goals."[28] Although this does not match the criteria of a statement of *dharma* for AI, it offers a starting point for its identity. In this chapter, we have argued that AI is a non-sentient information processing system modelled on the cognitive functions of the mind (as distinct from consciousness proper) and we have referred to it as "a problem-solving computational adjunct 'mind'". This is not sufficient as an identity in *dharma* terms. Dharmic identity must be expressed as a transitive relationship with beneficiaries, lacking which, that identity ceases to have meaning. Hence, we might suggest that AI is an entity with a transitive relationship and responsibilities to serve human society broadly. Based on the efforts of other experts, I suggest the following draft for the overall *dharma* of AI:

The dharma of AI is to serve human society broadly as an adjunct computational mind to enhance human capabilities for successfully addressing complex problems and thereby improving the life of all.

When asked in an interview if we should fear AI, Anil Seth, professor at Sussex University, retorted: “I’m much more worried about natural stupidity than artificial intelligence.”[29] A fair comment, but it is not an either/or scenario. There are two distinct aspects of the challenge that AI raises for human agency:

1. The generative and autonomous functions of AI as a non-sentient agent itself, and
2. The development, programming, and applications of AI by human agents.

A viable *dharma* statement offers the basis for evaluating how authentically AI is performing or being used within its defined *dharma*. The relevant criteria for such authenticity of *dharma* would include:

1. AI serves human society broadly—meaning that we intend it to serve all of humanity, not just privileged individuals, corporations, or nations.

2. It functions as a computational mind carrying out key functions such as: automating tasks, analysing data, and generating beneficial solution propositions which either assist human minds perform better or, when acting autonomously, act as a surrogate for human decision-making.

3. Its output and actions enhance human capabilities—meaning we need to be sure we are gaining additional benefits and better results (as judged by other criteria).

4. It enables us to successfully address problems (and not add new problems in its wake). This is a further qualitative requirement requiring consideration of suitable evaluation criteria.

5. It addresses complex problems. We should not accuse AI for failing in its *dharma* because it is regularly used for non-complex everyday tasks such as writing emails or our children' s homework. But we should remain mindful that our stated aspiration for AI is to address some of the major and profound problems facing our world. When so engaged in that use, it will be evermore critical that we are confident that AI is functioning true to its *dharma*.

6. It will improve the life of all. This may seem rather vague and impractical to monitor or achieve. Yet the UN, various nations, and NGOs have criteria and targets for how to gauge the progress of humanity on our planet. For instance, the UN’s 17 Sustainable Development Goals (SDGs) are intended to be a ”blueprint to achieve a better and more sustainable future for all.”[30]

Although our hope is for AI to operate to the high values we have defined as its *dharma*, it should be understood that AI is being, and may always be, misused by human carelessness and malevolence. That does not change its own *dharma*. Indeed, it calls on us to be more vigilant, to monitor AI’s own internal processes and outputs to ensure that they are not becoming malign because of human interference or negligence. It also accentuates the need, that alongside AI’s *dharma*, there must be *dharma* statements guiding developers, users, and regulators. The following is an attempt to draft in a single *dharma* statement the service responsibilities for all those responsible or involved in AI’s integration into society:

To serve humanity by developing AI as an accurate and beneficial tool for assisting us to address complex problems and by evaluating and approving its output of analysis and propositions by application of overall human supervision, decision-making and autonomous agency to ensure the value of AI for the benefit of all.

In addition to the criteria mentioned above for the *dharma* of AI, this second statement of *dharma* contains specific responsibilities for human personnel:

Oversight with Authority: Strong, effective oversight structures must be established to guarantee AI's accuracy and benefit to society, especially when addressing significant global challenges.

Caution in Global Applications: Stringent evaluation is essential for AI solutions to large-scale issues, where risks of exacerbating problems or unforeseen consequences are high.

Monitoring for Equity and Compliance: AI's contributions should reflect an equitable benefit for all, necessitating regulation and enforcement to prevent outputs that contravene legal or ethical standards.

Such a *dharma* framework for AI, while preliminary, proposes essential principles that support ethical guidance alongside existing laws and regulations promoting AI's responsible integration into society. *Dharma* establishes a high ethical standard that exceeds mere legal requirements. While legality sets a minimum baseline below which penalties are enforced, *dharma* advocates for positive ethical conduct that benefits society as a whole. Keeping the *dharma* standard in mind offers greater benefit than tolerating behavior barely within the bounds of the law.

Failures in AI systems or misuse by human handlers should generally be addressed under existing laws, covering issues like discrimination, fraud, or fiscal mismanagement. Like other technologies, responsibility should be traced through the chain of users, vendors, and manufacturers. Does blame lie with the end user, the vendor, the manufacturer, and so on? However, enforcing these laws is complex in digital and social media domains where jurisdictional challenges exist, especially for multinational corporations. It is hard to see what type of international agreements could be in place–especially when there are likely pressures of commercialization and government applications wishing to dominate such a powerful tool and weapon. One view is that the globe might divide into discrete zones of isolated tech and AI platforms choosing to insulate themselves from being hacked or simply out of a desire to control and immunize their people from the influences of other regions. We might be entering a new dawn of global reconfiguration in which AI is just one of the players amongst developing technologies and superpower corporations. The challenges to supervise and enforce legal standards are almost insurmountable. Yet, this only serves to reinforce the need to establish the common vision of a *dharma* standard to which we can aim to adhere.

Conclusion

AI is a fact of modern society and will continue to be integrated more and more into every facet of human life that involves information and thought. AI is a type of adjunct mind modelled on human cognitive processes. It emulates the psychological processing of data and decision-making that goes on in the human mind, and in that respect, it does many things we can do, but often better and more extensively. Many more human capabilities still elude AI. But even if AI advances to perform every cognitive task that humans are capable of, it will never be the same as us. Besides our abilities to think, rationalise, and create, we humans are aware of and certain of our existence as conscious entities; a sense

of being alive. Our experiences feel like something. Our mind generates the qualitative form of our inner experiences (qualia). Such qualia are irreducible to any property or process we know of in physics, biology, neuroscience, or computation. Hence, we conclude that regardless of what we call intelligence or thinking power within AI to whatever degree it could reach, there will remain a gulf between the actual experience of being subjectively sentient as a human and the cognitive, information-processing functions and outputs of computational systems. Hence, I have argued that we should proceed on the basis that there is no reason to suppose that AI is currently or will ever be sentient. This clarification is important because it establishes the primacy of human consciousness and agency as distinct and, if we want to use the term, superior to the nature of AI. Such clarity avoids promoting a demeaning message to the public that they are no different from a computer, robot, or software program; and indeed, worse: that they are less intelligent, less capable, and perhaps less valuable than those devices and applications. It also avoids us entering the quagmire of legislation and rights that might be claimed by and on behalf of AI-powered devices and applications. I have offered further rationale that animals demonstrated to undergo the experience of qualia are also distinctive from AI in terms of sentience, so there is no justification for extending the rights and protections for non-human biological life to computational systems. As an adjunct mind, AI can act both as a tool and, at times, a delegated agent. I have offered a *dharma* statement for AI:

> *To serve human society broadly as an adjunct computational mind to enhance human capabilities for successfully addressing complex problems and thereby improving the life of all.*

To ensure that AI remains true to its dharma, we must oversee its activity. Just as we, the conscious self possessed of free will, must monitor and deliberate on the propositions of our mind presenting ideas and propositions, so human society as the only agent with actual free will, must monitor and deliberate, stringently on the 'belief propositions'presented by our adjunct artificial mind. In this regard, allow me to offer a text from the Bhagavad-gita:

> "From wherever the mind wanders, due to its flickering and unsteady nature, one must certainly withdraw it and bring it back under the control of the Self."

This is our *dharma* as custodians and users of AI—to monitor, evaluate, and, whenever necessary, bring our minds, both natural and artificial, back under our control. The Gita offers a further note both of encouragement and of caution:

> "For one who has subdued the mind, the mind is the best of friends; but for one who has failed to do so, the mind will be the greatest enemy."

It is solely up to us whether AI will become our best friend or our worst enemy.

Bibliography

[1] Stryker, C. Kavlakoglu, E. (2024, August 16) What is artificial intelligence https://www.ibm.com/topics/artificial-intelligence

[2] Harari, Y. N. (2024) *Nexus: A Brief History of Information Networks from the Stone Age to AI*. Fern Press 2024

[3] Kurzweil, R. (2024). *The singularity is nearer: When we merge with AI*. Penguin Books.

[4] Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.

[5] Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450. https://doi.org/10.2307/2183914

[6] Searle, J. R. (1992). *The rediscovery of the mind*. MIT Press.

[7] Dennett, D. C. (1991). *Consciousness explained*. Little, Brown and Company.

[8] Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Roberts and Company.

[9] Descartes, R. (1641/1985). *Meditations on First Philosophy* (J. Cottingham, Trans.). In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), *The philosophical writings of Descartes* (Vol. 2, pp. 1-62). Cambridge University Press.

[10] Jīva Gosvāmi. (2001). *Tattva-Sandarbha: Sacred India's Philosophy of Truth* (S. Dasa, Trans.). Jiva Institute.

[11] Peirce, C. S. (1867). On a New List of Categories. *Proceedings of the American Academy of Arts and Sciences, 7*, 287–298. https://doi.org/10.2307/20179567

[12] Lewis, C. I. (1929). *Mind and the World-Order: Outline of a Theory of Knowledge.* Charles Scribner's Sons.

[13] Crick, F., & Koch, C. (1998). Consciousness and Neuroscience. Cerebral Cortex, 8(2), 97-107.

[14] Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach.* Roberts & Company Publishers.

[15] Nagel, T. (1974). What Is It Like to Be a Bat? The Philosophical Review, 83(4), 435-450.

[16] Jackson, F. (1982). Epiphenomenal Qualia. The Philosophical Quarterly, 32(127), 127-136.

[17] Searle, J. R. (1992). *The Rediscovery of the Mind.* MIT Press.

[18] Strawson, G. (1994). *Mental Reality.* MIT Press.

[19] Block, N. (1995). On a Confusion About a Function of Consciousness. Behavioral and Brain Sciences, 18(2), 227-247.

[20] Dennett, D. C. (1988). Quining Qualia. In A. Marcel & E. Bisiach (Eds.).

[21] Tye, M. (2007). Qualia. The Stanford Encyclopedia of Philosophy (Fall 2007 Edition), Edward N. Zalta (ed.).

[22] Elwood, R. W., & Appel, M. (2009). Pain experience in hermit crabs? Animal Behaviour, 77(5), 1243-1246.

[23] Seth, A. (2017). *Being You: A New Science of Consciousness.* Faber & Faber.

[24] Tononi, G. (2008). Consciousness as Integrated Information: A Provisional Manifesto. The Biological Bulletin, 215(3), 216-242.

[25] Tegmark, M. (2017). Life 3.0: Being Human in the Age of Artificial Intelligence. Knopf.

[26] Koch, C. (2019). The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed. MIT Press.

[27] Lumbreras S (2022) The Synergies Between Understanding Belief Formation and Artificial Intelligence. Front. Psychol. 13:868903. doi: 10.3389/fpsyg.2022.868903.

[28] European Commission, "Artificial Intelligence," 2021

[29] Anil Seth, interview with New Scientist, 2017.

[30] United Nations, 2030 Agenda for Sustainable Development. The 17 SDGs contained in this document aim to address global challenges such as poverty, inequality, environmental degradation, peace and justice.

[31] Prabhupada, S. B. (2006). *Bhagavad Gita As It Is*. (6.26) Intermex Publishing.

[32] Ibid (6.6)

The APA 7th edition format for references in academic papers follows specific guidelines depending on the type of source being cited. Below are examples for common types of references:

1. Books

Format:

Author, A. A. (Year). Title of the book: Subtitle if any. Publisher. DOI or URL (if available).

Example:

Smith, J. L. (2020). Understanding the universe: A guide to astronomy. Academic Press.

2. Journal Articles

Format:

Author, A. A., & Author, B. B. (Year). Title of the article. Title of the Journal, Volume(Issue), page range. https://doi.org/xxx

Example:

Doe, J., & Roe, R. (2019). The effects of mindfulness on stress. Journal of Psychology, 45(2), 123-135. https://doi.org/10.1234/jpsych.2019.0045

3. Chapters in Edited Books

Format:

Author, A. A. (Year). Title of the chapter. In E. E. Editor & F. F. Editor (Eds.), Title of the book (pp. xx-xx). Publisher. DOI or URL (if available).

Example:

Johnson, M. (2018). Quantum theories and realities. In T. Greene (Ed.), Modern physics debates (pp. 45-68). Science Press.

4. Webpages

Format:

Author, A. A. (Year, Month Day). Title of the webpage. Website Name. URL

Example:

National Institute of Mental Health. (2021, March 5). Anxiety disorders. NIMH. https://www.nimh.nih.gov/health/topics/anxiety-disorders/index.shtml

5. Reports

Format:

Author, A. A. (Year). Title of the report (Report No. xxx). Publisher. DOI or URL

Example:

World Health Organization. (2020). World health statistics 2020. WHO. https://www.who.int/data/gho/publications/world-health-statistics

General Guidelines

Author Names: Use initials for first and middle names (e.g., "Doe, J. R."). List up to 20 authors; for more than 20, list the first 19 followed by an ellipsis and the final author's name.

Italics: Titles of books and journals are italicized.

DOI or URL: Use a DOI if available; otherwise, include the URL. Do not include "Retrieved from" unless the source material is time-sensitive.

Hanging Indent: References should use a hanging indent (second and subsequent lines are indented).

If you have a specific source type in mind, feel free to ask for clarification!