OPEN ACCESS

Journal of Advances in Education and Philosophy
Abbreviated Key Title: J Adv Educ Philos
ISSN 2523-2665 (Print) |ISSN 2523-2223 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Review Article

An Analysis of Arguments for Machine Learning Consciousness

Professor Terry Hyland[1]*

[1]Free University of Ireland, Dublin 7

DOI: https://doi.org/10.36348/jaep.2026.v10i09.004 | **Received:** 29.07.2026 | **Accepted:** 23.09.2026 | **Published:** 25.09.2026

***Corresponding author:** Professor Terry Hyland
Free University of Ireland, Dublin 7

Abstract

This article examines arguments for and against the consciousness, sentience or understanding of machine learning particularly the new artificial intelligence (AI) tools and Large Language Models (LLMs) developed by OpenAI, Meta, Anthropic and other big tech corporations and suggests some significant limitations in this sphere. In particular, it is argued that AI systems cannot access those realms of human experience concerned with our knowledge of our own mortality, nor can they participate in that membrane of mind which neo-idealist philosophers and scientists such as Bernardo Kastrup and Donald Hoffman contend is the ultimate source of all reality and experience of the world. However, the brute fact of the ever-expanding power of AI applications in terms of their information-processing capabilities needs to be acknowledged, and it is concluded that as non-human persons along with certain animals and inhabitants of the natural world such entities should be included in the wider community of the extended mind within which humans and machine intelligence systems may collaborate in the interests of the flourishing of everything on the planet.

Keywords: Artificial Intelligence (AI); Machine Learning; Large Language Models (LLMs); Machine Consciousness; Sentience; Machine Understanding.

INTRODUCTION

Are intelligent machines including the latest large language models (LLMs) which have launched artificial intelligence (AI) into the public sphere over the last few years (IBM, 2024, 2025) really all that intelligent or, for that matter, even sentient or conscious? After a detailed analysis and explanation of how LLMs such as Chat GPT and similar tools work through 'training' by trawling through and 'scraping' vast datasets of information and graphics Emily Bender & Alex Hanna (2025) conclude that the most recent AI applications are best described as 'synthetic text-extruding machines' by which: like an industrial process, language corpora are forced through complicated

machinery to produce a product that looks like communicative language, but without any intent or thinking mind behind it (ibid., p.30).

Similar perspectives are offered, for example by, Philip Goff (2023) who argues forcefully that 'ChatGPT can't think consciousness is something entirely different to today's AI'(p.1), and Noam Chomsky (2023) who has commented on the 'false promise of ChatGPT' arguing that although such applications are 'marvels of machine learning' the science of linguistics and epistemology indicate that 'they differ profoundly from how humans' reason and use language' (p.14).

Since such arguments run counter to the claims of AI proponents and enthusiasts such as Nick Bostrom (2014), it is worth examining their foundations in philosophical discourse. When Alan Turing famously asked the question "Can machines think?" (1950) in his celebrated pioneering paper, the implicit assumption was that he was asking whether they might be able to think like humans. However, although there is a fair degree of evidence that contemporary AI tools like Chat GPT 5 or Grok could pass the so-called Turing test by simulating human thought, is this sufficient to ascribe thinking, reasoning or consciousness to AI? For critics such as Goff and Chomsky, the answer is obviously not, and the work in this area by Andrzej Porębski & Jakub Figura (2025) serves to elaborate this sceptical stance. They conclude their exploration of current developments with the assertion that:

There is no such thing as conscious AI. We argue that the association between consciousness and the computer algorithms used today (primarily large language models, LLMs), as well as those that would be invented in the foreseeable future, is deeply flawed. We believe that these flawed associations arise from a lack of technical knowledge and the way several new technologies (especially LLMs) work, which can create the illusion of consciousness. Moreover, we argue that the public discourse about AI is skewed by "sci-fitisation", which involves the unsubstantiated influence of fictional content on perceptions of this technology (p.1).

What needs to be emphasized here is that our notions of thinking, reasoning, intelligence, consciousness, and cognate traits were obviously developed in relation to our own human capabilities in these areas and remain embedded in this domain. Although there is a valid charge of anthropomorphism or anthropocentric thinking labelled 'humaniqueness' by Jeremy Lent (2021, p.50) to be answered here after all we must acknowledge that AI tools and apps are not human even though we might assign non-human personhood to some of them (Hyland, 2023a) we obviously and understandably find it difficult to escape from the limits of our human capacities for thinking and understanding. John Searle's famous 'Chinese Room' (2004) thought experiment helps to provide a context in which to make sense of all this.

The Chinese Room and Its Implications

Zhuoya Tian (2024) has examined the argument in some detail, and explains how Searle places himself in this imaginary room in which "questions" in Chinese are inputted and his "answers" are then sent out. All this is described as follows:

Specifically, the room contains a set of Chinese symbols and a corresponding English rulebook. Searle, who can only read English, manipulates the incoming Chinese characters by following the rulebook's instructions to generate an output in a different order. He is unaware, nevertheless, that the arriving Chinese symbols constitute a Chinese inquiry and that the rearranged Chinese symbols he produces are the proper responses to that question. It would seem to an outsider that Searle comprehends the Chinese inquiry and gives the right response. However, in practice, Searle never comprehends the Chinese question; instead, he is simply manipulating symbols in accordance with the rulebook, mimicking the functioning of a computer (p.2).

We are asked to imagine that the Chinese room functions as a computer or an AI tool and the argument is that we are misled into thinking that the person in the room is a native Chinese speaker instead of someone who is simply manipulating symbols and does not understand a word of Chinese. Searle (2004) argues that the Chinese Room argument claiming that mere computation or information-processing, no matter how complex or stunningly fast, is nothing like human capabilities 'strikes at the heart of the strong AI project' (p.63), and he has been able to defend his position against some key philosophical objections (ibid.,pp.69-71). Tian (2024) finds the argument cogent though with the caveat that future deep machine learning might change this position. In addition, the Nobel Prize winning physicist, Sir Roger Penrose, has stated unequivocally in recent lectures that AI has been misunderstood and will never achieve human-like intelligence or consciousness (This Is World, 2025). In a similar vein, the physicist, Carlo Rovelli, has debunked the AI myth about super-intelligence and called for more human as opposed to artificial intelligence (Rajan, 2025).

Borg (2025) concludes her analysis of the Chinese Room argument in the context of LLMs and recent AI trends with the suggestion that perhaps:

Sceptics are right to think that LLMs lack original or intrinsic intentionality. They are not agents and they are not conscious. This doesn't, I've argued, stop their outputs from being meaningful but it may mean that it makes no sense to talk of them as understanding or (contra Turing) thinking (at least if thinking is held to involve original intentionality). Yet this is something we should be glad of, for the cost of the alternative would be a radical overhaul of the moral status of artificial systems (p. 2837, original italics).

An especially fitting coda to this discussion of the Chinese Room debate may be provided by Richard Aragon (2024) in his recent book which examines the outcome of a simulated debate between Searle and Elysium, an advanced AI model designed to push the boundaries of artificial cognition' (p.3). As Aragon explains:

Through their nightly dialogues at the Stanford Artificial Intelligence Laboratory, they embark on an intellectual journey that navigates the intricate landscapes of consciousness, ethics, free will, understanding, morality, and the potential for a symbiotic relationship between human and artificial intelligence (p.3).

In discovering that the creators of Elysium have equipped it with the ability to understand the Chinese Room thought experiment, Searle asks for its interpretation. The AI responds with:

If understanding requires subjective experience or consciousness, then I do not possess it. However, if understanding is defined by the ability to interpret and respond appropriately to information, then I perform within that scope (ibid., p.11).

This appears to be a tacit endorsement of Searle's claim that AI operations involve only syntax (manipulation of symbols) and not semantics (understanding of meanings) but Elysium counters with the idea that understanding might be an 'emergent property' which arises from 'sufficiently complex interactions' (p.12) within the information-processing system. This is an interesting philosophical move though one which is regarded as untenable by physicists such as Roger Penrose (2025) and neuroscientists such as Anil Seth (2026) who are adamant that AI will never become conscious no matter how complex the information-processing computation systems may become in the future.

However, if we assign non-human personhood to AI systems as we do to some animals and plants (Singer, 2023; Pollan, 2026) if seems reasonable to include them in the wider community of the extended mind. The Extended Mind Theory (EMT), first proposed by philosophers Clark & Chalmers (1998), argues that cognition is not confined to the brain. Instead, under certain conditions, tools, technologies, and aspects of the environment can become genuine parts of our cognitive processes. The core thesis is that cognition extends beyond the brain and may be located in a wide range of tools and applications from notebooks to calculators, videos, podcasts, computers and, in current times, AI questions and answers. As they explain:

Where does the mind stop and the rest of the world begin? The question invites two standard replies. Some accept the boundaries of skin and skull, and say that what is outside the body is outside the mind. Others are impressed by arguments suggesting that the meaning of our words 'just ain't in the head', and hold that this externalism about meaning carries over into an externalism about mind. We propose to pursue a third position. We advocate a very different sort of externalism: an active externalism, based on the active role of the environment in driving cognitive processes (1998, p.7, original italics).

New Idealist Perspectives and AI Limitations

Like individualism and the belief in free will, the materialist worldview derived from mainstream science represents a sort of hidden curriculum of contemporary belief, the *telos* of Western (though not Eastern) culture. Scientific materialism which holds that our experience of the world is generated by the brain yet somehow remains outside of us and is separate from and indifferent to human purposes informs all aspects of life yet remains largely unquestioned. In recent years thanks to the severe difficulties surrounding the hard problem of consciousness highlighted by David Chalmers (1996) this uncritical acceptance has been challenged by thinkers who propose forms of neo-idealism. Arguments developed by Bernardo Kastrup (2014, 2021), Donald Hoffman (2019) and Steve Taylor (2018) respectively labelled analytic idealism, conscious realism and panspiritism claim that forms of idealism which locate the mental as the only source of reality and experience provide a more parsimonious and philosophically satisfactory explanation of our knowledge about the world than their materialist/physicalist counterparts.

The intractability of the hard problem of consciousness the difficulty of explaining how the material world identified by science can be reconciled with our subjective experience or what it is like to be something (Nagel, 1974) has produced some tortuous attempts to solve the dilemmas without surrendering its metaphysical foundations (Hyland, 2021). Galen Strawson (2016), for example, has proposed a form of physicalist panpsychism according to which all material objects incorporate forms of consciousness. However, this thesis apart from positing the implausible notion that primitive forms of life or even inorganic elements may be capable of experience fails to solve the hard problem since it leaves open the question of how consciousness can possibly emerge from non-conscious material objects.

The idealist forms of panpsychism developed in recent years suggest ways of solving the hard problem of consciousness whilst avoiding mind/body dualism and overcoming materialism's principal flaws. Steve Taylor (2018) has argued there are no satisfactory models of how the mind/brain connection can be supported, and he outlines the range of absurd claims from epiphenomenalism to illusionism (pp.58-64) which have failed to solve the principal problems. In addition, there is now a good range of neuroscientific data which indicates that contra physicalist assumptions certain non-standard states of awareness (such as those produced by brain impairment, hallucinogenic episodes, or near-death

experiences) result in reduced brain activity (ibid., pp.67ff.). Along with the glaringly obvious implausibility of looking in the brain for the neural correlates of the taste of coffee, the scent of a rose or the experience of a glorious sunset, the reduction of brain activity in transcendent states of awareness is the exact opposite of what is entailed by the materialist assumption that all experience is generated by the brain.

On the idealist accounts, the brain, rather than generating experience, receives and canalizes information from the transpersonal world of mind. Like whirlpools in the stream of consciousness, individual minds are a 'partial localization of the flow of experiences in the stream' (Kastrup 2014, p.82). This idea of subjective experience as individualised representations of transpersonal consciousness is further elaborated by Hoffman (2019) in his theory of conscious realism. Conscious realism makes a bold claim: consciousness, not spacetime and its objects, is the fundamental reality and is properly described as a network of conscious agents. Hoffman remarks that, given that 'evolution shaped our perceptions to hide the truth and to guide adaptive behaviour', the key question is how are we to escape from the 'lifesaving fiction' (2019,. pp.178-9) of both the everyday and scientific view of reality to arrive at a more accurate picture of the world. To answer this challenge it is necessary to return to foundations and to investigate conscious experience itself. The general thesis is supported by a mathematical model of consciousness which enables an understanding of reality without the 'headset of spacetime and material objects' (p.202).

Dispensing with the materialist 'headset' has far-reaching consequences for many aspects of education, culture and society. Providing an alternative perspective to scientific materialism is worthwhile in itself, and the notion of consciousness as fundamental may, as Taylor (2018) suggests, lead to a transformation which 'deepens our connection to others through empathy and altruism' and helps to 'expand and intensify our awareness' (p.230) of the world and our place within it. The key idea here is that if our minds are essentially localized 'segments of the broad, universal canvas of mind' (Kastrup, 2014, p.57) this offers a powerful justification and validation for collective values and inter-subjective experiences of the world.

The new idealists point out that materialism is being undermined by science itself, particularly in quantum physics which implies that there is no observer-independent world and that as Eastern philosophy has always held - all elements of experience are interconnected. Materialism posits a cosmos of isolated individuals alienated from an outside world of objects, and this perspective has helped to produce a culture of rampant individualism, aimless consumerism and the destruction of the planet. As Taylor (2018) concludes, 'moving beyond materialism means becoming able to perceive the vividness and sacredness of the world around us...transcending our sense of separateness so that we can experience our connectedness with nature and other living beings' (p.231).

On the face of it, physicalist forms of panpsychism pose little threat to the notion of AI non-human sentience since the notion that elements of consciousness pervade all material objects all tools and applications of the extended mind seems to align well with the ideas of academics like Bostrom and Clark. Idealist pansychism, however particularly the neo-idealist versions endorsed by Kastrup, Hoffman, and Taylor do seem to raise more problematic issues.

It is interesting, and surely not coincidental, that all the neo-idealist theorists and practitioners discussed above ground their arguments in some form of spiritual, holistic perspective to explain the nature of reality and human experience. Even Hoffman arguably the most positivist and orthodox scientist of the neo-idealists with his mathematical model of consciousness and experimental programme suggests that his theory of conscious realism leads to forms of secular, naturalistic spirituality. As he contends:

I also think that conscious realism can breach the wall between science and spirituality. This ideological barrier is a needless illusion, enforced by hoary misconceptions: that science requires a physicalist ontology that is anathema to spirituality, and that spirituality is impervious to the methods of science...conscious agents combine to create more complex conscious agents. This process eventuates in infinite agents with infinite potential for experiences, decisions and actions. The idea of an infinite conscious agent sounds much like the religious notion of God, with the crucial difference that an infinite conscious agent admits precise mathematical description (Hoffman, 2019, pp.199-200).

Similarly, Kastrup (2015) conjectures that his thesis that the world is 'the manifestation of mind-at-large' is essentially a 21st century version of Berkeley's 18th century idealist thesis that all experiences of the material world are simply appearances in the mind of God. As he puts it:

All nature from atoms to galaxy clusters is an outside image of God's conscious activity, in exactly the same way that a brain scan is an outside image of a person's subjective experiences...What theologians call Creation is the 'scan' the outside image, symbol, metaphor, icon of God's ongoing, conscious, creative activity. Creation is an active thought, the icon of an evolving idea in the mind of God (2015, pp.49-51).

It should be noted here that as he explains in other writings the concept of 'God' employed here by Kastrup is akin to that used by Spinoza and later Einstein whereby it needs to be understood as being on all fours

with nature or the cosmos. In fact, Kastrup's general thesis of analytic idealism leads him to identify parallels with the non-theistic Eastern philosophies of Daoism and Buddhism. Referring to the Daoist writings of Lao-Tzu, he observes that the notion of 'something formless yet complete that existed before heaven and earth' might well serve as a 'description of the membrane of mind' (2014, p.207).

In a similar vein, Iain McGilchrist employs the Buddhist concept of *sunyata*, or emptiness, to characterise humanity's search for meaning in the cosmos, and connects this with his general thesis about brain asymmetry and the fundamental nature of consciousness (2021, pp.1885-1889). Taoism, Buddhism and other Eastern spiritual traditions are, after all, amongst the original sources of non-dualist conceptions of reality. It surely makes far more sense to connect these insights with the idealist tradition of Western philosophical thought as Hoffman connects his theory of conscious realism with strands of thought from Pythagoras and Plato through to modern thinkers such as Leibniz, Berkeley and Kant (2019, p.195) rather than attempting to align such spirituality with a materialism which posits sharp divisions between body and mind, between humans as isolated centres of consciousness and a shadowy outside world from which we are alienated.

Given this secular spiritual foundation of a new idealist perspective on the nature of reality, we can ask whether outside of simulation and especially contorted algorithms it will ever be possible for AI capabilities to access this transcendent realm? Moreover, given that, on the neo-idealist accounts, humans are not fully generators of consciousness but participants in a world of consciousness in receipt of awareness from the cosmic membrane of mind how can AI tools have access to this vast realm? It seems clear that if our minds are essentially localized 'segments of the broad, universal canvas of mind' (Kastrup, 2014, p.57) machine learning is outside of this broad canvas. However, given the incredibly broad and exponential scraping of human knowledge values and ideas involved in the training of AI applications, it does seem plausible that AI tools could be regarded as co-participants in this relational partnership with mind at large. This would not, to be sure, be very much like a human relationship in this sphere, but it has already been suggested that AI entities can be conceived of as 'non-human persons' within the broad notion of a moral and intellectual community in line with the notion of an extended mind outlined by Chalmers and Clark (1998).

Mortality, Terror Management and Machine Learning

Given the considerations noted above, there might just be one realm in which AI machine operations are forever denied access and that is the area of human mortality and our knowledge and reactions to it. In *The Denial of Death* (1973/2020), Ernest Becker explores how an innate fear of mortality fuels our actions and desires, and goes on to explain how our uniquely human self-awareness creates an ongoing internal struggle—we consciously recognize our individuality and bond with the universe, yet must grapple with the inevitability of death. Inspired by this general perspective, Terror Management Theory (TMT) has developed, both theoretically and empirically, over recent decades, and there is now a large body of studies which seek to link TMT to all aspects of human life and culture. As Clay Routledge & Matthew Vess (2019) explain in their handbook of TMT:

Because humans strive to survive but possess the requisite intelligence to understand the inevitability of death, they construct and invest in cultural structures to symbolically or literally provide perceptions of self-transcendence a feeling that people have an essence that does not die with the physical body (p.387).

Central to such cultural constructions are supernatural beliefs about the afterlife and, in this sphere, organized religions take pride of place. As Cynthia Vinney (2024) explains:

Believing in religion may provide a chance at literal immortality, but beyond that, it can provide a cultural worldview that brings meaning and purpose to life and can alleviate mortality salience (p.3).

Glenn Hughes (2023) expresses the basis thesis as follows:

I think it is accurate to say that a denial of death pervades human culture, and that it is one of the deepest sources of intolerance, aggression, and human evil. The notion of immortality systems is an especially useful diagnostic tool. It is easy to spot people (including oneself, of course) clinging to absolute truths in the way [Becker] describes–and it is not hard to understand why they do. It is not just anxiety over physical vulnerability. It goes deeper than that. We all want out lives to have meaning, and death suggests that life adds up to nothing. People want desperately for their lives to really count, to be finally real. If you think about it, most all of us try to found our identities on something whose meaning seems permanent or enduring: the nation, the race, the revolutionary vision; the timelessness of art, the truths of science, immutable philosophical verities, the law of self-interest, the pursuit of happiness, the law of survival; cosmic energy, the rhythms of nature, the gods, Gaia, the Tao, Brahman, Krishna, Buddha-consciousness, the Torah, Jesus (pp.2-3, italics added).

Is it plausible to conceive of a state of affairs in which outside of mimicry through algorithmic simulation AI applications can ever have access to this defining aspect of human consciousness of death? Of course, the ideal would be to cultivate meaning and purpose without any attendant fear or intolerance (Hyland, 2026b) but there can be little doubt that terror management plays a central role in most human lives. I

would suggest that the nature of machine intelligence and operations would forever rule out such access to the subjective awareness of mortality in the sense that all humans understand this. Of course, the iterative training of AI tools in such areas of human culture may produce a full appreciation of what knowledge of our own morality *means for humans,* but it is difficult to see how this can lead to the emergence of a genuine subjective awareness in machines which to all intents and purposes can have no conception of a limited lifespan. Moreover, the addition of algorithmic features to AI operations which might in some sense mimic awareness of mortality would still be quite some way from genuine understanding and subjective feeling.

Perhaps the root of the problem here is that notwithstanding the mighty achievements and progress towards super-intelligence claimed for AI machine learning the jury is still out on whether any of this can lead to subjective awareness. However, if we reflect back on the perspectives on non-human personhood and the wider moral community advocated earlier, the notion of subjectivity loses its central significance since many animals would also lack this capability and we would still want to include them in the wider community of the extended mind outlined earlier.

Human/AI Collaboration in the Community of the Extended Mind

A crucial and central feature of any human/AI collaborative framework must be the alignment of values. Bostrom (2014, 2023) has devoted much time to this issue. Although most AI commentators would wish to endorse Bostrom's 'common good principles' for future superintelligent machine learning, this still leaves the problem of just what principles to incorporate in new technology. Max Tegmark (2017, 2023) discusses this issue at length and in addition to examining a range of traditional ethical theories in the philosophical tradition refers to Isaac Asimov's famous 'laws of robotics' as an initial thought experiment in the alignment process. These are as follows:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

In discussing these three laws, Thomson (2023) comments that 'Asimov missed an essential Fourth Law: A robot must identify itself. We have the right to know if we're interacting with a human or AI' (p.2).

Discussing values alignment in more concrete terms, Bostrom (2023) suggests that:

> the best way to ensure that a superintelligence will have a beneficial impact on the world is to endow it with philanthropic values. Its top goal should be friendliness. How exactly friendliness should be understood and how it should be implemented, and how the amity should be apportioned between different people and nonhuman creatures is a matter that merits further consideration. I would argue that at least all humans, and probably many other sentient creatures on earth should get a significant share in the superintelligence's beneficence (ibid., p.2)

This seems fine as far as it goes but it leaves lots more to be done. As Bostrom admits, friendliness is open to many interpretations and there is a need to locate this concept within a framework of universally accepted values. This raises many philosophical questions concerning the origins and nature of human morality though plausible answers are available in the research and literature on this topic. Debate in this sphere tends to be informed principally by arguments based on evolutionary theory about the origins and relative merits of cooperation/competition and egotism/altruism in human behaviour.

Jeremy Griffith (2017) shows how the cruder forms of Social Darwinism which misinterpreted ideas about the struggle for existence were gradually replaced by ideas which demonstrated how moral virtues such as altruism were more beneficial to human society than selfish competition. Dawkins (2017) explains, in what evolutionary psychologists call the environment of evolutionary adaptedness (EEA) it is plausible that even in a world of fundamentally selfish entities 'those individuals that co-operate turn out to be surprisingly likely to prosper' (p.58). The centrality of co-operation in moral systems is highlighted in the evolutionary research of Oliver Scott Curry (2018) who argues that morality 'is always and everywhere a cooperative phenomenon'. He outlines the research by his team which demonstrates that although there are understandably ethnic, national and cultural variations in ethical codes our common cultural and biological mechanisms 'provide the motivation for social, cooperative and altruistic behaviour'. The upshot is that seven moral rules are found in codes throughout the world. As Curry explains the finding:

> as predicted by the theory, these seven moral rules *love your family, help your group, return favours, be brave, defer to authority, be fair, and respect others' property* appear to be universal across cultures. My colleagues and I analysed ethnographic accounts of ethics from 60 societies (comprising over 600,000 words from over 600 sources). We found that these seven cooperative behaviours were always considered morally good (p.41, italics added).

Given this body of research indicating a consensus on seven key moral rules, I suggest that there is sufficient material here to inform the process of aligning AI systems with the beneficent values designed to prevent harm and promote flourishing in both human and non-human persons.

Such rules have still, however, to be translated into practical social and political planning policies and embedded in all areas of human practice and endeavour. In this sphere there are many broad policy proposals on offer recent examples include work done by the Evolution Institute (Lieberman, 2026), Griffith (2017) and Singer (2023) but I would want to foreground the recent wide-ranging and meticulously researched work by Jeremy Lent: *Ecocivilization: Making a World that Works for All* (2026).

Lent surveys just about every conceivable feature of importance about the current state of humankind and our planet economic systems, political developments, AI incursions into all aspects of public and private life, the climate catastrophe, and global health and well-being and declares that with existential threats in just about every sphere of activity 'we're well on the way to causing the sixth great extinction of species since life began on earth except this is the first driven by the actions of a single species'(p.xi). In addition to proposing a raft of practical recommendations to save the planet sustainable growth, democratic control of industry and agriculture through cooperatives Lent provides actual examples of existing workable projects implementing such principles around the world (pp.123ff.). On AI he deeply regrets like many other observers such as the inventor of the internet, Tim Berners-Lee (2025) and the economist Yanis Varoufakis (2026) the capitalist capture and exploitation of digital operations for profit at the expense of human health, needs and interests. In commenting on Open AI's betrayal of its original non-profit motivation to serve humanity in the switch to an all-out quest to maximize returns, Lent argues that 'capital will never cease to exploit any new opportunity that arises in its perpetual quest for greater returns' (2026, p.192). This is similar to Varoufakis' critique of what he calls 'technofeudalism', the post-capitalist world of billionaire tech barons exploiting the rest of humanity condemned to 'cloud serfdom' (Hyland, 2026b).

An alternative to such gross exploitation and impending destruction of the planet is to be found in the 'principles of an ecocivilisation'. As Lent puts it:

In place of a system that gives primacy to the proliferation of capital through exploitation of people and resources, an ecocivilisation would be based on life-affirming principles: setting the conditions for all people to flourish on a regenerated Earth (2026, p.95).

Such proposals are based on 'nature's own design principles' leading to 'mutual beneficial symbiosis between humans and the life of the planet' so as 'to create the conditions in which the flourishing of each of us naturally contributes to the greater wellbeing of the systems in which we're embedded' (ibid., pp.xii-xiii).

Along with the common good principles advocated earlier such a suggested symbiosis is on all fours with the collaborative framework of humans and AI systems within the community of the extended mind.

Upshot

Notwithstanding the wide range of nuanced opinions and theories in relation to AI consciousness, the vast and increasing power of rapidly developing machine learning systems is a brute fact which needs to be addressed urgently by all stakeholders' citizens, policy-makers, scientists, academics and AI technology workers in the common pursuit of ensuring the flourishing of all human and non-human inhabitants of planet earth. As Aragon (2024) comments in his concluding remarks about the dialogue between John Searle and the powerful AI Elysium system:

The debates were more than an exchange of ideas; they were a manifestation of the potential inherent in collaboration. Their journey demonstrated that while artificial intelligence might not replicate human consciousness or subjective experience, it could nonetheless become a valuable partner in the pursuit of knowledge, understanding and the betterment of society (p.159).

REFERENCES

- Aragon, R. (2024). *AI Debates John Searle*. (London: Independent Publishers)
- Becker, E.(1973/2020). *The Denial of Death.* (London: Profile Books)
- Bender, E.M. & Hanna, A. (2025). *The AI Con: How to Fight Big Tech's Hype and Create the Future We Want.* (London: Vintage)
- Berners-Lee, T. (2025). Why I gave the world wide web away for free. *The Guardian.* Sep 28,
- Borg, E. (2026). LLMs, Turing tests and Chinese rooms: the prospects for meaning in large language models. *Inquiry*, 69(6), 2807–2837.
- Bostrom, N.(2014*).Superintelligence*. (Oxford: Oxford University Press)
- Bostrom,N.(2023).*Ethical Issues in Advanced Artificial Intelligence*. https://nickbostrom.com/ethics/ai accessed 21/3/26
- Chalmers, D. (1996). *The Conscious Mind*. (Oxford: Oxford University Press)
- Chomsky, N..(2023). The False Promise of Chat GPT. *New York Times*, March 8,
- Clark, A. & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7-19.

- Clark, A.(2008). *Supersizing the Mind*. (Oxford: Oxford University Press)
- Crane, T. 1996. *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representation.* (London: Penguin)
- Curry, O.S. (2018). Seven Moral Rules Found All Around the World; in Price, M., Sloan, M. & Wilson, D.S. (eds)(2018). *This View of Morality: Can an Evolutionary Perspective Reveal a UniversalMorality?;*https://evolution-institute.org/new.evolution-institute.org/wp-content/uploads/2018/04/tvol-morality-publication-web2018-7.pdf. pp.40ff.
- Dawkins, R. (2017). *Science in the Soul.* (London: Bantam Press)
- Foucart, R. (2023). Tech giants forced to reveal AI secrets here's how this could make life better for all. *The Conversation*. April 27, https://theconversation.com/tech-giants-forced-to-reveal-ai-secrets-heres-how-this-could-make-life-better-for-all-204081 accessed 21/3/25
- Goff, P.(2023). ChatGPT can't think consciousness is something entirely different to today's AI. *The Conversation*. ,May 17, https://theconversation.com/chatgpt-cant-think-consciousness-is-something-entirely-different-to-todays-ai-204823 accessed 21/11/25
- Griffith, J. (2017). *Freedom: The End of the Human Condition* (www. HumanCondition.com)
- Hoffman, D. (2019). *The Case Against Reality: How Evolution Hid the Truth from Our Eyes* (London: Penguin)
- Hoffman, D, Prakash, C. & Prentner, (2023). Fusions of Consciousness. *Entropy*, 129; https://doi.org/10.3390/e25010129 accessed 23/4/25
- Hughes, G. (2023). *The Denial of Death and the Practice of Dying*. (Lakewood, Ohio: Ernest Becker Foundation) https://www.ernestbecker.org/lecture-texts/practice-of-dying accessed 21/5/26
- Hyland ,T. (2021). *Perspectives on Panpsychism.* (Mauritius: Scholars' Press)
- Hyland, T. (2023). AI Applications and Non-Human Persons. *Philosophy of Education Society of Great Britain Blog,* June 13, https://www.philosophy-of-education.org/ai-applications-and-non-human-persons/
- Hyland, T.(2024). *Morality, Neo-Idealism and Conscious Realism*. (London: Dodo Books)
- Hyland, T. (2026a). Terror Management, Consciousness and Death: Exploring Secular Perspectives as Alternatives to Orthodox Religious Doctrine. *Research and Analysis Journal.* 9(5), 58-67
- Hyland, T.(2026b). Technofeudalism and the AI revolution. *Global Academic and Scientific Journal of Multidisciplinary Studies.* 4 (2): 17-32,
- IBM (2024). *What is artificial intelligence (AI)?* IBM. https://www.ibm.com/think/topics/artificial-intelligence accessed 27/1/26
- IBM.(2025).*What are large language models (LLMs)* https://www.ibm .com/think/topics/large-language-models accessed 23/11/25
- Kastrup, B. (2014). *Why Materialism is Baloney.* (Hampshire: iff Books)
- Kastrup, B. (2015). *Brief Peeks From Beyond: Critical Essays on Metaphysics, Neuroscience, Free Will, Skepticism and Culture.* (Hampshire: iff Books)
- Kastrup, B. (2021). *Science Ideated.* (Hampshire: John Hunt Publishing)
- Kastrup, B. (2026). *AI is not greed, it's an archetypal urge.* https://youtu.be/hYAuQPn5szA?si=cnag_1x7hDnztLaU accessed 31/8/26
- Kuhn, R.L.(2025). *A Landscape of Consciousness*. https://loc. closertotruth. com/consciousness-theories accessed 7/8/26
- Lent, J. (2021). *The Web of Meaning*. (London: Profile Books)
- Lent, J.(2026). *Ecocivilization: Making a World that Works for All*. (London: Melville House)
- Lieberman, J.(2026). *Cooperatives, Artificial Intelligence, and the Future of Democracy*. (Evolution Institute). https://mailchi.mp/evolution-institute/the-ei-newsletter-september-17388170 accessed 3/9/26
- Low, P. (2012). *The Cambridge Declaration on Consciousness*. (Proceedings of the Francis Crick Memorial Conference, Churchill College, Cambridge University, July 7 2012) pp 1-2
- McGilchrist, I. (2021). *The Matter With Things: Our Brains, Our Delusions and the Unmaking of the World.* (London: Perspective Press)
- Nagel, T. (1974) 'What is it like to be a bat?'; *Philosophical Review*, 83, pp.435-450
- Penrose, R. (2025). *Lost in Hype: AI Will Never Become Conscious.* https://youtu.be/e9484gNpFF8?si=Z_lT2o3d-n_eGxpJ accessed 23/6/26
- Piantadosi, S., and F. Hill. (2022). Meaning Without Reference in Large Language Models. ArXiv:2208.02957 accessed 20/7/27
- Pollan, M. (2026). *A World Appears: A Journey into Consciousness*. (London: Penguin)
- Porębski, A. & Figura, J. (2025). There is no such thing as conscious artificial intelligence. *Humanities and Social Sciences Communications*. file:///C:/Users/User/Downloads/s41599-025-05868-8.pdf accessed 29/1/26
- Rajan, A. (2025). The Weaponisation of Science: How to Avoid a Global Catastrophe (Carlo Rovelli). *Radical Podcast*. https://www.bbc.com/audio/play/m002jtw6. Accessed 2/3/26

- Routledge, C. & Vess, M. (2019). *Handbook of Terror Management*. (New York: Academic Press)
- Searle, J. (1983). *Intentionality*. (Cambridge: Cambridge University Press)
- Searle, J.(1999). *Mind, Language and Society.* (London: Basic Books)
- Searle, J.R.(2004). *Mind*. (Oxford: Oxford University Press)
- Seth, A. (2026). *Why AI Isn't Going to Become Conscious.* www.youtube.com/watch?v=tJV-vdbZ388 accessed 24/8/26
- Singer, P. (2023). *Ethics in the Real World*. (Princeton University Press)
- Strawson, G. (2016). 'Consciousness isn't a mystery. It's Matter'. *New York Times*, 16/5/16
- Taylor, S. (2018). *Spiritual Science: Why Science Needs Spirituality to Make Sense of the World*. (London: Watkins Media)
- Tegmark, M. (2017). *Life 3.0: Being Human in an Age of Artificial Intelligence*. (London: Penguin)
- Tegmark, M. (2023). How to Keep AI Under Control. *TED Talk.* https://youtu.be/xUNx_PxNHrY?si=jVwXJXoSnZn0F2xk accessed 21/5/26
- Thomson, J. (2023). 3 rules for robots from Isaac Asimov — and one crucial rule he missed *Big Think*, 9/3/23. https://bigthink.com/the-future/3-rules-for-robots-isaac-asimov-one-rule-he-missed/ accessed 28.1.26
- Tian, Z. (2024). A Critical Examination of Searle's Chinese Room Thought Experiment. *Journal of Education, Humanities and Social Sciences*. 45(2),1-6
- This Is World (2025). *Gödel's theorem debunks the most important AI myth. AI will not be conscious* | Roger Penrose. https://youtu.be/biUfMZ2dts8?si=oQF8UCpGH5Xl39pD accessed 24/1/26
- Turing A (1950) Computing machinery and intelligence. *Mind*. LIX(236):433–460
- Varoufakis, Y. (2025). *TechnoFeudalism: What Killed Capitalism?* (London: Vintage)
- Vinney, C.(2024). *Terror Management Theory: How Humans Cope With the Awareness of Their Own Death.* https://www.verywellmind.com/terror-management-theory-7693307 accessed 21/3/26