

# Artificial Intelligence (AI) Developments and their Implications for Humankind: A critical analysis of contemporary trends and perspectives

---

## Abstract

The recent appearance of widely available artificial intelligence (AI) applications such as ChatGPT and Bard has fuelled a flurry of popular and academic discussions about the implications of such AI tools for all aspects of contemporary life and culture. After examining some aspects of recent developments, this article goes on to review some salient critiques of the emerging AI debate with the intention of analyzing some of the key themes in the current discourse. The key objective will be to deflate some of the more overblown and alarming perspectives informed by anthropomorphising AI developments. **In particular, there is an attempt to steer a middle course between doom-laden pessimism and futuristic optimism.** Drawing on a range of philosophical positions, it is concluded that AI applications are best conceptualised as powerful tools which need to be utilised pragmatically and regulated ethically in partnership with humans in the best interests of all of us.

Key words: Artificial intelligence, AI applications, consciousness, idealism, education

## 1. Introduction

The appearance in recent years of Open AI's ChatGPT series and Google's Bard application (AI News, 2023) has resulted in a spate of articles analysing such developments reflecting views ranging from catastrophic hyperbolic theories to incredulous debunking. Renaud Foucart (2023) offers a representative illustration in his comment that:

AI is expected to affect every aspect of our lives – from healthcare, to education, to what we look at and listen to, and even **how** well we write. But AI also generates a lot of fear, often revolving around a god-like computer becoming smarter than us, or the risk that a machine tasked with an innocuous task may inadvertently destroy humanity. More pragmatically, people often wonder if AI will make them redundant (p.1)

Many of the concerns about the new AI tools have been expressed by educators who fear that teaching and learning will be damaged by the easy access to the large language models (LLMs) like ChatGPT which can write essays and answer assignment questions in a matter of minutes. As Will Douglas Heaven (2023) observed in the *MIT Technology Review*:

Just days after OpenAI dropped ChatGPT in late November 2022, the chatbot was widely denounced as a free essay-writing, test-taking tool that made it laughably easy to cheat on assignments. Los Angeles Unified, the second-largest school district in the US, immediately

blocked access to OpenAI's website from its schools' network. Others soon joined. By January, school districts across the English-speaking world had started banning the software, from Washington, New York, Alabama, and Virginia in the United States to Queensland and New South Wales in Australia. Several leading universities in the UK, including Imperial College London and the University of Cambridge, issued statements that warned students against using ChatGPT to cheat (p.2).

On a more general level, an open letter signed by, amongst other leading figures in the digital technology world, Elon Musk and co-founder of Apple, Steve Wozniak, called for a pause to current machine learning AI developments whilst the wider implications are evaluated carefully. The letter published by the *Future of Life Institute* (2023) warns that:

AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research and acknowledged by top AI labs... *Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable...* Therefore, we call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4. This pause should be public and verifiable, and include all key actors. If such a pause cannot be enacted quickly, governments should step in and institute a moratorium. (p.1, original italics)

In a similar vein – in an announcement which stunned the AI tech world – Geoffrey Hinton, the so-called “godfather of AI” quit his job at Google warning of the dangers of unregulated and unsupervised trends in the field. Among his comments he observed that:

Some of the dangers of AI chatbots were “quite scary”, warning they could become more intelligent than humans and could be exploited by “bad actors”. “It’s able to produce lots of text automatically so you can get lots of very effective spambots. It will allow authoritarian leaders to manipulate their electorates, things like that. But, he added, he was also concerned about the existential risk of what happens when these things get more intelligent than us. I’ve come to the conclusion that the kind of intelligence we’re developing is very different from the intelligence we have. So it’s as if you had 10,000 people and whenever one person learned something, everybody automatically knew it. And that’s how these chatbots can know so much more than any one **person**. (Taylor & Hern, 2023)

All of these concerns – especially those being expressed by experienced experts and practitioners in the AI field – merit serious consideration. However, a core aspect of such consideration needs to include a more precise analysis of the concepts and claims being made about the nature of AI in terms of knowledge, computation, intelligence and consciousness. In this respect, it would be useful to examine the issues within the framework of recent writings by philosophers and scientists concerned with mind, neuroscience and consciousness.

## **2. AI and Consciousness**

Studies in this field in recent years have been dominated by the so-called ‘hard problem’ (Chalmers, 1996) of consciousness. **Susan Blackmore (2011)** has defined this problem in terms of the question: ‘how can objective, physical processes in the brain give rise to subjective experience?’(p.25). Within philosophy of mind, this ‘mind-body problem’ goes back at least as far as Descartes and his infamous dualist analysis of the mental and physical worlds which leaves unexplained exactly how they may be connected. More generally it

results in the long-standing problem of how to explain subjective mental phenomena such as hopes, wishes, intentions, emotions, or just ordinary sense data – as Nagel (1974) puts it, simply what it is like to be something – in a world which, according to science, consists only of material objects, forces and processes. A number of solutions in the form of reconciliation strategies have been proposed in relation to the hard problem including the idea that there is no serious problem since the mind and mental events are simply what the brain does (hence a form of extended materialism; see Dennett, 1991) or, alternatively, that all material objects are imbued with forms of consciousness which evolve more fully within complex systems.

This latter view is what contemporary panpsychism (Hyland, 2021) has largely come to mean and – in its materialist or physicalist form – has been championed most prominently by Galen Strawson (2006). As against this materialist solution to the hard problem, neo-idealist theories which posit the notion that consciousness is an ontological primitive – the foundation of all reality – have been advanced by Bernardo Kastrup (2014) and Donald Hoffman (2019). These neo-idealist perspectives will be drawn on extensively below in the critique of current AI claims but, at this stage, such claims about the nature and extent of AI capabilities are worth analysing more forensically.

### *2.1 AI, Intelligence, and Human Reasoning*

Much discussion about AI applications and devices makes use of language which is metaphorical/analogical rather than literal/factual. For example, we might hear about home thermostats “knowing” when to switch off in response to temperature changes, security cameras “sensing” when objects come within their ambit or, at a more advanced level, self-driving cars behaving like “intelligent” motorists in following routes and avoiding collisions. All of this behaviour is achieved by forms of AI but the language used to describe them can be ambiguous and misleading. Knowing, sensing, cognition, and intelligent behaviour are definitively human traits and, when extended to machines, tools and applications such as Google or ChatGPT, need to be understood in extended metaphorical terms.

In examining the metaphorical anthropomorphising of AI devices, it is worth entering the caveat that there is no intention here of exalting human intelligence, ability and consciousness by placing it above the rest of the natural world. In recent work, Jeremy Lent (2021) makes the useful distinction between what he describes as ‘animate intelligence...the original AI’ found in abundance in the natural world and in the lives of our ancestors, and ‘conceptual intelligence’ which is broadly those forms of reasoning distinctive of human behaviour and measured by IQ tests (pp.33-4). Lent goes on to explain that animate intelligence can be detected even at the level of microscopic cells which are observed to be ‘acting purposefully to maintain and propagate’ their lives. Similarly, sentient awareness of surroundings and intelligent behaviour is fully on display in the ‘networked intelligence of plants’ and animals throughout the natural world (pp.40-43). In the valuing of conceptual intelligence and consciousness based on human cognition – what is criticised as ‘our mainstream culture’s self-congratulatory obsession with humaniqueness’ (p.50) – to the neglect of animate forms found in abundance throughout nature, we are, Lent argues, overlooking so much that can benefit the whole planet. What is required is an integration of animate and conceptual forms of awareness; as he puts it:

Perhaps the greatest challenge to human intelligence today is not how to accomplish the next technological breakthrough or build the most advanced AI, but how to integrate human ingenuity with our own animate intelligence and that of the natural world (p.55).

With this framework in mind, it is worth looking more closely at some of the key concepts in this field – in particular, intelligence and consciousness – with a view to clarifying the discussion about contemporary AI trends.

### Intelligence

Intelligence, according to Anna Cianciolo (2014), is a term:

“Commonly applied to the capacity of humans (and sometimes that of other higher mammals) to accomplish a wide range of mental tasks, including comprehension, analysis, abstraction, and prediction among others. It is named as a key factor in learning and success in the academic, occupational, personal and social domains” (p.417) [Anna Cianciolo, 2014].

She then goes on to identify intelligence as test performance, as higher-order thinking, and as social function, and there is an extensive literature on the nature/nurture debate on cultivating intelligence and on **Howard Gardner’s theories of ‘multiple intelligences’ (1993)**

It seems that intelligence – like sentience, consciousness, emotional sensitivity, moral sense, and almost all traits and capabilities – **needs to be viewed** in terms of a spectrum of factors, and at one far end of the spectrum is to be found the notion of ‘superintelligence’ identified and analysed extensively by Nick Bostrom (2016), the philosopher and Director of the Strategic Artificial Intelligence Research Centre at the University of Oxford. Bostrom writes that ‘we use the term “superintelligence” to refer to intellects that greatly outperform the best current human minds across many very general cognitive domains’ (p.63). He then goes on to define a range of different types of superintelligence such as:

- Speed superintelligence: A system that can do all that a human intellect can do, but much faster
- Collective superintelligence: A system composed of a large number of smaller intellects such that the system’s overall performance across many very general domains vastly outstrips that of any current cognitive system
- Quality superintelligence: A system that is at least as fast as a human mind and vastly qualitatively smarter (ibid., pp.64-69)

Bostrom goes on to contend that ‘artificial intelligence already outperforms human intelligence in many domains’ (ibid, 14) and cites AI wins in chess, Go and other games over human grand masters, a claim that has much support within educational circles (Heaven, 2023; Chen, 2023).

### Consciousness

In discussing the hard problem of consciousness earlier, reference was made to the widely accepted conception of consciousness in terms of phenomenal experience or what it is like to be something (Searle, 2004). If we add to this picture the idea of the spectrum mentioned in relation to intelligence, it seems reasonable to suggest – in addition to degrees of consciousness in humans, say from simple wakeful awareness to advanced problem solving through to the meta-consciousness or thinking about thinking – that there are many levels of consciousness in the natural world. In addition to what Lent described as conceptual awareness, there is the animate consciousness found throughout the natural world but where do we draw the boundary lines? As Anneka Harris (2019) comments:

An organism is conscious if there is *something that it is like* to be that organism...Is it *like something* to be a grain of sand, a bacterium, an oak tree, a worm, an ant, a mouse, a dog? At some point along the spectrum the answer is yes, and the great mystery lies in why the “lights turn on” for some collections of matter in the universe (p.5, original italics)

If we apply this idea of degrees, levels and when the ‘lights turn on’ to AI applications, this takes us to the heart of the contemporary debate. Responding to recent developments in the field, critics such as Noam Chomsky (2023) have 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 (14). Similarly, Philip Goff (2023) argues forcefully that ‘ChatGPT can’t think – consciousness is something entirely different to today’s AI (1). Against this, Bostrom’s recent comments on developments are more favourable to AI sentience. In a recent interview he claimed a fair degree of sentience for LLMs such as ChatGPT, and went on to say that:

“I also think it’s not doing them justice to say they’re simply regurgitating text...They exhibit glimpses of creativity, insight and understanding that are quite impressive and may show the rudiments of reasoning... If an AI showed signs of sentience, it plausibly would have some degree of moral status... This means there would be certain ways of treating it that would be wrong, just as it would be wrong to kick a dog or for medical researchers to perform surgery on a mouse without anesthetizing it” (Al-Sibai 2023, 1-2).

In order to locate such disagreements within a philosophical/scientific framework, it is worth examining the central issues against the background of the physicalist/idealist dichotomy mentioned earlier.

### **3. Physicalism, Idealism and AI Consciousness**

Before examining the key issues in this sphere in terms of physicalist and idealist conceptions of our relation to the world, it would be worthwhile examining some general questions concerning the possibility of machine consciousness against the background of the sceptical positions illustrated above in the comments of Chomsky and Goff. A useful framework for such an examination might be provided by the famous Turing Test, a thought experiment devised by the computer pioneer Alan Turing in 1950 as a means of determining whether any form of AI reasoning or intelligence could be considered on a par with or indistinguishable from that displayed by humans.

#### *3.1. AI and Turing Consciousness*

Eric Kleppen (2023) explains that the Turing Test:

grew out of a thought experiment devised by computer scientist Alan Turing in which he devised what he initially named The Imitation Game. This test pits human respondents against a machine in order to test the machine’s ability to exhibit human-like responses and intelligence. To this day, the Turing Test is widely considered a benchmark for measuring the success of AI research (p.1)

Kleppen notes that ‘while no machine has ever passed the Turing Test flawlessly, several machines have fooled judges to some extent’. However, he goes on to suggest that ‘beyond the limitations of the test itself, many AI researchers feel the Turing Test is irrelevant today’.

The contemporary AI context is rather different from the machine intelligence available in Turing's day. As Kleppen concludes:

Whether or not the Turing Test is truly relevant remains a hotly debated topic for AI researchers. That said, many feel AI is still a long way from achieving human-like general intelligence and the Turing Test remains one of the many ways in which humans can evaluate a dimension of an AI's abilities. When companies like Google create large language models and push the boundaries of chatbot technology, they still use human evaluators to ask a series of questions to determine its abilities. In this way, some form of Alan Turing's thought experiment remains culturally relevant to the advancement of artificial intelligence (ibid.,p.2).

At least part of the reason for the scepticism of critics such as Chomsky and Goff may be located in the tendency noted earlier to exalt human capabilities above all others – perhaps informed by the assumption that only living organisms can exhibit consciousness – thus debarring all claims to sentience and human-like reasoning from any synthetic constructions based on computational information processing such as supercomputers, robots or Chatbots. John Searle's famous 'Chinese Room' thought experiment can serve to throw some light on these issues. The key thrust of the argument is explained clearly by John Horgan (2021):

[Searle] asks us to imagine a man who doesn't understand Chinese sitting in a room. The room contains a manual that tells the man how to respond to a string of Chinese characters with another string of characters. Someone outside the room slips a sheet of paper with Chinese characters on it under the door. The man finds the right response in the manual, copies it onto a sheet of paper and slips it back under the door. Unknown to the man, he is replying to a question, like "What is your favorite color?," with an appropriate answer, like "Blue." In this way, he mimics someone who understands Chinese even though he doesn't know a word. That's what computers do, too, according to Searle. They process symbols in ways that simulate human thinking, but they are actually mindless automatons (p.2).

Searle (2004) argues that the Chinese Room argument – claiming that mere computation or information-processing, no matter how complex or stunningly fast – '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). However, it is worth noting that the argument here leans heavily on understanding language – particularly the differences between syntax (grammar) and semantics (meaning) – and, though applicable to the LLMs discussed earlier, might not be as relevant to the more general and advanced forms of machine learning currently emerging. Also, in identifying language as the *sine qua non* of human reasoning the argument might be said to fall foul of the anthropocentric "humaniqueness" mistake referred to above. Pointing to the extensive animate intelligence displayed by animals throughout the natural world, critics might wish to suggest that sentience and consciousness can exist without an understanding of human language.

### *3.2.Physical and Idealist Conceptions*

Strawson (2006) is one of the leading proponents of the physicalist or materialist perspective on the nature of consciousness. He asserts that:

Consciousness... [by which] I mean what most people mean in this debate: experience of any kind whatever...is the most familiar thing there is, whether it's experience of emotion, pain, understanding what someone is saying, seeing, hearing, touching, tasting or feeling. It is in

fact the only thing in the universe whose ultimate intrinsic nature we can claim to know. It is utterly unmysterious (p.1)

Strawson then goes on to assert that the so-called objective and unmysterious nature of the physical world is, in fact, far from the truth. As he comments:

The nature of physical stuff, by contrast, is deeply mysterious, and physics grows stranger by the hour. (Richard Feynman's remark about quantum theory — "I think I can safely say that nobody understands quantum mechanics" — seems as true as ever.) Or rather, more carefully: The nature of physical stuff is mysterious *except insofar as consciousness is itself a form of physical stuff* (ibid., original italics)

Although this strategy goes some way in solving the hard problem, it still has to deal with the legacy of Cartesian idealism: how does non-mental matter give rise to purely mental phenomena such as the taste of coffee or the smell of a rose.? In order to deal with this thorny problem, a number of philosophers and cognitive scientists in recent years have adopted idealist positions which posit the notion of consciousness as an ontological primitive, the foundation of all reality and experience of the world. In the 19th century the idealistic turn was pioneered by William James whose idea was that in the origins and history of the cosmos there is no clear cut-off point at which consciousness or subjective experience suddenly appears out of a past which contains no such experience. He thought that 'experience was present at the very origin of things' (1983, p.152), and was, thus, a fundamental feature of nature. Similar ideas were put forward by A.N. Whitehead whose vision – following the monism of Leibniz and Spinoza – was founded on the idea that mental entities may emerge from non-mental ones in the sense that 'the many become one and are increased by one' (1978, p. 32).

Harris (2019) discusses the problem of explaining how mental qualities emerge naturally from non-mental ones without resorting to mystical 'New Age' mythology or endorsing absurd views about the consciousness of worms, bacteria or smart phones. She explains that self-evidently there are different forms and levels of consciousness – from the minimalist mental states of animals and plants to the sophisticated subjective experiences characteristic of humans – and it is a mistake to conflate all the various forms of experiential states in nature, especially if this involves the anthropomorphism which involves attributing mind-states to plants and animals. Harris observes that:

Unfortunately, it seems quite hard for us to drop the intuition that consciousness equals complex thought. But if consciousness is in fact a more basic aspect of the universe than previously believed, that doesn't suddenly give credence to your neighbour's belief that she can communicate telepathically with her ficus tree. In actuality, if a version of panpsychism is correct, everything will still appear to us and behave exactly as it already does (p.83).

In a similar vein, David Skribina (2017) warns us not to conflate human consciousness with a 'certain universal quality of physical things in which both inanimate mentality and human consciousness are taken as particular manifestations' (p.17).

Such fundamental category mistakes are what Bernardo Kastrup (2020) points to in his re-reading and re-interpretation of Schopenhauer's metaphysics. It is of crucial importance, he argues, to distinguish between the consciousness of all animals which consists in raw

awareness or brute experience of nature, and human self-awareness based on abstract ‘re-representations’ of experience and ‘levels of meta-consciousness’ (pp.32ff).

Amidst the diverse range of panpsychist perspectives referred to above it is possible to discern a number which lean more towards the mentalist or idealist interpretations of the grounds of knowledge and experience than the standard materialist versions alluded to by Strawson. This trend is both rational and understandable in terms of a determination to follow the Occam’s Razor principle of ‘simplest is best’ to its logical conclusion. Bruenstrap advocates such a position as a form of ‘ontological simplicity’ (2016, pp.2ff) which may be justified by Bruno’s famous dictum that even if the basic idea turns out not to be true it is, at least, well conceived.

As indicated above, Kastrup and Hoffman are among the leading proponents of a neo-idealist metaphysics as an answer to the hard problem of consciousness. Kastrup (2014) articulates his thesis that “the cosmos is mental and everything is mind by means of a series of ingenious metaphors and analogies which seek to explain the world revealed to us through experience in ways which are both cogent, precise, and more epistemologically and metaphysically satisfying than the mainstream materialist theories. We are asked to picture the ultimate primitive mind or cosmic consciousness as a ‘thin, mirror-like membrane with some rigidity, but also some elasticity’ such that the ‘qualities of experience now correspond to the specific patterns of vibration of the membrane’ (p.138). There is, thus, ‘nothing to reality but the medium of mind itself’ (ibid.) and all our experiences of the world may be explained in terms of the vibrations and oscillations of the membrane of mind. Subjective individualised experiences of the world may be correlated with the ripples and loops of this membrane which brings about segmented islands of consciousness. The metaphor is thus intended to explain both why we seem to have limited control over the unfolding of events in the world and also why we seem to be separated from each other in terms of our individualised states of consciousness”.

In other work, however, Kastrup (2015) is concerned to emphasise that “both of these characteristics of subjective consciousness – lack of control and individual ego states – are actually largely illusory and, as such, present us with a confused and partial perspective on reality. In order to escape such confuse it is necessary to wield Occam’s Razor forcefully and accept that everything is a modification of consciousness”. As he explains:

*I claim that we do not need more than consciousness to explain reality: all things and phenomena can be made sense of as excitations of consciousness itself. According to this more parsimonious view, the ground of all reality is a transpersonal flow of subjective experiences that I metaphorically describe as a stream. Our personal awareness is simply a localization of this flow: a whirlpool in the stream (2015,.p.13, original italics).*

Like Kastrup’s radical idealist perspectives, this position is taken to its logical conclusion by Hoffman (2019) in this theory of conscious realism. Hoffman’s startlingly radical thesis incorporates ideas and data from evolutionary theory, cognitive psychology, neuroscience, quantum physics and philosophy to establish a position which suggests that our assumptions about our knowledge of the objective world are mistaken and, moreover, that forms of consciousness are fundamental to everything that we may claim to know, think and experience. There are two principal aspects of Hoffman’s approach: one drawn from evolutionary game theory which purports to explain why our perceptions of reality are mistaken, and another strand which attempts to move beyond the hard problem of

consciousness by offering a conception of interacting conscious agents supported by a mathematical model of consciousness.

In dealing with the counter-intuitive notion that our senses deceive us as to the nature of reality – why would evolution, after all, not favour true perceptions of an objective world – Hoffman uses the metaphor of a computer interface (p.xii). The purpose of a desktop interface, he argues, is not to reveal the “truth” of the computer in terms of its various circuits, voltages and layers of software but to hide this truth to enable the pragmatic task of writing emails and completing internet research. This metaphor is then applied to evolution and our experience of the world in the following way:

This is what evolution has done. It has endowed us with senses that hide the truth and display the simple icons we need to survive long enough to raise offspring...You may want truth, but you don't need truth. Perceiving truth would drive our species extinct. (ibid.,pp..xii-xiii).

This argument from evolution is reinforced by data from the field of evolutionary game theory to construct an operationally pragmatic theorem which Hoffman labels ‘Fitness-Beats-Truth (FBT)’ which is itself based on universal Darwinism by which survival, adaptation and reproduction trumps all other considerations. Applying game theory models to this construction we arrive at the astonishing conclusion that ‘fitness drives truth to extinction’ (Hoffman, 2019, p.61). After examining various speculations – most notably those proposed by Bostrom and others – that consciousness may arise out of a computer simulation (see Hyland, 2019), Hoffman employs the analytical tools mentioned in earlier sections to conclude (as Kastrup does also) that ‘all attempts at a physicalist theory of consciousness have failed’ (2019, p.183). He reasons that:

Occam's Razor, applied to the science of consciousness, counsels a monism over an amphibious dualism, a theory based on one kind rather than two...If we grant that there are conscious experiences, and that there are conscious agents that enjoy and act on experiences, then we can try to construct a scientific theory of consciousness that posits that conscious agents – not objects in spacetime – are fundamental, and that the world consists entirely of conscious agents (ibid.,pp.182-3).

### 3.3. *Idealist Consciousness and AI*

Applying the idealist framework established by Kastrup and Hoffman to contemporary AI developments helps to clarify a number of key issues. In the first place, the degrees and levels of consciousness outlined by a number of commentators allow us to place AI applications such as ChatGPT somewhere along a spectrum of consciousness. The precise location of various forms of AI on a consciousness spectrum will necessarily be an open-ended question informed by the theoretical frameworks outlined above. If we are inclined towards a more physicalist account, the placement might be of the following:

*micro-organisms – lower animals – plants - primates (chimpanzees/bonobos) and mammals (dolphins/whales) – AI – humans*

Idealist versions, on the other hand might place AI lower on the spectrum:

*micro-organisms – lower animals – plants – AI – primates and mammals - humans*

However, noting the dangers of anthropocentrism mentioned earlier, it is feasible to place certain animals and plants at a level higher than humans in terms of animate intelligence (Lent, 2022) though, obviously humans would rate the highest place on conceptual intelligence since this is a specifically human construction.

The lower position of AI on the consciousness spectrum for idealist positions may be explained by the problematic nature of AI claims when considered within the monistic perspectives developed by Hoffman, Kastrup and similar commentators such as Steve Taylor (2019) and Philip Goff (2019). For Kastrup – who conceives the subjective consciousness of humans and animals as individual segments of a larger mental membrane – the computational /information-processing power of AI tools must always fall short of consciousness since such material objects are just constructions of human minds which themselves are the only recipients of universal consciousness. Kastrup, like Hoffman, does not dispute the fact that AI applications will potentially outstrip human capabilities in all spheres – including art and poetry as well as science and mathematics – but, as a computer scientist, he suggests that there is absolutely no justification for entertaining the hypothesis that objects made from silicon chips running on algorithms designed by us can achieve the consciousness associated with Nagel’s ‘what-is-like-to-be’ forms of subjective experience (Institute of Art and Ideas, IAI, 2023).

For Hoffman, we assess the consciousness of other beings through the portal provided by our evolutionary history. Thus, we gain hardly any information about the consciousness or otherwise of, say, rocks and minerals, a little information from invertebrates, even more from observing primates and mammals, and quite a lot from interacting with other humans as conscious agents. Hoffman is clear that the question does not turn on the difference between organic living organisms and inorganic synthetic objects, or between the intentions and motivations of humans as against the programmes of AI tools – all of this may be achievable through the complex algorithms of machine intelligence – but, rather, our interactions with other potential sources of consciousness in a universe of conscious agents (IAI, 2023). AI applications will be, like all other aspects of reality, part of our user interface which provides only partial and limited information about the ultimate nature and source of experience.

Broadly similar sentiments inform Goff’s panpsychist perspective which allows for many types and degrees of consciousness throughout the natural world. In terms of the mainstream Integrated Information Theory of consciousness the processing of computers might not ‘be integrated enough to give rise to consciousness’ in terms of subjective awareness yet, Goff speculates, ‘might such computers not be programmed to believe that they have feelings and experiences?’ (2019, p.103). His recent comments which claim that ChatGPT is not an example of consciousness at work indicates that he thinks that AI developments have not quite reached that stage (2023) but – especially in the light of the enormous speed of machine learning increments and in anticipation of major developments in quantum computing (Hyland, 2019) – both the Turing Test and the subjective experiential aspects of consciousness are potentially achievable.

Many of the panpsychist/neo-idealist theses are informed by a secular spirituality which, on the face of it, seems to exclude machine intelligence from its ambit. Writing about what he calls ‘spiritual science’, for example, Taylor (2019) comments that:

Rather than being just biological machines, human beings are, both mentally and physically, expressions of spirit, or consciousness. You can say that our physical bodies are an external

expression of universal consciousness, while our minds (or beings) are an inner expression (p.228).

Thus, on this account – though AI may outstrip human capabilities in all spheres – it would seem to be excluded from this spiritual realm. In a similar vein, Lent (2021) argues that ‘the most pervasive mistaken metaphor of life in common currency is that it’s merely a very complicated machine’ (p.142). He goes on to suggest that:

A fundamental error arising from this metaphor is the notion that our minds are ‘software’ that can be separated from our physical ‘hardware’. With computers, information is substrate independent, which means you can upload your files to the cloud, download them somewhere else, and they will be exactly the same. Life is not substrate independent. The so-called ‘information’ carried by your DNA can’t be separated by how it’s expressed within the cell; the ‘information’ transmitted through your brain is inextricably bound up in each neuron’s own molecular structure and its dynamic relationships with other neurons (ibid.,p.143).

Even though these forms of spiritual idealism do seem to exclude much of AI from the consciousness worldview, it is plausible – as I suggest later – to regard AI applications as incipient ‘non-human persons’ (Singer, 2009; Hyland, 2023) which, like higher primates and mammals, can be included in the moral community.

#### **4. Ethical Matters**

Given the widespread controversy surrounding recent AI developments surrounding Chatbot applications, the ethical – and more generally, the social, political and economic – implications of burgeoning machine intelligence must be considered as paramount in the current debates. As the public debate about the rapid advancement of AI becomes ever more hysterical a letter signed by – among others, Sam Altman, CEO of OpenAI, the Microsoft-backed AI research lab that is behind ChatGPT, and the so-called godfather of AI who recently left Google, Geoffrey Hinton – warned of the dangers in apocalyptic terms in the statement that:

Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war (Romo, 2023, p.1).

However, although it is important to take note of such warnings from experts in the field, Nello Cristianini (2023) has called for calm and clarity in the field. As he comments:

It is certainly true that, along with many benefits, this technology comes with risks that we need to take seriously. But none of the aforementioned scenarios seem to outline a specific pathway to extinction. This means we are left with a generic sense of alarm, without any possible actions we can take...Except for weaponisation, it is unclear how the other – still awful – risks could lead to the extinction of our species, and the burden of spelling it out is on those who claim it (pp.1-2).

The plea here is for greater clarity and precision and we can approximate to this by analysing the threats and opportunities within an ethical framework under the following headings.

##### *4.1. Economic Matters*

The chief concern in this area is that AI tools and applications will replace many jobs currently performed by humans thus leading to mass redundancy and unemployment. British

Telecom announced recently that it would be cutting its workforce by 55,000 with 11,000 of these jobs replaced by AI (BBC News, 2023). On a more dramatic scale, a recent report by Goldman Sachs predicted that around 300 million jobs would in future be lost or degraded by forms of AI (Kelly, 2023).

There are, however, alternative views of these developments which offer a less catastrophic and more hopeful vision of the future. A report by the World Economic Forum (2023), for example, makes the following important points about what commentators are calling the ‘fourth industrial revolution:

- Around 40% of all working hours could be impacted by AI large language models (LLMs) such as ChatGPT-4, says a report from Accenture.
- Many clerical or secretarial roles are seen as likely to decline quickly because of AI, according to the World Economic Forum's *Future of Jobs Report 2023*.
- But roles for AI and machine learning specialists, data analysts and scientists, and digital transformation specialists are expected to grow rapidly, the report adds.
- Reskilling people to use AI effectively will be the key to companies being able to use the technology successfully, says Accenture.

The key message here is that:

Success with generative AI requires an equal attention on people and training as it does on technology...This means both building talent in technical competencies like AI engineering and enterprise architecture, and training people across the organization to work effectively with AI-infused processes (p.1).

In a similar vein, Jonathan Aitken (2023) reminds us that:

“The development of technology and its associated impact on job security has been a recurring theme since the industrial revolution. Where mechanisation was once the cause of anxiety about job losses, today it is more capable AI algorithms. But for many or most categories of job, retaining humans will remain vital for the foreseeable future” Jonathan Aitken (2023).

**Aitken** goes on to suggest that:

This means that, as workers, we need to look to harness the capability of AI systems and use them to their full potential. This means always questioning what we receive from them, rather than just trusting their output blindly...If we apply a sceptical mindset to how we use this new tool, we’ll maximise its capability while simultaneously growing the workforce – as we’ve seen through all the previous industrial revolutions (pp.1-2).

This message of careful regulation, focus and planning will need to be applied to all impacts of AI technology, particularly in the socio-political, cultural and educational spheres.

#### *4.2.Social/Political Issues*

The flurry of concerned and critical activity referred to earlier surrounding the appearance of AI LLMs has been accompanied by feverish government action to keep pace with the rapid developments. The White House has issued a *Blueprint for an AI Bill of Rights* (2023) which

includes the following five fundamental principles to guide future implementation and use of the new AI applications (pp.3-6):

- You should be protected from unsafe or ineffective systems
- You should not face discrimination by algorithms and systems should be used and designed in an equitable way
- You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used
- You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you
- You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter

A similar set of policy principles have been proposed recently by the UK Government which have been described as a ‘pro-innovation approach’ designed to ensure that AI regulation does not interfere with investment in the new technologies. Albert Sanchez-Graells (2023) has pointed out that the ‘plans have been criticised for being too lax, already outdated, and lacking in meaningful detail’ (p.1). He suggests that the policy proposals have more to do with ensuring a post-Brexit AI technology market for Britain rather than protecting the public from potential harm, and concludes that:

Only by implementing effective protections and showing strong and decisive action domestically can the UK government hope to build the credibility needed to lead international efforts of AI regulation (ibid.,p.2).

All such proposals seem to be rather light-touch - relying too on the goodwill of AI technology companies and the ability of the public to understand the new developments – and there is a growing consensus that regulation will need to be tighter. In terms of public understanding, the role of schools and colleges discussed below will be crucial. On the matter of state legislation in this field the European Commission’s legal framework governing medicinal products for human use (2023) might serve as a potentially effective model in this respect. In addition, all the current legislation for public protection in the sphere of medical practice and pharmaceutical development – which all countries seek to implement and monitor – can be effectively applied to AI policy alongside the regulation of the leading companies in the machine intelligence field.

Though there may be disagreements about degrees of control and regulation of AI uses, few commentators disagree about the control of AI use in warfare. Several examples of near nuclear disasters avoided by humans are on record – a recent film was made about the most famous one, the Russian, Stanislav Petrov (Myre, 2018) – and given the range of drones, tanks, fighter planes and missiles currently operating with AI components, the control and regulation of such applications is vital. Both military bodies and tech experts are now fully aware of the dangers here, and there are urgent calls in the US for ‘Congress to put guardrails in place to ensure [AI] is not misused’ (Khalil, 2023).

#### *4.3. AI, Technology and Medicine*

Amidst the furore surrounding the recent introduction of Chatbots it is easy to overlook the fact that just about every aspect of modern life has for a long time been utterly dependent upon AI applications in one form or another. From laptops, smartphones, cash dispensers, home security devices to modern automobiles, contemporary life would be unthinkable without machine intelligence devices. Critical ethical issues concerning all this only tend to be highlighted when problems arise as, for example, social media is used for criminal or harmful purposes, or when systems are hacked leading to failures of public health or national security systems. The recent introduction of driverless cars has again foregrounded this human/AI interface, and Sanksshep Mahendra (2023) explains what is involved in order to achieve so-called ‘autonomous’ driving:

AI needs to plan and execute actions without the influence of a human driver. The AI is equipped to perform the same functions as a human driver. It has recognition and decision-making abilities, sensory functions, and the ability to model data with deep learning algorithms. Armed with these innovations, the AI-powered vehicle can perform autonomously (p.1).

Of course, such driverless cars are still on experimental trial and no doubt many iterative adaptations will have to be made to ensure safety on the roads, just as they did when automobiles first replaced horses on public highways.

In terms of the most promising utilisation of AI power, the role of AI in medicine and health care is, perhaps, the field which is most exciting for proponents. Liz Kwo (2021) has outlined a wide range of advances and gains made through the introduction of AI tools in healthcare settings – from drug discovery to rapid diagnosis – and concludes her review of the field by observing that by:

improving workflows and operations, assisting medical and nonmedical staff with repetitive tasks, supporting users in finding faster answers to inquiries, and developing innovative treatments and therapies, patients, payers, researchers and clinicians can all benefit from the use of AI in healthcare (p.1).

Just recently, it was announced that AI had been used ‘to discover *abaucin*, an effective drug against *A baumannii*, bacteria that can cause dangerous infections’ (Yang, 2023), and the range of opportunities for improving all aspects of medical treatment through AI expands exponentially (Davenport, 2019).

#### 4.4. Education and Culture

As indicated in the introduction, many of the concerns about the new AI tools have been expressed by educators who fear that teaching and learning will be damaged by the easy access to the large language models (LLMs) like ChatGPT which can write essays and answer assignment questions in a matter of minutes (Heaven, 2023). A letter to *The Times* signed by leading educators from the state and private sector described AI developments as ‘bewildering’ and announced the launch of a review body of experts to advise schools on which areas are ‘beneficial, and which are damaging’ since ‘we have no confidence that the large digital companies will be capable of regulating themselves in the interests of students, staff and schools’ (Shad, 2023, p.2). At college and university level, academics and

administrators have been scrambling to try to understand and manage the impacts of AI applications such as ChatGPT for students and staff.

Among the prophets of doom in this sphere, however, can be found more measured voices which suggest a productive partnership between academia and AI. Will Douglas Heaven (2023), for example, comments that initially educators were worried that ‘ChatGPT would undermine the way we test what students have learned, a cornerstone of education’. However, it seems that many teachers have now adapted to the new applications and discovered some positive ways of working with them. As he comments:

Far from being just a dream machine for cheaters, many teachers now believe, ChatGPT could actually help make education better. Advanced chatbots could be used as powerful classroom aids that make lessons more interactive, teach students media literacy, generate personalized lesson plans, save teachers time on admin, and more (pp.1-2).

Similarly, Karen Lancaster (2023) urges university lecturers to embrace AI applications such as ChatGPT working with students to eliminate errors and achieve the best results. She concludes her plea for a working partnership between AI tools and academia by observing that:

if universities accept the use of AI software for essay-writing, they should increase the expected level of scholarship accordingly, in the same way that maths tests for people *with* calculators should demand a higher level of aptitude than maths tests for people *without* calculators. But simply behaving as if the technology doesn’t exist, or decreeing that its use amounts to misconduct, is probably not a prudent way forward (p.4).

Similar arguments have been advanced by Claire Chen (2023) of Stanford University who reviews a wide range of insights which can be gained from working creatively with AI tools in ways which can enhance teaching and learning in many fields of learning. It is well to acknowledge all the increments in learning and development – in just about every sphere of activity – that have been made in recent decades through AI tools. The fact that humans have improved at the fiendishly complex game of Go since *DeepMind’s AlphaGo* finally defeated the world’s best players (Rosenblum, 2023) is just one small, perhaps emblematic, indication of what educators and policymakers might gain by working in partnership with AI applications.

## **5. Conclusion:**

The foregoing review of threats and opportunities inherent in AI applications has pointed towards a compatibilist perspective which would seek to avoid both catastrophic doomsday warnings and futuristic over-optimism. Clearly strict guardrails need to be in place to control AI uses in national defence, medicine and scientific experimentation but, as the EU Commissioner, Margrethe Vestager, commented recently, ‘although the existential risk from advances in AI may be a concern, it was unlikely, whereas discrimination from the technology was a real problem’. She told the BBC that ‘Probably [the risk of extinction] may exist, but I think the likelihood is quite small. I think the AI risks are more that people will be discriminated [against], they will not be seen as who they are’ (Milmo & Hern, 2023, p.1).

On the issue of the consciousness or otherwise of AI applications, the question is still an open one probably determined by prior philosophical stances on materialism, idealism, spirituality

and the nature of human sentience and awareness. As indicated earlier, although Chomsky and Goff are adamant that the new advances do not parallel human consciousness or reasoning, Bostrom is clear that emergent AI can be regarded as both sentient and super intelligent. As Peter Cave argues in his ingenious thought experiments in *Can a Robot Be Human?* (2007), because we have no access to minds other than our own, we are naturally sceptical about whether machines constructed by humans – no matter how much they are able to behave like humans in every possible way – are actually capable of our inner awareness and experience. Yet, how can we ever know this for certain since our only evidence for the consciousness of other people is precisely the observable criteria we would apply to AI applications?

However, although it is reasonable to maintain an agnostic position on the consciousness or otherwise of AI, the power of the new applications require us to take a stand on the ways they are applied in the social, economic, medical, educational and broad cultural spheres of life. In short, we need to consider the ethical framework of AI developments and, in this sphere, I would suggest that we can gain much by looking at Peter Singer's conception of 'non-human persons'. Initially, Singer wanted to argue that certain animals – higher primates and mammals in particular – should be brought into the moral community on the grounds that they were sentient beings who could suffer (2009). In recent work, he has examined the status of AI machines and robots in the light of the wider questions of rights and responsibilities surrounding the human/AI interface. Anticipating contemporary developments, Singer (2023) asks 'if machines can and do become conscious, will we take their feelings into account?' This is considered a pertinent **question since our treatment of the**

only non-human sentient beings we have encountered so far – animals – gives no ground for confidence that we would recognize sentient robots not just as items of property, but as beings with moral standing and interests that deserve consideration (p.382).

After all, like corporations and similar artefacts, there are solid grounds for granting the legal personhood of AI tools and applications. As Visa Kurki (2019) argues, there is no reason to doubt that an 'AI can function as a legal person, it can be granted legal personhood on somewhat similar grounds as a human collectivity' (p.175).

We have little to lose and much to gain from acknowledging the moral and legal status of AI and constructing working partnerships with applications and tools in education, science, and the workplace. It goes without saying that we must acknowledge and deal with what Bostrom calls the 'control problem' of superintelligent AI, that is, the building in of mechanisms which protect all stakeholders. He provides a comprehensive analysis of control methods and offers what he calls the 'common good principle' for the regulation of the human/AI interface which is that 'superintelligence should be developed only for the benefit of all humanity and in the service of widely shared ethical ideals' (2016, p.312).

## References

AI News (2023). <https://www.artificialintelligence-news.com/categories/ai-applications/ai-chatbots/> accessed 18.5.23

- Aitken, J. (2023). AI could threaten some jobs, but it is more likely to become our personal assistant . *The Conversation*, May 25, <https://theconversation.com/ai-could-threaten-some-jobs-but-it-is-more-likely-to-become-our-personal-assistant-206297> accessed 18.6.23
- Alhazmi, A.K., Alhammadi, F., Zain, A.A., Kaed, E., Ahmed, B. (2023). AI's Role and Application in Education: Systematic Review. In: Nagar, A.K., Singh Jat, D., Mishra, D.K., Joshi, A. (eds) *Intelligent Sustainable Systems. Lecture Notes in Networks and Systems*, Vol. 578. Singapore: Springer. [https://doi.org/10.1007/978-981-19-7660-5\\_1](https://doi.org/10.1007/978-981-19-7660-5_1) accessed 7.5.23
- Al-Sibai, N.(2023).Nick Bostrom says that AI Chatbots may have some degree of sentience. *The Byte*, April 13,<https://futurism.com/the-byte/nick-bostrom-ai-chatbot-sentience> accessed 29.4.23
- BBC (2023). BT to cut 55,000 jobs with up to a fifth replaced by AI. *BBC News*, May 18, <https://www.bbc.com/news/business-65631168>. Accessed 9.6.23
- Blackmore, S. (2011). *Zen and the Art of Consciousness*. (Oxford: Oneworld Pubs)
- Bostrom, N. (2016). *Superintelligence*. (Oxford: Oxford University Press)
- Bruenstrup, G. (2016). *Emergent Panpsychism*. <https://www.hfph.de/hochschule/lehrende/prof-dr-godehard-bruentrup-sj/articles/2016b.pdf> accessed 24.5.23
- Cave, P.(2007). *Can a Robot be Human?* (Oxford: Oneworld Publications)
- Chalmers, D. (1996) *The Conscious Mind*. (Oxford: Oxford University Press)
- Chen, C. (2023). AI will transform and teaching and learning. Let's get it right. *Stanford University AI+ Education Summit*, March 9. <https://hai.stanford.edu/news/ai-will-transform-teaching-and-learning-lets-get-it-right> accessed 3.4.23
- Chomsky, N. (2023). The False Promise of Chat GPT. *New York Times*, March 8, 2023
- Cianciolo, A.T. (2014). Intelligence: History and Controversies. In Phillips, D.C.(ed). *Encyclopedia of Educational Theory and Philosophy* (London: Sage) pp.417-421
- Cristianini, N.(2023). If we're going to label AI an 'extinction risk', we need to clarify how it could happen. *The Conversation*. May 31, <https://theconversation.com/if-were-going-to-label-ai-an-extinction-risk-we-need-to-clarify-how-it-could-happen-206738> accessed 3.6.23
- Davenport, T. (2019). The potential for artificial intelligence in healthcare. *National Library of Medicine*, 6(2): 94-98, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/> accessed 23.6.23
- Dennett, D.(1991). *Consciousness Explained*. (Boston: Little Brown)
- European Commission. (2023). Legal framework governing medicinal products for human use in the EU. *European Commission*. [https://health.ec.europa.eu/medicinal-products/legal-framework-governing-medicinal-products-human-use-eu\\_en](https://health.ec.europa.eu/medicinal-products/legal-framework-governing-medicinal-products-human-use-eu_en). Accessed 23.6.23

- 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 19.5.23)
- Future of Life Institute (2023). *Pause Giant AI Experiments – An Open Letter* <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (accessed 21.5.23)
- Gardner, H. (1993). *Multiple Intelligences: The Theory in Practice*. (Port Moody, BC: Pac-Rim Bookservices)
- Goff, P.(2019). *Galileo’s Error: Foundations for a New Science of Consciousness*. (London: Rider)
- 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 24.5.23)
- Harris, A.(2019). *Conscious: A Brief Guide to the Fundamental Mystery of the Mind*. (New York: Harper Collins)
- Heaven, W.D. (2023). ChatGPT is going to change education, not destroy it. *MIT Technology Review*, April 6, <https://www.technologyreview.com/2023/04/06/1071059/chatgpt-change-not-destroy-education-openai/>
- Hoffman, D. (2019). *The Case Against Reality: How Evolution Hid the Truth from Our Eyes* (London:Penguin)
- James, W. (1983 edn). *The Principles of Psychology*. (Cambridge: Harvard University Press)
- Horgan, J.(2021). Quantum Mechanics, the Chinese Room Experiment and the Limits of Understanding. *Scientific American*, March 9, <https://www.scientificamerican.com/article/quantum-mechanics-the-chinese-room-experiment-and-the-limits-of-understanding/> accessed 1.6.23
- Hyland, T. (2019). *Adventures in the Matrix*. Mauritius: Scholars’ Press
- 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/> accessed 15.6.23
- IAI (2023). Consciousness in the Machine: Kastrup, Schneider, Hoffman Debate on Consciousness. *Institute of Art and Ideas*. <https://www.youtube.com/watch?v=VmQXpKyUh4g> accessed 8.6.23
- 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. (2020). *Decoding Schopenhauer’s Metaphysics*. (Winchester: Iff Books)

- Kelly, J. (2023). Goldman Sachs Predicts 300 Million Jobs Will Be Lost Or Degraded By Artificial Intelligence. *Forbes*. May.31, <https://www.forbes.com/sites/jackkelly/2023/03/31/goldman-sachs-predicts-300-million-jobs-will-be-lost-or-degraded-by-artificial-intelligence/> accessed 12.6.23
- Khalil, J. (2023). Military, tech experts raise concerns about AI weaponization: ‘We have to be very concerned’ . *The Hill*. March 6, [https://thehill.com/homenews/nexstar-media\\_wire/4033091-military-tech-experts-raise-concerns-about-ai-weaponization-we-have-to-be-very-concerned/](https://thehill.com/homenews/nexstar-media_wire/4033091-military-tech-experts-raise-concerns-about-ai-weaponization-we-have-to-be-very-concerned/) accessed 15.6.23
- Kleppen, E. (2023). What is the Turing Test? *Built In*. January 13, <https://builtin.com/artificial-intelligence/turing-test> accessed 25.5.23
- Kurki, V.A.J. (2019). *The Legal Personhood of Artificial Intelligences*. (Oxford: Oxford Academic) <https://academic.oup.com/book/35026/chapter/298856312?login=false> accessed 16.6.23
- Kwo, L.(2021). Contributed: Top 10 Use Cases for AI in Healthcare . *MobiHealth News*. July 1, <https://www.mobihealthnews.com/news/contributed-top-10-use-cases-ai-healthcare> accessed 17.6.23
- Lancaster, K. (2023). Why universities should embrace AI essay-writing software. *PESGB Blog*, February 24. <https://www.philosophy-of-education.org/why-universities-should-embrace-ai-essay-writing-software/> accessed 5.5.23
- Lent, J. (2021). *The Web of Meaning*. (London: Profile Books)
- Mahendra, S.(2023). AI and Autonomous Driving. *Artificial Intelligence*. February 7, <https://www.aiplusinfo.com/blog/ai-and-autonomous-driving/> accessed 3.7.23
- Milmo, D. & Hern, A. (2023). Discrimination is a bigger AI risk than human extinction – EU commissioner. *The Guardian*, June 14, <https://www.theguardian.com/technology/2023/jun/14/ai-discrimination-is-a-bigger-risk-than-human-extinction-eu-chief> accessed 21.6.23
- Myre, G.(2018). Stanislav Petrov, 'The Man Who Saved The World,' Dies At 77. *NPR-The Two Way*. <https://www.npr.org/sections/thetwo-way/2017/09/18/551792129/stanislav-petrov-the-man-who-saved-the-world-dies-at-77> accessed 13.6.23
- Nagel, T. (1974). What is it like to be a bat?; *Philosophical Review*, 83, pp.435-450
- Romo, V.(2023). Leading experts warn of a risk of extinction from AI. *NPR Technology*. May 30, <https://www.npr.org/2023/05/30/1178943163/ai-risk-extinction-chatgpt> accessed 1.6.23
- Rosenblum, A. (2023). Humans have improved at Go since AIs became best in the world. *New Scientist*. March 13, <https://www.newscientist.com/article/2364137-humans-have-improved-at-go-since-ais-became-best-in-the-world/> accessed 7.4.23
- Sanchez-Graells, A. (2023). The UK wants to export its model of AI regulation, but it’s doubtful the world will want it. *The Conversation*. June 7, <https://theconversation.com/the->

[uk-wants-to-export-its-model-of-ai-regulation-but-its-doubtful-the-world-will-want-it-206956](#)  
accessed 15.6.23

Saunders, N. (2023). AI: evolution is making us treat it like a human, and we need to kick the habit. *The Conversation*. May 16, <https://theconversation.com/ai-evolution-is-making-us-treat-it-like-a-human-and-we-need-to-kick-the-habit-205010>

Searle, J.R. (2004). *Mind*. (Oxford: Oxford University Press)

Shad, N.(2023). Schools bewildered by AI advances, say head teachers. *BBC News*. <https://www.bbc.com/news/uk-65655010> accessed 20.5.23

Singer, P.(2009). *Animal Liberation*. New York: Harper Perennial Modern Classics

Singer, P.(2023). *Ethics and the Real World*. Princeton: Princeton University Press

Skribina, D. (2017). *Panpsychism in the West*. (Cambridge, MA: MIT Press)

Strawson, G. (2006). Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies*, 13 (10/11), pp.3-31; also Strawson, G. (2016). Consciousness isn't a mystery. It's Matter. *New York Times*, 16/5/16

Taylor, S. (2019). *Spiritual Science: Why Science Needs Spirituality to Make Sense of the World*. (London: Watkins Media)

Taylor, J. & Hern, A. (2023). 'Godfather of AI' Geoffrey Hinton quits Google and warns over dangers of misinformation. *The Guardian*, 2/5/23, <https://www.theguardian.com/technology/2023/may/02/geoffrey-hinton-godfather-of-ai-quits-google-warns-dangers-of-machine-learning> accessed 30.5.23

The White House. (2023). *Blueprint for an AI Bill of Rights*. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> accessed 11.6.23

Whitehead, A.N. (1978). *Process and Reality: An Essay in Cosmology*. (New York: Free Press)

World Economic Forum. (2023). *These are the jobs most likely to be lost – and created – because of AI*. May 4, <https://www.weforum.org/agenda/2023/05/jobs-lost-created-ai-gpt/> accessed 12.6.23

Yang, M.(2023). Scientists use AI to discover new antibiotic to treat deadly superbug. *The Guardian*. May 25, <https://www.theguardian.com/technology/2023/may/25/artificial-intelligence-antibiotic-deadly-superbug-hospital> accessed 28.6.23