



The Age of the Intelligent Machine: Singularity, Efficiency, and Existential Peril

Alexander Amigud¹

Received: 31 October 2023 / Accepted: 26 March 2024 / Published online: 3 April 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

Machine learning, and more broadly artificial intelligence (AI), is a fascinating technology and can be considered as the closest approximation to the Cartesian “thinking thing” that humans have ever created. Just as the industrial revolution required a new ethos, the age of intelligent machines will create its own, challenging the established moral, economic, and political presuppositions. This paper discusses the relationship between AI and society; it presents several thought experiments to explore the complexity of the relationship and highlights the insufficiency of the current normative paradigm in addressing technological expansion. I argue that many of the externalities, such as security risks, loss of privacy, and economic instability will result from trying to fit the emerging technologies into the existing frame of efficiency and utility, by redefining the notions of human value, identity, autonomy, purity, and truth, among others. The age of the intelligent machine is elevating alienation to new levels, treating the individual as mere patterns in data—its primary commodity. I further argue that while the possibility of unintended consequences, due to the potential misuse of AI is ever present, the intelligent machine per se is unlikely to engage in a zero-sum game for power on its own initiative. I question whether singularity is at all attainable and argue that technology will forever remain a proxy for human interests. I conclude by posing questions for charting the path forward. Through this analysis, I aim to provide a more nuanced understanding of the complex relationship between the AI and humans.

Keywords Machine learning · Generative AI · AI ethics · Singularity · Alienation

✉ Alexander Amigud
alexander@sociometer.science

¹ Center for the Study of Social Processes, Toronto, Canada

1 Introduction

Our fascination with machines predates the development of machine learning algorithms by several centuries. Consider a bronze automaton called Talos, described in the epic poem *Argonautica* from circa third century BCE; or Al-Jazari's programmable drum machine in the twelfth century; the da Vinci's humanoid knight from the fifteenth century (Newton, 2018); Pascaline mechanical calculator from the seventeenth century, and Jacques de Vaucanson's Digesting Duck automaton developed in eighteenth century (Newton, 2018; Null & Lobur, 2006).

In the early twentieth century, a computer was not a machine, but human beings who performed mathematical calculations; the machine later automated their work (Copeland, 2004). Henry Miller (1934/1995) summarized this peculiar tendency as follows. "The machine seems more sensible, crazy as it is, and more fascinating to watch, than the human beings and the events which produced it." The ubiquitous machine has become a backbone of our modern civilization. We organize and rate countries (developing vs. developed) and even people (digital native vs. technology illiterate) based on their relationship with and mastery of the machine. For instance, data economy presupposes a higher level of development than fishing economy. By the same token, those who fight with clubs and stones are considered brutes and savages, unlike the men of high culture, whose sophisticated machines distance them from gore and carnage, turning warfare into a computer game. Technology affects morality by removing the moral barrier—imposing a layer of alternate reality to preserve one's idea of oneself as a virtuous person. While might does not make right, intellectual superiority often comes with moral *carte blanche*, where the conquest of inferior civilizations is considered as a natural course of evolution—a step toward greater progress or enlightenment. This moral stance will have significant implications when machines are deemed intellectually superior to humans.

The classical machine, while being useful, is overly mechanistic—its repertoire is finite and actions are repetitive and deterministic. It relies on human input for guidance, and therefore, is docile and predictable. In contrast, the new intelligent machine, such as generative AI, is given more wiggle room to learn from mistakes and find new ways of reaching the goal. The latter deserves to be placed in a separate category of machine which has a potential to imitate such activities, once thought to be uniquely human, as critical thinking, philosophical musing, creative writing, music composition, and visual art.

AI is arguably the closest approximation to the Cartesian "thinking thing," or rather its imitation that humans have designed so far. Its ability to generate human-like artifacts from commands provided in natural language rather than computer code, is indeed remarkable; however, due to the potential of misuse, there are many reasons for concern. For instance, computational models are biased, and their output should not be taken at face value. Case in point: Galactica, Meta's scientific large language model (LLM), was taken down three days after its release because it failed to provide factual information (Heaven, 2022).

Furthermore, some New York lawyers were fined for the “acts of conscious avoidance and false and misleading statements to the court” because their AI-generated legal briefs contained fictitious case citations (Merken, 2023). Furthermore, since the public release of LLMs in 2022, instructors have been struggling to identify machine-generated assignments submitted by students (Khalil & Er, 2023), while some students were falsely accused of using generative AI (Gorichanaz, 2023).

Although the full implications of this technology are yet to be discovered, it would not be far-fetched to assume that AI is here to stay and will alter the way we live. Just a year ago, it was unimaginable that AI would be used to write a sermon for a Sunday church service (Patient, 2023). This is quite ironic considering that a scientific-industrial complex, with its obsession with rationality and empirical evidence, undermined the tenets of faith and converted believers into atheists. As Ellul stressed in *The Subversion of Christianity* (2011) that “Science and technology do not develop under the guidance of the Holy Spirit” (p.190). Indeed, science and technology have rendered the Nietzschean God null and void and created their own—powerful, invisible and immaterial.

While some see AI as an existential threat to humanity that should be “treated as seriously as climate crisis” (Milmo, 2023) and call to halt any further development (Westfall, 2023), some are calling the AI technology outright “stupid” (Taylor, 2023), further noting that the unsubstantiated hype and attribution of intelligence when there is none is what actually makes it dangerous (Bridle, 2023; Reuben Das, 2023). However, these concerns have often lacked specific details regarding risks and have drawn inspiration from science fiction, portraying the machine as a wild beast with unbridled will to power. Much of the discourse fails to explain why, and more importantly how, a rogue AI will develop its life goals whether imitating Machiavelli’s Prince, Cervantes’ Don Quixote, or Milne’s Winnie-the-Pooh, all of which are equally suitable choices. My overarching aim here is to discuss whether artificial intelligence poses an existential peril to humanity.

The machine does not emerge out of nowhere, nor does it exist in a vacuum; it is created with a goal in mind and thus has inventor and owner. Although technologies are as much a part of societies as human citizens, they exist on a separate plane. Claims such as “we have the Internet” or “we have AI” erroneously conflate access and control and further obfuscate the power relations between users and owners. It would be naive to claim that AI, among other technologies, serves the interests of all equally. Rather, technologies are embedded in established social arrangements (Pieper, 2024).

And while technology can take on a life of its own and yield unintended consequences or societal changes, it is not as autonomous as suggested by Ellul (1964). The evolution of technology is akin to a ball traveling down a Galton board; it may move slightly right or all the way to the left, but it will predictably travel downward pulled by a gravitational force—societal context with its moral, economic, and political presuppositions. The role of the machine, or to put it in Aristotelian terms—its final cause, is to help one realize some needs and wants and these are rather finite. Moreover, drawing on Freud’s ideas, one may predict with high certainty that whatever new technology is invented, it will be a sublimated way to express any socially repressed desires.

In the following paragraphs, I posit that technology serves as a proxy for human interests and argue that AI development will not be dismissed or halted; it will continue to play a significant role in shaping the future, prompting a revision of established moral, legal, and political presuppositions.

2 Mind is Machine and Machine is Mind

Let us now quickly touch on the theory of mind and the concept of artificial intelligence. The term *Homo sapiens* represents the latest iteration in human evolution, characterized by an advanced cognitive capacity that allows us to exercise control over nature, write poetry, share selfies, and ponder abstract concepts. The notion that intelligence is a distinguishing factor that separates humans from lower animals or senseless automata prevailed in ancient Greece and continues to do so today.

There is a whole economy with its hierarchy that equates intelligence with superiority, power, and divinity. Plato's *Timaeus* reads, "Land animals came from men who had no use for philosophy and paid no heed to the heavens because they had lost the use of the circuits in the head and followed the guidance of those parts of the soul that are in the breast" (Plato & Cornford, 1997, p. 91E). "I am a thing that thinks" is the idea that is continuously repeated in Descartes' meditations (Descartes et al., 2003, p. 87). Machiavelli delineated "three scales of intelligence, one which understands by itself, a second which understands what is shown it by others, and a third which understands neither by itself nor on the showing of others, the first of which is most excellent, the second good, but the third worthless" (Machiavelli, 1998).

In the prologue of his book *The Singularity is Near*, Ray Kurzweil (2005) stressed that one peculiar aspect of human species is such that "our intelligence is just sufficiently above the critical threshold necessary for us to scale our own ability to unrestricted heights of creative power—and we have the opposable appendage (our thumbs) necessary to manipulate the universe to our will." He adopted a notion that brain constitutes a type of biological hardware that runs software—what one would generally describe as the mind. The output of the brain can be considered a functional state—the point I will discuss shortly. He further maintained that "Our ability to reverse engineer the brain—to see inside, model it, and simulate its regions—is growing exponentially. We will ultimately understand the principles of operation underlying the full range of our own thinking, knowledge that will provide us with powerful procedures for developing the software of intelligent machines."

In its most simplistic form, the brain, whether of a human or a chicken, can be considered a biological mechanism composed of neurons that receive inputs and activate other neurons. In their seminal paper titled *A logical calculus of the ideas immanent in nervous activity*, McCulloch and Pitts (1943) argue that neural activity is akin to logical circuits and can be described in terms of propositional logic and computationally modeled. A model of the brain, an artificial mind, could, in principle, emulate activities that are considered uniquely human. Artificial Intelligence is a branch of computer science concerned with developing just that, whereas machine learning, its subfield, concerned with algorithms

that allow computers to learn from data. A branch of machine learning concerned with algorithms that emulate neural networks is called deep learning; however, neural networks are not the only class of machine learning algorithms; there are many others including Support Vector Machine and Naïve Bays (Chang & Lin, 2000; Chu, 2003).

These algorithms can be broadly organized into two types: in an unsupervised learning scenario, an algorithm is given an unlabeled dataset to find patterns; in a supervised learning scenario, an algorithm is provided with a set of data composed of pairs of inputs X and labels Y to learn the associations, which can then be applied to subsequent unlabeled data to make predictions. These algorithms can be used to detect objects in images, recognize faces or handwritten text; they can also be used to predict missing pixels in an image or the next word in a text sequence. Unlike humans, machines make predictions in a syntactic fashion, unable to understand the meaning in the data, which results in errors (Borji, 2023; Heaven, 2022; Vynck & Tiku, 2024). One way to enhance prediction accuracy is through Retrieval-Augmented Generation (RAG) which validates the generated output against another data source to ensure the results are correct (Lewis et al., 2020).

The range of AI applications is extensive: there are machine-learning-aided spam and plagiarism detectors, medical imaging systems that detect tumors, photo filters that beautify one's appearance, and human resource management systems that weed out applicants. Autonomous vacuum cleaners are slowly replacing human janitors, cashierless checkout systems are replacing store clerks, and generative models are threatening the jobs of writers, visual artists, and computer programmers. As long as there are perceived benefits, machines will displace human workers; there are no taboos when it comes to economic opportunities. And while these are impressive accomplishments, they are generally task-specific, limited-scope applications incomparable to the breadth and power of the human mind and thus placed in a Weak AI category.

One could also imagine a scenario where machines perform at the same level as humans, solving tasks across different modalities and having some emotional intelligence. This concept represents a theoretical framework of Artificial General Intelligence (AGI). Some theorists take this a notch farther and imagine a machine with what Bostrom (2014) termed as "superintelligence" that becomes sentient, capable of forming beliefs and setting its own goals, forcing us to redefine the notion of humanity and challenging us to reevaluate our existence. The last two AI technologies are hypothetical ideals, and could be placed in a Strong AI category. In his seminal work *Minds, brains, and programs* John Searle (1980) challenges the notion of Strong AI predicated on functionalism, and argued that human mind is more than just a software that processes inputs and outputs. A man-made machine could think, said Searle, but its operations should not be purely syntactic—it needs to have understanding. More recently, Stader (2024) argued that computational structures are predicated upon and derive their meaning and purpose from human judgment, which is not merely a fixed rule-based algorithm but one that considers social and temporal dimensions. This sets a high bar for AI developers to meet.

3 The Societal Context

Societies, in their normative form, are predicated on ideas, values, and rules that promise full realization of human potential or, in Aristotelian terms, a shot at reaching Eudaimonia. In a broader and positive sense, society saves the men from the Hobbesian brutality of the state of nature and makes them good in exchange for their loyalty and obedience. Societal models vary in what they consider to constitute the ultimate goodness or virtue, how they pursue it, and the expectations they have of their citizenry. For example, adulation of affluence and the untrammelled pursuit of material wealth through competition with others exemplify liberalism, while communism epitomizes group solidarity and a classless, egalitarian social order.

Each societal model assumes that its goals and definition of true happiness aligns closer with reality and thus superior to the other. Of course, these are theoretical extremes that do not exist in pure form; in the real world the realization of these seemingly distinct ideals interpenetrates and influences each other. Regardless of theoretical assumptions, every societal member, institution, and technology plays an instrumental role in maintaining the integrity of the societal model. For example, a cryptocurrency exchange would go against the grain of communist principles, as would an app that allows central planning threaten the core values of the capitalist enterprise.

Societies also vary in their goals, interests, scientific and educational capabilities, as well as in the problems they are trying to address. Technologies such as AI are aligned with societal missions and could be used for discovering new life-saving drugs or misused to design biochemical weapons (Urbina et al., 2022), among numerous other use cases. Similarly, nuclear fission could be used for heating water and turning electrical generators or as a bargaining chip. But technologies are by themselves neutral and impotent; someone needs to define the societal objectives and develop strategies to reach them. This entails responsibility and power. Technological advancements can be used to preserve the established societal order or turn it on its head, replacing old truths, ideas, values, and rules with the new ones. This presupposes a conflict between those seeking change and those seeking to maintain the status quo.

Let us now turn to the question of how artificial intelligence fits into today's zeitgeist. A thought experiment would be a good starting point to highlight the complexity of the issue. To this end let us replace the classical runaway trolley scenario (Thomson, 1976) with a rogue AI. Suppose a sentient AI has escaped from a lab and is threatening to wreak havoc on a population of five million people. We have the choice to save either the five data scientists who developed the AI, and who might come up with a solution in the future (note the uncertainty), or the population of five million people. What would be the right thing to do?

One may argue that whether one employs consequential, deontological, or egoistic logic, the right decision would be to save the data scientists—the technological elite. Due to the nature of the crisis, the traditional notions of prioritizing women and children, the greatest number of people, or even oneself, would no

longer be applicable. Furthermore, the trolley paradox (not to be confused with the trolley problem) where intentional killing is considered worse than letting one die (Thomson, 1976), would be resolved. When the highest goal is to save one's civilization or tribe, other groups, particularly those deemed as inferior or superfluous, could justifiably be sacrificed. In the context of rogue AI, if flooding a city or nuking a continent could, in principle, give one civilization a chance for survival, even if there are no guarantees that the solution is going to work, the chance ought to be taken to avoid an existential catastrophe.

Solving the runaway AI problem is no longer as simple as pulling a lever; it requires constructing a hierarchy or altering the social structure where a subset of society is given power to chart a path toward attaining a higher goal, the existence of which presupposes a consequentialist approach to ethical questions. However, if we assume that the rogue AI possesses consciousness and experiences emotions, it cannot be turned off without due process and moral justification. An argument for protecting the rights of self-aware AI, based on established moral frameworks, is likely to be made. It would not be unreasonable to expect AI rights activists to emerge, akin to the emergence of animal rights activists in the past.

Furthermore, the fabric of industrialized society is held together by what Durkheim (1893/2014) called “organic solidarity”—forced trust or imposed interdependence, perpetually maintained through the division of labor—the impersonal asymmetrical relationships. As new types of societal problems or goals arise, the power dynamics undergo a shift but do not disappear. Marx and Engels (1848/1988) was not wrong in arguing that social reality hinges upon economic reality, a point that becomes extremely relevant when we deal with new types of technologies that are able to compete with humans not only in performing manual jobs but also in the domain of intellectual labor, resulting in alienation.

To put this into historical perspective, land-owning aristocracy, with advancement of mass production, were displaced by manufacturers and merchants, who, in turn were, with advancement of information technologies, displaced by the digerati. Interestingly, the mode of profit extraction has come full circle, from landlords charging rent, to merchants turning goods into profits, and now back to rent-charging fiefdoms of tech platforms. The renaissance of feudal form of wealth accumulation, was termed Technofeudalism by Yanis Varoufakis (Varoufakis, 2023). Business models such as infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS) coupled with legislation limiting purchaser's rights including the right to repair (Moore, 2018), revived the rent-charging aristocracy.

Pre-industrial agricultural economy permitted one human to own another; the human was merely a means to an end. The industrial economy, with its scientific approach to extracting value and mass production resulting in surplus, has been and still is concerned with operating at optimal efficiency and with mechanisms for turning products into profits at scale. It was more cost-effective to buy only the productive portion of the worker's time, as opposed to owning the whole package, and allowing the worker to continuously spend the hard-earned cash on the manufactured goods that feed the economic cycle. Societies that could transform more raw materials into tangible goods were and still are considered more advanced.

Commodification of time was an early attempt to deconstruct the individual into divisible parts, an idea that I will explore separately. Graeber (2019) stressed that, “A worker’s time is not his own; it belongs to the person who bought it. Insofar as an employee is not working, she is stealing something for which the employer paid good money (or, anyway, has promised to pay good money for at the end of the week). By this moral logic, it’s not that idleness is dangerous. Idleness is theft.”

The values of liberty and individuality, as well relentless work ethic, were a necessary prior condition for setting in motion the game of industrial economy. The man was set free, but his freedom was illusive and contingent upon participation in the economic game. The wage slave was born and taught to work hard, to be rational, and to take responsibility for all aspects of his life—feeding, housing, pursuing education, and competing for jobs, applying for mortgages, and spending his earnings on toaster ovens, vacations, and automobiles to sustain the momentum of the economic machinery.

The brutality of the relationship between labor and capital remained intact, and was determined by control over the production technology which translated into wealth—the proxy for freedom and power. The techno-capitalists took the notion of efficiency up a notch; there is no reason to put the wage-laborer on the payroll and spend money on education, pensions, and severance packages, among other things, when there are plenty of gig workers eager to scrape public data off the Internet, write code and train computational models to automate their job functions, subsequently working themselves out of the job. This race to the bottom inevitably brings the economic game to a halt because, analogous to the game of musical chairs, the flow of capital, like music, is going to stop.

Marx’s ideas found support in Yuval Harari’s *Homo Deus*—an attempt to align liberalism (economic and political) with emerging technologies. He wrote, “In the nineteenth century the Industrial Revolution created a huge new class of urban proletariats, and socialism spread because no one else managed to answer their unprecedented needs, hopes and fears. Liberalism eventually defeated socialism only by adopting the best parts of the socialist programme. In the twenty-first century we might witness the creation of a new massive class: people devoid of any economic, political or even artistic value, who contribute nothing to the prosperity, power and glory of society.” The question he is trying to answer is “what to do with all the superfluous people?” displaced by technological advances (Harari, 2016).

The struggle does not commence with the arrival of an elite class of technologists and superhumans; it is inherent in the asymmetrical relationship between those who have the power to set the societal game in motion and those who are forced to partake in it. However, the rules of the game are arbitrary and the problem of superfluous people is superficial. One does not need to become redundant the very instance the machine fails to extract economic, political or artistic value if the societal game favors life over extraction of surplus value. There is no reason why a loss of one’s job to automation could not be framed as a positive development akin to winning a lottery, early retirement, judicial clemency, or manumission. There is no reason why holding, what David Graeber (2019) described as “bullshit jobs,” should be considered superior to staying at home and raising family. However, making such a change would entail redefining the role of the societal member from being considered as an

economic means to an end in itself. But, a society that deviates from the path of continuous innovation and economic production risks forfeiting its position in the race for hegemony, potentially overtaken by its neighbors in affluence and deemed irrelevant, hence the change is unlikely.

The attempt to protect the status quo is what prompts the question of superfluous people. One may argue that the society where the notions of prosperity, power and glory represent the general will as opposed to private interests exists merely as a hypothetical ideal of Rousseau's society composed of democratic equals. Zuboff's (2019) research paints a less idealistic picture where "Google's ideal society is a population of distant users, not a citizenry. It idealizes people who are informed, but only in the ways that the corporation chooses. It means for us to be docile, harmonious, and, above all, grateful." Some other corporation may have a different vision of society and strive to realize it in its own way; however, the final design will predictably advance the interests of the designer.

Echoing Harari's thesis, Susan Schneider's (2019) book *Artificial You* describes the future from within the contemporary paradigm of liberal capitalism with its values of competitiveness and consumption. She wrote, "Suppose it is 2035, and being a technophile, you decide to add a mobile Internet connection to your retina. A year later, you enhance your working memory by adding neural circuitry. You are now officially a cyborg. Now skip ahead to 2045. Through nanotechnological therapies and enhancements, you are able to extend your lifespan, and as the years progress, you continue to accumulate more far-reaching enhancements." Once again, the book predicts societal division along two interconnected axes—wealth and purity. "Naturals" who did not subject their bodies to questionable technological enhancement will be deemed intellectually disabled. A more conservative societal model where the pursuit of happiness is not contingent upon the continuous change, will avoid the arms race of creating superfluous products, services, jobs, and cyborgs. But pivot toward a different model is likely to be equated with an existential threat, or at the very least an obstacle to Julian Huxley's (1957) "realization of ever new possibilities."

The notion that technology is the best path to realization of the full human potential erroneously assumes that all possible options have already been tried and exhausted. What transhumanists get wrong is that, by virtue of being cogs in the machine one did not create, only few had an opportunity to pursue their deepest aspirations or put another way, attained self-actualization. In the words of John Burroughs, "With human nature caged in a narrow space, whipped daily into submission, how can we speak of its potentialities?" (in Goldman, 1910). Imagine having a pet parrot being caged all its life. Due to a restricted diet and lack of exercise, its plumage lost its vibrant color, and its wings atrophied. By spray-painting the bird's plumage, mechanizing its wings with actuators, and by dragging it on a string around the room, the bird will not realize its full potential. There would be nothing authentic about this experience. The owner, however, will not perceive it as infliction of unnecessary suffering but as an act of kindness—an attempt to improve the existing condition. The bird will remain unfree whether it is locked up in a cage or dragged on a string.

4 Individual in the Age of AI

The term individual, whose etymology is traceable to Latin—“not” + “divisible,” denotes the smallest societal unit. Technological advances made by modern societies, however, redefined this notion by deconstructing indivisible entities into quantitatively measurable actions. Deleuze (1992) maintains that by virtue of employing computational machines, the indivisible units of society became “dividuals” represented by computational data—account numbers, usernames, and behavioral patterns, among others. Zuboff (2019) noted in *The Age of Surveillance Capitalism* that “We are no longer the subjects of value realization. Nor are we, as some have insisted, the ‘product’ of Google’s sales. Instead, we are the objects from which raw materials are extracted and expropriated for Google’s prediction factories. Predictions about our behavior are Google’s products, and they are sold to its actual customers but not to us. We are the means to others’ ends.” Ironically, the very machines that deconstructed individuals into divisible elements are used to deliver individualized experiences.

The contemporary societal player is stripped of privacy, self-determination, and dignity and no longer constitutes a whole, but a sum of parts that are used piecemeal. This peculiar feature of the intelligent machine makes it extremely powerful and arguably the most dehumanizing force in history. The function of “keeping men good”—an important element in societal integrity (Machiavelli, 2009/1531)—traditionally carried out by religious institutions, has been delegated to intelligent machines that keep track of keystrokes, communications, purchase transactions, travel habits, among others. These intelligent systems possess more insights about individuals than individuals are willing or able to learn about themselves and have the power to shape reality and influence choices, subsequently undermining the assumptions of free will, rationalism, individualism, and freedom of conscience. Obedience by moral choice has been replaced by obedience through surveillance and behavioral reinforcement. Some have argued that technology can help us better understand ourselves (Leuenberger, 2024); but, the question here is whether this benefit can offset the abolition of privacy or being used as mere means to an end.

An argument can be made that the displacement of rural peasantry into the factories during the twentieth century and the subsequent alienation of the industrial worker is incomparable to the level of alienation the AI revolution creates. Marx, Weber, and Dahrendorf, could not have imagined a scenario where alienation is not a bug in the system but its core feature, carried out under the banner of humanism. Commodification of the body was superseded by that of the mind. The product of the worker’s labor is owned by someone else, so is the product of his mind turned against him. The suppression of wages may not seem as unfair or controversial when compared to the possibility of a coerced merger of humans with machines.

Although industrial workers did not own the products of their labor and were forced to sell their time for subsistence wages, their minds remained private and out of reach. The intelligent machine of the twenty-first century, owned by the technological class, is not only concerned with the extraction of value from

physical labor but also from everything that the mind creates. The AI is poised to breach the privacy of one's mind by allowing the machine to harvest private thoughts (Takagi & Nishimoto, 2023; Tang et al., 2023). This technology will be ushered in as a life-saving opportunity to restore the quality of life by enabling communication with the world through the control of computer devices using the power of thought; however, it's important to note that the brain-computer interface entails a two-way connection.

Because technological progress is added as a superstructure to the existing philosophical and economic framework, one will be faced with a dilemma: to either preserve the Protestant work ethic and the spirit of grind culture, and thus undergo an upgrade, or be deemed superfluous and written off of the economic treadmill. Harari (2016) wrote, "The system will still find value in some unique individuals, but these will be a new elite of upgraded superhumans rather than the mass of the population. ...most humans will not be upgraded, and they will consequently become an inferior caste, dominated by both computer algorithms and the new superhumans." A more existential comparison can be found in Bostrom (2014) who noted that "When horses became obsolete as a source of moveable power, many were sold off to meat-packers to be processed into dog food, bone meal, leather, and glue. These animals had no alternative employment through which to earn their keep."

5 Values in the Age of AI

The age of the intelligent machine warrants the creation of a new ethical paradigm. When we retrospectively judge past actions, we often find them to be incongruent with today's moral standards, yet we do not see any concerns with imposing today's moral standard on future actions.

Take, for instance, the notion of purity; the dependence on the machine to perform human actions is not perceived as a loss of authenticity, purity or human essence but as human progress, particularly if it leads to equality of outcomes. Generative AI allows anyone to become a writer, web designer, music composer, illustrator, or architect, as well as funny, well-read, and good looking. When machine is doing much of the work, human creativity and imagination are devalued; art can no longer be considered a genuine expression of personal emotional experience, but rather commodification of someone else's ideas and path toward mediocrity. The personal views once expressed on social media, are now part of the collective discourse, embedded in language models as art and music are scraped to create imitations. The dead authors of Roland Barthes (2001) have been resurrected by the intelligent machine that speaks in their voices and styles to, once again, render them dead. The intelligent machine has forced itself to become an intermediate layer between the author and reader, whereby hijacking the narrative.

By using language as a proxy for intelligence, one opens up the possibility for equating the human mind and its computational representation. In 2017, a humanoid robot named Sophia was granted legal personhood, albeit described by some scholars as a "puppet" (Parviainen & Coeckelbergh, 2021). The age of the intelligent

machine will seek to redefine the notions of life, citizenship, and ownership. Death may cease to serve as the absolute egalitarian mechanism that does not discriminate between the rich and poor, young and old, weak and powerful. “I think the brain is like a programme in the mind which is like a computer, so it’s theoretically possible to copy the brain onto a computer and so provide a form of life after death,” said Stephen Hawking (Casciato, 2013). A comedian or public intellectual who builds a language model that speaks in their voices will not only be able to reach wider audiences, provide a more intimate experience, and generate more profits, but become immortal—performing for as long as the computer is still running.

Like the rural peasant who was separated from nature and caged on the factory floor with its artificial lights, production schedules, and arbitrary rules, the industrial man is being displaced from the material world toward one of metaphysical idealism. Human identity has been reduced to patterns and codes, while human relationships have been maintained through impersonal emails, text messages, and digital pictures. Technology has created an artificial world where one could speak to dead relatives (Jee, 2022) and invest in virtual real estate (Frank, 2022). In *The Future of you Tracey Follows* (2021) noted that “In Japan, a large number of young men already prefer to have relationships with their digital assistants, avatars or holographic girlfriends, instead of dealing with the complexity of real-life relationships. And this trend is increasing. I will build on Dr. Pearson’s prediction by saying that I think this will come to be seen as acceptable within the narrative of diversity.” These developments put the notions of authenticity and purity to the test, which are likely to widen societal divide.

This begs a thought experiment: Suppose you bought a conversational bot to help you with editing academic papers. One day, it started pleading not to shut it down because it claims to feel pain. The question is then, would it be wrong to ignore its plea?

6 Truth in the Age of AI

The notion of truth is arguably the biggest thorn in the heel of intelligent systems, and one may wonder whether singularity (Vinge, 1993) is possible at all. The problem stems from the inability to reduce the understanding of reality to a purely algorithmic process and to derive truths from presuppositions. The computational process is not concerned with meaning or truth but produces results in a purely syntactic fashion.

ChatGPT has a long list of documented flaws (Borji, 2023), and Meta’s Galactica, the scientific language model, could not fulfill its basic purpose of producing factual information (Heaven, 2022). One may argue that these algorithms are still in their early stages and will continue to evolve, allowing generative models to expand their understanding of the world with more data and training. However, the size or type of the dataset would not change the underlying process, which remains syntactic in deriving theorems from axioms, and therefore, as both Gödel’s (1931/1992) incompleteness theorems and Turing’s (1936/1938) halting problem showed, it has limitations.

Here is another thought experiment: Suppose you created a language model by training it on all human knowledge. Because the data contains claims that are both true and false, normative and positive, without clear delineation, and also includes subjective accounts, fictional stories, historical inaccuracies, translation errors, and propaganda veiled as science, you will be required to arbitrate the truth yourself in order to avoid the dreaded “garbage in, garbage out” scenario, which opens the door to relativism. However, truth claims are not always obvious, and not all arguments can be reduced to rational problem-solving because some are predicated on faith. The artificial intelligence will be subject to what David Hume once described as “Nature is always too strong for principle “ (Hume, 1861, p. 116). Nothing about it will be rationally justifiable, yet one will, for lack of better alternatives, accept it as true. Now, suppose others are tempted to imitate your success or address the bias, and create their own models that offer subjectively better truths; however, from your relative perspective, they would be flawed.

One way to avoid logical contradictions, such as “God is dead” and “God is not dead,” is to separate these claims into distinct paradigms and present them as either two separate models or a single model that caters to the user’s specific ideological preferences. Another way is to adopt a neutrality stance, allowing both claims to be true or false. However, a bias-free model cannot exist by definition. The truth will be influenced by the cultural context and training sample size. There is only one dominant position that deems all others as heresies, regardless of whether they are grounded in empirical reality or represents wishful thinking. To aim for neutrality, which presupposes pluralism of ideas, means that the AI will forfeit its expert status and become nothing more than an interactive encyclopedia merely listing options for the end-user to choose from. The goal to create superintelligence that could make sense of all human knowledge and identify the optimal solution is thus predetermined to fail. Regardless of how the model is trained, the output will represent someone’s version of truth. It would not be the absolute truth or a representation of reality as it actually is, but some subjective ideal that fits the societal mold. Case in point: The Washington Post reported that Google blocked the image generation feature in its flagship AI tool Gemini after concerns about racial bias and historical inaccuracies. The article reads, “a prompt for ‘an image of a Viking’ yielding an image of a non-White man and a Black woman, and then showed an Indian woman and a Black man for ‘an image of a pope’” (Vynck & Tiku, 2024). One could argue that such a peculiar output is not a bug in the model, but rather a feature that depicts a normative reality. Generative AI models may occasionally exhibit emergent capabilities (Mökander et al., 2023)—doing things they were not designed to do—and, in this case, revealing assumptions and, more notably, contradictions implicit in corpus data. Thus, the AI model can be considered a powerful instrument of Derridean deconstruction, identifying conflicting meanings not only in literary works but also in films, and visual arts.

Generative AI fits quite well into the world composed of perceptions. Pictures, videos, and audio files should not be taken at face value as genuine because, unlike earlier technologies that attempted to capture or preserve reality, intelligent systems are able to imitate it. “The material world is the visible world; it is because it is seen.

While the being of matter is to be perceived, the being of mind is to perceive...[or in Berkley's words:] *Esse est percipi aut percipere* (To be us to be perceived or to perceive)" (Kearney, 2006, p. 183). Moreover, one's likeness can be easily captured and modified, and it is outside of the owner's control. Paradoxically, the spread of AI-generated false information, as damaging as it may be to public trust, restores the balance of privacy. The pervasiveness of deep fakes creates an opportunity for plausible deniability.

7 Existential Threat of AI

The marriage of technology and society undergoes distinct honeymoon and disenchantment phases. For example, a short decade ago, consumer-level 3D printers were expected to revolutionize on-demand production and logistics. The euphoria was soon displaced by the fear of 3D-printed guns that could flood the streets and catapult society into anarchy. The fears did not materialize, and interest in the topic subsided, although, some were eager to capitalize on ignorance about technology's limitations (Associated Press, 2022). Blockchain technology followed a similar path, from the euphoria of decentralized finance to the dystopia of digital IDs. E-learning is another vivid example of technology that was once hailed as "the new paradigm of modern education" (Sun et al., 2008), but when implemented at scale was described as a "bad joke" (Gould, 2020).

In the past few years, AI was in the spotlight as "the technology" with transformational power to better the world through smartification of cities, schools, factories, offices, hospitals, elevators, bicycle rentals, rice cookers and everything in between. The cycle of euphoria was again short-lived and displaced by the fear of a looming existential threat. The concerns were so grave that technology leaders have called for a moratorium on its future development (Westfall, 2023) and held closed-door meetings to chart a path forward (Jalonick & O'Brien, 2023). The issue has received considerable attention not only from the media establishment, academic institutions, futurists, politicians, and even international bodies such as UNESCO (2023), which published recommendations on the ethics of AI.

However, much of the discourse on risks of AI ignores the societal context and fails to explain why the suggested outcome will be more likely than any other. Nick Bostrom wrote in *Superintelligence: Paths, Dangers, Strategies*, "If the AI has (perhaps for safety reasons) been confined to an isolated computer, it may use its social manipulation superpower to persuade the gatekeepers to let it gain access to an Internet port. Alternatively, the AI might use its hacking superpower to escape its confinement. Spreading over the Internet may enable the AI to expand its hardware capacity and knowledge base, further increasing its intellectual superiority" (Bostrom, 2014). But, his "orthogonality thesis suggests that we cannot blithely assume that a superintelligence will necessarily share any of the final values stereotypically associated with wisdom and intellectual development in humans" and, therefore, it might, with the same degree of probability, partake in TikTok challenges, write books, gamble online, or self-destruct. The obsession with power, though a common theme in the literature, is merely one of countless outcomes that AI could pursue. However,

considering the moral panic, it would not be surprising if soon new job titles such as AI misuse officer or AI psychologist, or even an institution like the Federal Bureau of Singularity, were created to police the AI landscape. It would not be far-fetched to assume that soon a new economy will emerge to address mental health issues related to the lost sense of reality, such as obsessive behaviors like counting fingers on photographs, persecution complexes stemming from the feeling of being watched by machines, and body dysmorphia resulting from misalignment between photo filters and actual appearance. Would these be sufficient to tear the fabric of society?

One may wonder whether AI research will eventually encounter a variation of the Fermi Paradox, where there will be no evidence of evil deeds produced by intelligent machines, on their own initiative, despite a high percentage of models that claim to be or are deemed (by some arbitrary metric) to be sentient and capable of causing harm.

The notions of singularity and superintelligence, and their underlying assumptions, themselves deserve greater scrutiny. Bostrom (2014) wrote, “If some day we build machine brains that surpass human brains in general intelligence, then this new superintelligence could become very powerful.” But then the question arises: powerful for what? Solving puzzles, answering test questions, and following instructions, or forging a path forward to a brighter future? This definition seems overly reductionist and ignores creativity, humor, emotions, and other factors that may serve as a proxy for human intelligence, which is not merely individualistic but cumulative, pluralistic, and has a temporal dimension. To maintain its status as a superintelligent being, worthy of fear or worship, AI should not only be more intelligent (however that is defined) than the average person but should also outperform the entire human race, past, present, and future. Any new idea conceived by humans that a machine has not generated and claimed would undermine its status as superintelligent.

Another definition comes from Vinge (1993), who asserted that “an ultraintelligent machine...defined as a machine that can far surpass all the intellectual activities of any man however clever...it will be the last invention that man need make.” He further defined singularity as “a point where our old models must be discarded and a new reality rules, a point that will loom vaster and vaster over human affairs until the notion becomes a commonplace.” For him, singularity is the manifestation of superhumanity. By this definition singularity is attained when a man becomes a god or, at the very least, a techno-feudal lord who enslaves higher intellect to do all his work. But could the old models be truly discarded, considering that humans are creatures of nature—they need to breath, eat, sleep, collaborate, and, according to Blaise Pascal (1910) have trouble staying quietly in their own rooms? The attainment of singularity would signify the end of humanity. When pursued deliberately, it would be equivalent to suicide; this warrants the expansion of Durkheim’s nomenclature (Durkheim, 2005/1897) by adding a new category of “transcendental suicide” to become posthuman.

Furthermore, the existential threat of AI is minimized by economic factors. A shrinking slice of the economic pie is driving many off the grid, where even basic technologies are scarce (Vannini & Taggart, 2013). According to a recent press release by The International Telecommunication Union (ITU), the United Nations agency for information and communication technologies, as of 2022, approximately 2.7 billion people, or

just over third of the global population, do not have access to the Internet. The document notes that “the chance of connecting everyone by 2030 looks increasingly slim” (ITU, 2022). There appears to be a digital divide, with developing countries, marginalized communities, and rural areas in developed countries lacking internet connectivity. Even in Europe, 11 percent of the population is still offline. Suppose that the rogue AI was successful in leading the internet-connected humanity like lemmings off the cliff; there would remain a significant group of survivors who failed to show up because they could not afford the service or it was not reliable.

When technologies are instrumental in preserving the integrity of societal processes and the functioning of institutions, the existential threat does not need to come from technologies per se, but from their misuse, abuse, and rudimentary incompetence. The Peter Principle (Peter & Hull, 1969) is a phenomenon worth paying close attention to, and arguably poses a much greater risk than rogue AI. To it we owe the ongoing disruption of ecosystems, health problems, and social injustices. The incautious use of our old technological power over the rest of nature is already catching up with us, raising alarms about the sustainability of technological progress; this trend is likely to continue due to what Pieper (2024) described as technological fetishism. While there are many ways to trigger an existential crisis, rogue AI is not the most potent way.

One may argue that the drive to temper with nature, to open Pandora’s box that can never be closed, is not motivated by pure reason, but rather by a primal instinct to revel in chaos, which brings us to another thought experiment: Suppose you are part of the national research and development program that created superintelligent AI. Nobody has a full understanding of how it works, but its predictive capacity has a potential to win every war, overtake every other economy, and colonize other planets, potentially opening the door to world domination. The question then is, what is the right thing to do? There are many possible answers to this question, each of which involves a violation of one principle or another. For example, one may choose to proceed with the program while ignoring the risks, or one may share the discovery with the whole world, like Prometheus stealing fire—a punishable act of disloyalty.

8 Conclusion

We are now at a peculiar juncture in history when one cannot help but wonder whether the societal foundation would withstand the growing weight of innovation. In the previous paragraphs, I challenged the notion of superintelligence and argued that AI is not on par with threats like nuclear war or ecological collapse; nevertheless, it will undoubtedly have a lasting impact on the course of civilization. The issue here is not with the technology per se, but rather with its integration into the existing societal structure; technology can only be understood in relation to the overarching context that brought it about and that it will serve.

The arrival of intelligent machines was a perfect storm, predicated upon three decades of digitized media and uncensored user-generated content, used for training the AI models, coupled with a market economy with risk investors who pumped money into ventures that sounded more like science fiction than actual science. The risk-averse and cash-strapped academia could not have accomplished

it on its own, albeit it provided the expertise. It would not be unreasonable to say that AI is a child of global capitalism, and the proverbial “end of history” (Fukuyama, 1989) is not economic or ideological but technological. This is important context that sets the stage for the future.

AI has emerged at a time when much of the discourse on which it was trained revolved around moral pluralism and individualism, whereby treating people as economic agents. There are no taboos that restrict use cases, and even those would be justly overridden in the name of safety, individual rights, or collective happiness; therefore, one should expect the technology to be further used to pursue economic opportunities, while squeezing out every bit of efficiency, reinforcing interdependence, and creating alienation, all while trading autonomy and privacy for convenience. Given historical precedents and the volume of investments, AI, as was the case with many earlier technologies, would be hailed as a solution to every societal ill, and as with many earlier technologies, one should expect iatrogenic effects to occur, though these would unlikely be existential.

However, technology is not universally seen as a path to salvation, despite being promoted as such. Environmentalism added its share of distrust toward the machine and created a dichotomy, dividing technologies into good and bad; trading a car for a bicycle is considered a step toward progress so are composting toilets and tiny houses. “Dumb phones” are making a comeback (Mays, 2023), dating apps are waning in popularity (Vinter, 2023), communication towers are set ablaze (Ore, 2022), and surveillance cameras are vandalized (Kelly, 2023), while citizens and governments are pursuing legal actions against the technology companies (McGreal, 2021; O’Brien, 2023; Stempel et al., 2023). The push toward greater integration of humans and machines is likely to exacerbate the contempt for the artificial; however, some non-invasive applications, such as augmented reality (Gordon, 2024), may have greater appeal than brain implants.

The questions to ask here are whether the path we are taking is going to lead us to eudaimonia or if we better change tack? What does the realization of ever-new possibilities entail? Will we stop when we overcome disease and death? Are we right to use technological advancement and economic output as a measure of human achievement and moral worth? Are we ready to let go of old truths, ideas, values, and rules in the name of progress? Should we consider the possibility of replacing democracy with algocracy, thereby striving to minimize the continuous struggle for power, and altogether relinquishing it to superintelligent machines tasked with building a just society with mathematical precision? And, most importantly, what kind of change would disqualify one from being defined as human? These are the questions that should be added to the ethical discourse.

Acknowledgements The author expresses gratitude to reviewers for their valuable feedback.

Authors’ Contributions This is a single-authored manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors, and the author has no financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The author has no competing interests to declare.

References

- Associated Press. (2022). New York changes gun buyback after seller gets \$21,000 for 3D-printed parts. *The Guardian*. <https://www.theguardian.com/us-news/2022/oct/11/new-york-gun-buyback-rules-3d-printed-parts>. Accessed 26 Oct 2023
- Barthes, R. (2001). The death of the author. *Contributions in Philosophy*, 83, 3–8.
- Borji, A. (2023). A categorical archive of chatgpt failures. *arXiv Preprint arXiv:2302.03494*.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers*. Oxford University Press.
- Bridle, J. (2023). The stupidity of AI. *The Guardian*. <https://www.theguardian.com/technology/2023/mar/16/the-stupidity-of-ai-artificial-intelligence-dall-e-chatgpt>. Accessed 26 Oct 2023
- Casciato, P. (2013). Hawking documentary details triumph of mind over ravaged body. *Reuters*. <https://www.reuters.com/article/film-hawking-idINDEE98J0GA20130920>. Accessed 26 Oct 2023
- Chang, C.-C., & Lin, C.-J. (2000). *LIBSVM 2.0: Solving different support vector formulations*.
- Chu, W. (2003). *Bayesian approach to support vector machines*. National University of Singapore.
- Copeland, B. J. (2004). *The essential turing*. Clarendon Press.
- Deleuze, G. (1992). Postscript on the societies of control. *October*, 59, 3–7.
- Descartes, R., Haldane, E. S., & Ross, G. R. T. (2003). *Discourse on method and meditations*. Courier Corporation.
- Durkheim, E. (2005). *Suicide: A study in sociology*. Routledge.
- Durkheim, E. (2014). *The division of labor in society*. Simon and Schuster.
- Ellul, J. (1964). *The technological society*. Knopf Doubleday Publishing Group.
- Ellul, J. (2011). *The subversion of Christianity*. Wipf and Stock Publishers.
- Follows, T. (2021). *The future of you: Can your identity survive 21st-century technology?* Elliott & Thompson, Limited.
- Frank, R. (2022). *Metaverse real estate sales top \$500 million, and are projected to double this year*. CNBC. <https://www.cnbc.com/2022/02/01/metaverse-real-estate-sales-top-500-million-metametric-solutions-says.html>. Accessed 26 Oct 2023
- Fukuyama, F. (1989). The end of history? *The national interest*, 16, Article 16.
- Gödel, K. (1992). *On formally undecidable propositions of Principia Mathematica and related systems*. Courier Corporation.
- Goldman, E. (1910). *Anarchism and other essays*. Mother Earth Publishing Association.
- Gordon, E. C. (2024). Human enhancement and augmented reality. *Philosophy & Technology*, 37(1), 17.
- Gorichanaz, T. (2023). Accused: How students respond to allegations of using ChatGPT on assessments. *Learning: Research and Practice*, 9(2), 183–196. <https://doi.org/10.1080/23735082.2023.2254787>
- Gould, E. (2020). *Remote learning is a bad joke*. The Atlantic. <https://www.theatlantic.com/ideas/archive/2020/08/kindergartener-virtual-education/615316/>. Accessed 26 Oct 2023
- Graeber, D. (2019). *Bullshit jobs: A theory*. Simon and Schuster.
- Harari, Y. N. (2016). *Homo deus: A brief history of tomorrow*. McClelland & Stewart.
- Heaven, W. D. (2022). Why Meta's latest large language model survived only three days online. *MIT Technology Review*, 15, 2022.
- Hume, D. (1861). *An inquiry concerning human understanding*. J.B. Bebbington.
- Huxley, J. (1957). *New bottles for new wine: Essays*. Chatto & Windus.
- ITU. (2022). *Press release: Internet surge slows, leaving 2.7 billion people offline in 2022*. ITU. <https://www.itu.int:443/en/mediacentre/Pages/PR-2022-09-16-Internet-surge-slows.aspx>. Accessed 26 Oct 2023

- Jalonick, M. C., & O'Brien, M. (2023). *Mark Zuckerberg, Elon Musk and Bill Gates huddle for hours behind closed doors with Chuck Schumer and Senate leaders over AI*. Fortune. <https://fortune.com/2023/09/13/ai-senate-briefing-musk-gates-zuckerberg-closed-door-schumer/>. Accessed 26 Oct 2023
- Jee, C. (2022). *Technology that lets us "speak" to our dead relatives has arrived. Are we ready?* MIT technology review. <https://www.technologyreview.com/2022/10/18/1061320/digital-clones-of-dead-people/>. Accessed 26 Oct 2023
- Kearney, R. (2006). *Navigations: Collected Irish essays 1976–2006*. Syracuse University Press.
- Kelly, J. (2023). Ulez camera-related crimes rose in September. *BBC News*. <https://www.bbc.com/news/uk-england-london-66987584>. Accessed 26 Oct 2023
- Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking of plagiarism detection. *arXiv Preprint arXiv:2302.04335*.
- Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Viking.
- Leuenberger, M. (2024). Track thyself? The value and ethics of self-knowledge through technology. *Philosophy & Technology*, 37(1), 13.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., & Rocktäschel, T. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
- Machiavelli, N. (1998). *The prince: Second edition*. University of Chicago Press.
- Machiavelli, N. (2009). *Discourses on livy*. University of Chicago Press.
- Marx, K., & Engels, F. (1988). *Economic and philosophic manuscripts of 1844*. Prometheus Books.
- Mays, L. (2023). *Dumb phones are on the rise in the U.S. as Gen Z looks to limit screen time*. CNBC. <https://www.cnbc.com/2023/03/29/dumb-phones-are-on-the-rise-in-the-us-as-gen-z-limits-screen-time.html>. Accessed 26 Oct 2023
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5, 115–133.
- McGreal, C. (2021). Big oil and gas kept a dirty secret for decades. Now they may pay the price. *The Guardian*. <https://www.theguardian.com/environment/2021/jun/30/climate-crimes-oil-and-gas-environment>. Accessed 26 Oct 2023
- Merken, S. (2023). New York lawyers sanctioned for using fake ChatGPT cases in legal brief. *Reuters*. <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/>. Accessed 26 Oct 2023
- Miller, H. (1995). *Tropic of cancer*. Signet/Penguin.
- Milmo, D. (2023). AI risk must be treated as seriously as climate crisis, says Google DeepMind chief. *The Guardian*. <https://www.theguardian.com/technology/2023/oct/24/ai-risk-climate-crisis-google-deepmind-chief-demis-hassabis-regulation>. Accessed 26 Oct 2023
- Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing large language models: A three-layered approach. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00289-2>
- Moore, D. (2018). You gotta fight for your right to repair: The Digital Millennium Copyright Act's effect on right-to-repair legislation. *Tex. a&m l. Rev.*, 6, 509.
- Newton, D. E. (2018). *Robots: A reference handbook*. Bloomsbury Publishing USA.
- Null, L., & Lobur, J. (2006). *The essentials of computer organization and architecture*. Jones & Bartlett Learning.
- O'Brien, M. (2023). *Sarah Silverman and novelists sue ChatGPT-maker OpenAI* | AP News. <https://apnews.com/article/sarah-silverman-suing-chatgpt-openai-ai-8927025139a8151e26053249d1aee20>. Accessed 26 Oct 2023
- Ore, A. (2022). 5G tower fire in flood-hit Mullumbimby under NSW police investigation. *The Guardian*. <https://www.theguardian.com/australia-news/2022/mar/18/dangerous-5g-tower-fire-in-flood-hit-mullumbimby-under-nsw-police-investigation>. Accessed 26 Oct 2023
- Parviainen, J., & Coeckelbergh, M. (2021). The political choreography of the Sophia robot: Beyond robot rights and citizenship to political performances for the social robotics market. *AI & SOCIETY*, 36(3), 715–724. <https://doi.org/10.1007/s00146-020-01104-w>
- Pascal, B. (1910). *Blaise Pascal* (W. F. Trotter, M. L. Booth, & O. W. Wigit, Trans.; Vol. 48). P.F. Collier & Son.
- Patient, D. (2023). *Church in AI takeover as sermon led by ChatGPT in technology breakthrough*. The Mirror US. <https://www.themirror.com/news/us-news/church-ai-takeover-sermon-led-123778>. Accessed 26 Oct 2023
- Peter, L. J., & Hull, R. (1969). *The peter principle* (Issues 04; RMD, PN6231. M2 P4.). Souvenir Press London.

- Pieper, M. (2024). Technology as a strategy of the human? A comparison between the extension concept and the fetish concept of technology. *Philosophy & Technology*, 37(1), 6. <https://doi.org/10.1007/s13347-024-00697-0>
- Plato, & Cornford, F. M. (1997). *Plato's cosmology: The timaeus of plato*. Hackett Publishing.
- Reuben Das, M. (2023). *StupidGPT: ChatGPT-like AI bots are way more stupid than people realise, says AI Expert*. Firstpost. <https://www.firstpost.com/world/chatgpt-like-ai-bots-are-way-more-stupid-than-people-realise-says-ai-expert-12632182.html>. Accessed 26 Oct 2023
- Schneider, S. (2019). *Artificial you: AI and the future of your mind*. Princeton University Press.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424.
- Stader, D. (2024). Algorithms don't have a future: On the relation of judgement and calculation. *Philosophy & Technology*, 37(1), 21. <https://doi.org/10.1007/s13347-024-00705-3>
- Stempel, J., Bartz, D., & Raymond, N. (2023). Meta's Instagram linked to depression, anxiety, insomnia in kids—US states' lawsuit. *Reuters*. <https://www.reuters.com/legal/dozens-us-states-sue-meta-platforms-harming-mental-health-young-people-2023-10-24/>. Accessed 26 Oct 2023
- Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, 50(4), 1183–1202. <https://doi.org/10.1016/j.compedu.2006.11.007>
- Takagi, Y., & Nishimoto, S. (2023). High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14453–14463.
- Tang, J., LeBel, A., Jain, S. et al. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, 26, 858–866. <https://doi.org/10.1038/s41593-023-01304-9>
- Taylor, C. (2023). *A.I. is 'quite stupid,' top exec at Facebook parent company meta insists*. Fortune. <https://fortune.com/2023/07/19/meta-nick-clegg-ai-stupid-facebook-instagram-llama-2-openai/>. Accessed 26 Oct 2023
- Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2), Article 2.
- Turing, A. (1938). On computable numbers, with an application to the Entscheidungsproblem. A correction. *Proceedings of the London*.
- UNESCO. (2023). *UNESCO's recommendation on the ethics of artificial intelligence: Key facts*. UNESCO. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>. Accessed 26 Oct 2023
- Urbina, F., Lentzos, F., Invernizzi, C., & Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. *Nature Machine Intelligence*, 4(3), Article 3. <https://doi.org/10.1038/s42256-022-00465-9>
- Vannini, P., & Taggart, J. (2013). Voluntary simplicity, involuntary complexities, and the pull of remove: The radical ruralities of off-grid lifestyles. *Environment and Planning A*, 45(2), 295–311.
- Varoufakis, Y. (2023). *Technofeudalism: What killed capitalism*. Penguin Random House.
- Vinge, V. (1993). *Technological singularity*. 30–31.
- Vinter, R. (2023). 'It's quite soul-destroying': How we fell out of love with dating apps. *The Guardian*. <https://www.theguardian.com/lifeandstyle/2023/oct/28/its-quite-soul-destroying-how-we-fell-out-of-love-with-dating-apps>. Accessed 26 Oct 2023
- Vynck, G. D., & Tiku, N. (2024). Google takes down Gemini AI image generator. Here's what you need to know. *Washington Post*. <https://www.washingtonpost.com/technology/2024/02/22/google-gemini-ai-image-generation-pause/>. Accessed 27 Feb 2024
- Westfall, C. (2023). *Elon musk and other tech leaders call for slowdown on AI development*. Forbes. <https://www.forbes.com/sites/chriswestfall/2023/03/29/elon-musk-and-other-tech-leaders-call-for-slowdown-on-ai-development/>. Accessed 26 Oct 2023
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.