Philosophy & Technology (2026) 39:65
https://doi.org/10.1007/s13347-026-01065-w

RESEARCH ARTICLE

The Eclipse of Reason in the Age of Artificial Intelligence: Why we are Failing to Cope with AI Development and Steer it Toward Sustainability

Federico Cugurullo[1,2]

Received: 30 July 2025 / Accepted: 21 February 2026
© The Author(s) 2026

Abstract
Artificial Intelligence (AI) is causing major disruptions, triggering waves of social, cultural, political, economic and environmental problems that risk undermining the very foundations of human society and the totality of the environment surrounding it. Two problems in particular stand out. The first one is that there is generally no vision of development behind the production of AI technologies, but rather individualistic political and economic strategies benefitting a few oligarchs and private companies. The second problem is that there is often little or no comprehension of the AI technologies that are being produced, resulting in several forms of distress and confusion cutting across culture, economics and politics. The argument advanced in this paper is that these two problems derive from one broader problem: the *eclipse of reason*. In order to unpack and explain this problem, the paper draws on the philosophy of Max Horkheimer and makes three distinct contributions. First, it develops a theoretical framework to illustrate the nature of the eclipse of reason and how it unfolds nowadays in the age of AI. Second, it expands Horkheimer's philosophy, by taking into account and integrating matters of education, ideology and governance, which are crucial for understanding not simply how the eclipse of reason takes place, but also how it *stays* in place, thereby chronically clouding our capacity to comprehend AI and govern its deployment in a sustainable manner. Third and finally, the paper proposes a philosophically grounded roadmap for navigating the changes that are needed to reverse the eclipse of reason.

Keywords Artificial intelligence · Reason · Critical theory · Sustainability · Governance · Acceleration

✉ Federico Cugurullo
federico@nus.edu.sg

[1] Department of Geography, National University of Singapore, AS2-03-01, 1 Arts Link, 117568 Singapore, Singapore

[2] Department of Geography, Trinity College Dublin, Dublin, Ireland

Published online: 30 March 2026

1 Introduction

It is a matter of fact that Artificial Intelligence (AI) is causing major disruptions, triggering waves of social, cultural, political, economic and environmental repercussions that risk undermining the very foundations of human society and the totality of the environment surrounding it (Coeckelbergh, 2024; Cugurullo et al., 2025; Johnson & Acemoglu, 2023). The spectrum of these repercussions is so broad and heterogeneous, ranging from the depletion of natural resources to the erosion of democracies, and the scale of their impact so vast, from the behavior of individuals to the metabolism of entire nations, that the root cause that they share is not immediately visible. As a result, many AI-related problems appear to be isolated and distant from each other, in a way that makes it increasingly difficult to connect the dots and find solutions targeting their source. This paper addresses this issue by advancing a theoretical framework grounded in critical theory and designed to expose what is at the core of some of the most pressing problems caused by AI development.

More specifically, the paper focuses on two interconnected problems. The first one is that there is generally no vision of development behind the production of AI technologies. Governments from all over the world are investing a lot in AI, but many of these investments do not seem to be actually directed towards broader development goals rooted in the needs of society as a whole and in the preservation of the natural environment. Instead, it is evident that what is often at play are the interests of a few oligarchs and private companies, as the Trump administration exemplifies emblematically. In January 2025, Trump announced a $500 billion investment in AI infrastructure as part of a strategic plan of technological development featuring prominently Sam Altman, CEO of OpenAI and Larry Ellison, chairman of Oracle, at the beginning of a presidency heavily influenced by Elon Musk (White House, 2025). In Trump's words, this plan and its outcome will be "incredible" (CNBC, 2025). He talks about the "construction of colossal data centres" and "massive structures" that he describes as "beautiful buildings", making direct reference to real estate (CNBC, 2025). Furthermore, he says that "it will lead to something that could be the biggest of all" (CNBC, 2025). However, what that "something" is or means is never specified. No clear vision of development is presented and, apart from a small mention of the power of AI to cure diseases hitherto considered untreatable and of the obvious economic and geo-political benefits of onshoring the AI industry, Trump's speech fails to take into account the needs of citizens and of the ecosystems that they inhabit (CNBC, 2025).

The second interconnected problem is that there is generally little or no comprehension of the AI technologies that are being produced, resulting in several forms of distress cutting across culture, economics and politics. As investments in AI innovation keep flowing, and as many different AI technologies enter our society and mediate most aspects of our life, a lot of people are struggling to cope with the rapid and radical changes that AI is triggering. Examples of this problem include the numerous individuals who are starting romantic relationships with AIs *à la* ChatGPT as part of an emergent neo-animism which denotes a lot of confusion regarding the extent to which AI technologies are conscious and should be regarded as sentient beings (Cugurullo, 2024; Pan & Mou, 2024). From an economic point of view, there is also

confusion about the actual capabilities of AI (particularly complex Generative AIs), and this is leading to uncertainty as to what jobs should be automated or better left in the hands of humans (Agrawal et al., 2023). Most problematically, from a political perspective, society is struggling to respond to the escalation of AI development in a manner that can actually steer it towards the benefits of the greatest number of people, without harming the natural environment in the process. There are of course important exceptions to take into account and these include recent AI regulations as well as the work of NGOs pushing for a progressive, democratic and sustainable use of AI (Cugurullo et al., 2025; Floridi, 2021; Roberts et al., 2023). Yet, there is a growing feeling that, by and large, the construction and deployment of AI is serving the purpose of accruing the power of already powerful economic and political actors and, overall, empirical evidence of the social and environmental sustainability of AI remains scarce (Coeckelbergh, 2024; Johnson & Acemoglu, 2023; Zuboff, 2019).

The argument advanced in this paper is that these two problems derive from one single broader problem called the *eclipse of reason*. In order to unpack and explain this problem, the paper draws on the philosophy of Max Horkheimer and makes three distinct contributions. First, it develops a theoretical framework to illustrate the nature of the eclipse of reason (Sect. 2) and how it unfolds nowadays in the age of AI (Sect. 3). Second, it expands and updates Horkheimer's philosophy, by taking into account and integrating matters of education, ideology and governance, which are crucial for understanding not simply how the eclipse of reason takes place, but also how it *stays* in place, thereby chronically clouding our capacity to comprehend and govern AI (Sect. 4). Third and finally, the paper proposes a philosophically grounded roadmap for navigating the changes that are needed to reverse the eclipse of reason (Sect. 5).

2 Theoretical Framework: The Eclipse of Reason in Horkheimer's Philosophy

In *Eclipse of Reason*, originally published in 1947, Max Horkheimer (1895–1973) one of the founders of the so-called Frankfurt School, investigates the nature of reason in modern times, in an attempt to shed light on the forces the drive the actions of single individuals as well as society as a whole, by paying particular attention to matters of technological development. In unpacking the concept of reason, Horkheimer (2013) makes and proposes a crucial distinction between *subjective* and *objective* reason. For Horkheimer (2013), subjective reason is the reason of the individual, which focuses on the self and what the self needs and desires. This is thus a form of reason leaning toward personal interests and gains and, eventually, triggering actions whose outcomes benefit a given individual. Horkheimer (2013) argues that there is nothing wrong with subjective reason per se, inasmuch as it is a part of human nature that originates in our primordial instinct of self-preservation without which, put it simply, we would not care about our own basic needs, such as food or shelter, and therefore we would not survive. However, Horkheimer (2013) invites us to reflect on the dangers that an excessive dose of subjective reason might cause, since an extreme attention to the needs, wishes and benefits of the self risks translating into a disregard

for what is around the self. In other words, if we take subjective reason to extremes, it can make us blind to the needs of other individuals and those of the surrounding environment.

Objective reason is different. Horkheimer (2013: 2) describes it as a "force" driving "relations among human beings and social classes, social institutions and nature." This form of reason thus goes beyond the sphere of the self and extents its scope to groups of individuals, to society as a community in which interdependent individuals live, to the institutions that facilitate and regulate their lives and, last but not least, to the surrounding natural environment. Seen from Horkheimer's perspective, reason becomes therefore a matter of scale, with subjective reason pertaining to the small and narrow scale of the self, and objective reason encompassing much larger and broader scales, including cities, regions, nations and the whole planet in an effort to care for humanity and nature. This is because, for Horkheimer (2013), objective reason focuses on a plurality of needs and interests and, unlike subjective reason, it is not restricted to self-preservation. Instead, it is sensitive to the preservation of society as a whole and it seeks to find an equilibrium between the interests of many different selves, without forgetting that our species is part of complex and delicate ecosystems that have needs too. In these terms, although Horkheimer does not explicitly engage with the notion of posthumanism, his idea of objective reason resonates with contemporary posthuman philosophies and their attention to non-anthropocentric models of development and more-than-human needs (Braidotti & Bignall, 2018; Clark, 2023).

In his explanation of objective reason, Horkheimer (2013) stresses that this is a form of reason which is based on *higher concepts*. A higher concept is one that captures concerns, needs, aspirations and scales that go beyond the sphere of the self. Here Horkheimer (2013) draws on Plato's (2007) philosophy and proposes *justice* and *equality* as examples of higher concepts. Similarly, he also mentions *happiness* as a higher concept and, in so doing, establishes a connection with Aristotle's (2000, 2004) theory of *eudaimonia*. For Aristotle (2000) achieving happiness or, in his words, a condition of eudaimonia is a matter of development, relations and scales. It is a matter of development because it begins with single individuals coming together as a family; then several families forming a village and eventually the union of a few villages leading to the formation of the *polis* (πόλις), which is for Aristotle (2000) the ideal type of human community. It is also a matter of scale and relations because, as Aristotle (2000) argues, single individuals cannot fully realize their potential, flourish and, thus, become happy, by themselves. They need other people as well as the resources that the polis provides (Aristotle, 2000). From an Aristotelian perspective, happiness is relational and the governance of the polis is precisely an act meant to maximize not simply the eudaimonia of individuals, but of them as citizens: as members of the polis and part of a collective (Aristotle, 2000).

Returning to Horkheimer, the process that he describes as the eclipse of reason is twofold. First is the tendency of subjective reason to overpower objective reason. This is a tendency that, according to Horkheimer (2013), is strong in modern and technologically advanced societies where subjective reason triumphs to the detriment and, ultimately, negation of objective reason. Paraphrasing Horkheimer, this means that one side of the coin is missing. For Horkheimer, subjective and objective reasons are two sides of the same coin, i.e. reason, and they are both important. What

he laments is that the negation of objective reason removes a crucial part of reason, thereby maiming reason itself. In practice, this situation creates a problematic lacuna in the compass the orientates our actions as well as our models of development which, as Horkheimer (2013) argues, are negatively influenced by the supremacy of subjective reason and its emphasis on individual needs over those of society and the environment.

Second is "the complete transformation of the world into a world of means, rather than of ends" (Horkheimer, 2013: 66). This is a key passage that requires some explanation. For Horkheimer, subjective reason tends to pursue the development of means or, put it simply, instruments for fulfilling personal objectives (which are the main concern of this strand of reason). In the philosophy of Horkheimer (2013), this tendency is connected to technological development and innovation whereby more and better instruments are created, eventually leading to the production of advanced technology. For example, the need for shelter has led over the course of millennia to the construction of rudimentary dwellings made of wood around two million years ago; and subsequently to the erection of brick houses around 7,000 BCE, along an impressive process of techno-domestic development which is now culminating in domotics (Clark, 2003; Cugurullo et al., 2023). A similar need has also led, since the invention of money and the diffusion of market economies, to the pursuit and accumulation of different forms of capital. This is because money becomes *de facto* an instrument in our hands that we can use to buy a home when we do not want or cannot build one ourselves.

Objective reason is different and, consequently, it leads to another type of development. Being objective reason based on higher concepts, this strand of reason is for Horkheimer (2013) intrinsically connected to philosophical development. This is because higher concepts, such as *justice*, *equality* and *happiness*, are difficult to define and understand, and philosophy is thus key to reflect upon their meaning. The complexity of a higher concept becomes apparent in comparison with a basic need targeted by subjective reason. The need for food, for instance, that each of us feels can be simply explained as the need that our body has for energy in order to move, grow and repair. In contrast, a concept like justice requires a much more sophisticated and longer explanation: one that started with philosophers including Plato (2007) in ancient Greece and Confucius in ancient China and that, after many centuries, is still in the making (Duvert, 2018; Miller, 2013). The notion of *sustainability* as an example of a recent higher concept is another case in point. Sustainability is a complex concept, inasmuch as it comprises a multi-species perspective on development, multiple scales ranging from local ecosystems to global environmental changes, and an intricate connection of matters of economic growth, ecology, justice, culture and politics (James, 2014). Capturing this complexity and elucidating what a sustainable form of development means cannot be done without rigorous philosophical reflections. This is further exemplified by recent works in philosophy of technology in which philosophers interrogate the meaning of sustainability in relation to AI (Mazzi & Floridi, 2023).

Furthermore, it is crucial to note that this line of philosophical work is not just meant to provide explanations or capture meanings. Here philosophy also identifies goals, or *ends* in the words of Horkheimer (2013). Justice, for example, is not simply

a concept: it is a societal goal that a plethora of philosophers have identified, for a long time, in a way that should orientate the development of society (Risse, 2020). Similarly, sustainability is supposed to be a goal too. In this regard, philosophical reflections on sustainability and AI are meant not only to develop an understanding of what a sustainable AI is but, above all, to clarify that nowadays an important goal for governments from all over the world is to deploy AI specifically to achieve a condition of sustainability (Cowls et al., 2023).

Means and ends are of course compatible. Building houses to fulfil the need for shelter would not be possible without technological tools. Similarly, the goals of sustainable development can be supported by technology in the shape of AIs trained to monitor environmental loss or to predict future patterns of energy consumption and carbon emissions, for example (Cowls et al., 2023; Silvestro et al., 2022). What Horkheimer (2013) denounces in *Eclipse of Reason* is that when one side of reason is missing, we end up with a constant process of technological development leading to more and increasingly powerful technologies. Here, put it simply, *means* abound. However, without the guidance of objective reason, this type of development is not pursuing an end, intended as a goal inspired by higher concepts whose realization would benefit society as a whole without harming the environment. Therefore, in Horkheimer's (2013: 66) words, we eventually end up living in a "world of means, rather than of ends." This is a dangerous condition because a constant technological development, lacking a clear social and environmental vision capable of guiding its direction and giving meaning to it, is a blind form of development.

2.1 Discussion: AI Development from the Perspective of Horkheimer's Philosophy

The theoretical framework illustrated above is helpful to discuss and understand some of the major problems that the development of AI is posing nowadays. The eclipse of reason theorized by Horkheimer is not a sudden phenomenon. It is a process that unfolds gradually and, from this perspective, we can notice how the emergence of different AI-related problems corresponds to different phases of the eclipse of reason. This paper deals with two interconnected problems, in particular, which were introduced in Section One. These are first the lack of a broader vision of development behind stratospheric investments in AI technology; one that is rooted in the needs of society as a whole and the environment, instead of benefitting mostly a minority of oligarchs and corporations at the expense of the planet's ecosystems. The second one pertains to a general lack of comprehension of the AI technologies that the aforementioned investments are producing. As it was stressed in the introduction, this is evidenced on a personal level by romantic relationships that a growing number of individuals are starting with AIs under the false belief that such synthetic intelligences are conscious; by a widespread confusion about the actual capabilities of AI and, last but not least, by a global impasse regarding what political strategies can, in theory and practice, actually steer AI development towards a condition of sustainability.

The first problem can be explained in relation to the first phase of the eclipse of reason described by Horkheimer. The development of AI technology and the massive investments that today are being made in the field of AI are often driven by *subjective*

reason. This is a form of reason that, as the German philosopher reminds us, leans toward personal interests and gains and, as a result, any outcome rooted in subjective reason benefits only a small number of individuals. What is lacking is the influence of *objective* reason as a force driving "relations" among people and between them and the natural environment (Horkheimer, 2013: 2). Without objective reason, development cannot go much further than the sphere of the self and is ultimately unable to reach an equilibrium between the interests of multiple selves and those of the more-than-human world.

Most importantly, without objective reason, development is neither inspired nor framed by higher concepts such as justice, equality, happiness and sustainability. These are concepts that should orientate the genesis and deployment of AI technology, and their absence simply means that AI development would not be able to equally spread its benefits, and to boost people's eudaimonia without harming the natural environment. Overall, anytime we hear political discourses, such as Trump's AI strategy (White House, 2025), that lack a broad vision of development beyond the obvious economic gains of few stakeholders, and when we come across evidence of the socio-environmental unsustainability of AI that is now abounding in the shape of surveillance capitalism, digital authoritarianism, ghost work, environmental exploitation, e-waste and mounting carbon emissions generated by energy and water-intensive data centres (Brevini, 2021; Coeckelbergh, 2024; Cugurullo et al., 2025; Falk et al., 2024; Zuboff, 2019), the explanation from the perspective of Horkheimer's philosophy is simple: subjective reason has overpowered objective reason.

Relatedly, the root of the second interconnected problem lies in the second and final phase of the eclipse of reason identified by Horkheimer (2013: 66), namely the "transformation of the world into a world of means, rather than of ends." Nowadays, we can observe the development of AI technologies as means to pursue individualistic goals. Once again, real-life examples abound. Tech companies using AI tools to extract behavioral data that are later bought by advertisers (Zuboff, 2019). Predictive AI systems deployed at the local level as instruments of disciplinary power to monitor citizens' activities, so to anticipate and control their next moves (Xu et al., 2024). Elections manipulated via AI (Coeckelbergh, 2024). Urban AI experiments such as Neom whereby the Saudi elite is seeking to attract international investments and, at the same time, eliminate any form of dissent and maintain the status quo (Cugurullo, 2026). It is here that AI becomes an instrument of surveillance, a tool to influence public opinion and gain political power, or simply a commodity that is commercialized and sold to make money. From the perspective of Horkheimer's philosophy what is missing is philosophical development and inquiry to critically reflect on the meaning of AI, its actual capabilities, the extent to which it can foster a condition of sustainability and, if so, how. In practice, what we are witnessing is a form of technological development that is by and large decoupled from philosophical development.

What is also missing are clear goals or, in Horkheimer's (2013: 66) words, "ends". AI development is generating plenty of instruments, but it is not equally generating societal goals. According to the German philosopher this is a task that technological innovation alone cannot perform and one which requires philosophical inquiry. Overall, Horkheimer's thesis is helpful to recognize that the two problems discussed in this paper are interconnected, because they share the same root cause and repre-

sent different stages of the formation of one broader problem. Horkheimer calls this problem the "eclipse of reason" which, as the previous section has explained, consists of a twofold process whereby development becomes dominated by the interests of a minority of individuals (and their subjective reason), eventually leading to the production of technological instruments whose deployment is not targeting overarching societal goals (or *ends* in his words) identified through a meticulous philosophical inquiry. However, Horkheimer's thesis needs to be expanded and updated, given that *Eclipse of Reason* is a relatively short contribution written at the end of the Second World War, and today we are facing a different political and technological landscape. It is with the aim of updating and expanding Horkheimer's philosophy that the next section begins.

3 Segue: An Expansion of Horkheimer's Philosophy in the Age of AI

In his analysis of the technology-development nexus, Horkheimer tends to present a series of binary scenarios. Most prominently, he counterposes subjective and objective reason in a dichotomic manner, to then present technological development as a process following a track that is different than that of philosophical development. As it was stressed in the previous section, Horkheimer's approach brings noteworthy advantages because of its capacity to grasp the big picture and identify the crux of the matter in relation to major social and technological problems. However, the same approach is unable to capture some important nuances that risk to remain hidden by the binaries established in *Eclipse of Reason*.

For a start, the "transformation of the world into a world of means, rather than of ends" denounced by Horkheimer (2013: 66), does not necessarily lead to a situation in which philosophical development and inquiry are completely absent and irrelevant. The state of affairs is more complex than Horkheimer's account, and this is evident nowadays as a number of philosophical studies are, to some extent, influencing AI development and its impact on society. Many of these studies are precisely in the tradition of Horkheimer's critical theory and serve the important purpose of demystifying AI and clarifying its actual capabilities. The work of Natale (2021), for example, convincingly demonstrates that numerous AIs might appear to be conscious because they are strategically designed in a deceitful manner to give the illusion of possessing humanlike properties. In a similar critical vein, Pasquinelli's (2023) studies help us reconsider the autonomy of AI technologies on the basis of their dependency on human labour and knowledge, while recent philosophical inquiries within the Marxist tradition have shown that AI is an ideology hiding a project of social and environmental domination (Cugurullo, 2025). Beyond the realm of critical theory, we also find a number of philosophers, particularly in the field of AI ethics, who are actively working on real-life regulations and policies together with governments and supranational organizations (Cocchiaro et al., 2025; Floridi et al., 2018, 2022).

Therefore, the problem is not that philosophical inquiry and critical theory are completely absent or decoupled from technological development, as Horkheimer's thesis would suggest. These intellectual perspectives are indeed present, but they are not evolving and being disseminated as rapidly as technological innovation. In con-

ceptual terms, we can grasp this issue by drawing on Rosa (2013) to understand that there are different kinds of acceleration. One is *technological acceleration* which, as Rosa (2013) explains, is the speed at which technologies are invented, produced and distributed. Another type of acceleration is the velocity at which technology is being studied and comprehended by the society that is deploying it, here defined as *epistemological acceleration*. In the age of AI, technological acceleration has substantially surpassed epistemological acceleration. In crude politico-economic terms, instead, we can simply and safely state that more investments are pouring into the making of AI technologies, rather than in their philosophical study.

Relatedly, there is the question of culture and education. This is something that Horkheimer does not discuss in *Eclipse of Reason*, although it is of fundamental importance. Philosophy is indeed crucial and the critical theory advocated by Horkheimer is necessary to evaluate the capabilities of a given technology and, above all, to reflect on the extent to which they can serve the purpose of realizing higher concepts such as justice and happiness. However, without effective educational systems and programmes there is a very concrete risk that the knowledge produced by critical theory remains isolated and unevenly distributed, thereby losing its potential for actual impact. Returning to some of the issues discussed so far in this paper, there is a clear tension between the fact that a growing number of people consider some AI technologies to be conscious to the point of even starting romantic relationships with them (Pan & Mou, 2024), and the fact that critical studies show that the same technologies are not actually conscious, but appear as such since they are deceitfully designed to fake humanlike properties like consciousness (Natale, 2021). It is also problematic that most governments are investing heavily in AI under the belief that it will fix all environmental problems as a silver bullet, while critical scholars have denounced such belief as a dangerous ideological illusion hiding colossal amounts of energy, water, carbon emissions, e-waste and raw materials, intrinsic to the ferocious metabolism of the AI industry (Cugurullo, 2025). These tensions exist because the knowledge coming from critical theory has not been sufficiently disseminated via education on a large scale. In a nutshell, what we observe is that critical theory has generated precious knowledge about technology, in a manner that would be in line with Horkheimer's aspirations, but the same knowledge has not become part of (mainstream) culture.

It is not a coincidence that so-called *AI literacy*, namely the capacity to properly understand, evaluate and use AI technology, has become one of the key factors shaping how people approach AI and the extent to which they decide to integrate it into their life (Almatrafi et al., 2024). It should not come as a surprise that, in this context, those who are more literate show greater care in the deployment of AI and awareness concerning concrete risks, including data breaches, glitches, system failures and algorithmic bias, instead of fearing unrealistic sci-fi scenarios populated by sentient technologies (Cugurullo & Acheampong, 2024; Tully et al., 2025). Ultimately, literacy is only tangentially connected to philosophy and remains chiefly a matter of education, which is a situation that invites us to expand Horkheimer's thesis and its binary representation of development composed of technological innovation and philosophical inquiry.

Another binary scenario postulated in *Eclipse of Reason* pertains to higher concepts which in Horkheimer's (2013) thesis are either present or absent depending on the presence or absence of objective reason in processes of technological development. The way Horkheimer illustrates this scenario requires more nuance and an extra layer of philosophical reflections, since presence and absence can actually overlap in complex ways. Higher concepts can indeed be present, but only nominally at a very superficial level. If we take a higher concept like *sustainability*, as an example, it becomes quickly evident that the notion of making development sustainable is present, almost universally, in countless policies, political agendas, projects and masterplans (Wheeler, 2023). However, there is also substantial evidence, particularly at the local level, that many of the initiatives that are officially pursued under the banner of sustainability, are not actually targeting and achieving environmental preservation and social justice (Kuntsman & Xin, 2024). In these cases, the term *sustainability* is present merely as an empty signifier, in a way that is detrimental to *sustainability* as a higher concept, with the latter eventually losing meaning and practical relevance (Brown, 2016; Swyngedouw & Kaika, 2014).

In postmodern philosophy, the situation described above has been recurrently critiqued and explained by several scholars; most prominently by Foucault (2002) who talks about discourses to stress how language can be a treacherous medium to hide strategies targeting one thing only: power. On a related note, an analogous discrepancy between statements and facts has also been highlighted in Kirkham's (1992) *correspondence theory* according to which a given claim does not inevitably correspond to what is actually happening on the ground. Overall, returning to the paper's main topic, artificial intelligence, a similarly ambiguous situation can be observed in relation to sustainability as a higher concept. On the one hand, the notion of sustainability is present in AI development, in the shape of many local, regional and international initiatives in favour of *sustainable AI* (Van Wynsberghe, 2021). These are initiatives that are explicitly aligned with the ethos of sustainability under the belief that AI technology can and will make the spaces that we inhabit, cities in particular, more environmentally sustainable (Cugurullo, 2025).

On the other hand, however, a growing number of critical studies are showing how behind the same initiatives are often plans to use AI to preserve the power of the dominant elite, without any real care for the natural environment (Cugurullo, 2025; Hähnel & Müller, 2025). From the perspective of Marxian philosophy, this is ideology in action and, more specifically, *AIdeology* whereby visions of sustainable cities powered by AI hide projects of socio-environmental domination (Cugurullo, 2025). Emblematic is the case of Neom in the Kingdom of Saudi Arabia where new cities filled with AIs ranging from robots to driverless vehicles and autonomous urban operating systems, are being built under the banner of sustainability, in sharp contrast with a reality characterized by social injustice and loss of natural habitat (Cugurullo, 2026). In addition, it is important to remember that even when there is a genuine interest in using AI for sustainability purposes, the environmental costs of producing and maintaining AI technologies in terms of energy, water, carbon emissions and rare earth minerals and metals, remain substantial (Brevini, 2021; Cugurullo et al., 2025; Falk et al., 2024).

Such a problematic and ambiguous situation in which the sustainability of AI can be nominally present while being in practice absent, goes to show that the critical theory advocated by Horkheimer is not enough for the purpose of aligning technological development with higher concepts. Critical theory has to go hand in hand with critical empirical studies to shed light on the extent to which projects and initiatives that are officially leveraging AI to find sustainable solutions, are *de facto* advancing a condition of sustainability. This synergistic approach offers the potential of generating a virtuous circle connecting theory and practice, with philosophers and scholars working in cognate disciplines learning more about how supposedly sustainable AIs perform on the ground, and revising their theories and policy recommendations accordingly.

Finally, it is important to recognize that comprehending a higher concept does not automatically lead to its realization. In his philosophical system, Horkheimer (2013) essentially boils development down to two tracks. One is technological development through which technologies are designed and created. The other one is philosophical development whereby higher concepts are formulated and understood. In the mind of the German philosopher, these two lines of development should intersect and go hand in hand, with the latter framing the former so that philosophical inquiry can inspire the creation of technology and orientate its deployment towards higher concepts such as happiness and justice. A lacuna in this thesis is that while higher concepts, as Horkheimer (2013) points out, are indeed very complex to understand and thus require philosophical research to flesh out their meaning, they are also extremely difficult to realize.

We can refer again to the higher concept of sustainability, as an example, to illustrate this tension. From a conceptual perspective, there is the enormous challenge of coming up with a sound understanding of sustainability, capable of integrating the needs of heterogeneous individuals, societies and ecosystems, and encompassing social, political, environmental and economic spheres; one that has global resonance, while also focusing on specific socio-environmental issues, such as the loss of natural habitat and climate change, so to avoid becoming an empty signifier (Brown, 2016). Assuming that, over the years, a collective philosophical effort manages to overcome such conceptual challenge, there would still be an even bigger challenge around the corner, in relation to how a sustainable form of development can be practiced globally. In other words, in this scenario, the goal of development (or the *end* in Horkheimer's words) would be finally clear, but there would not be any clarity regarding the kind of strategy that is necessary to achieve it. Considering the plethora of human and more-than-human needs involved in the realization of sustainability at an international scale, this is an issue that only a very advanced system of global governance can tackle.

In the current age of AI, this is now more evident than ever as supranational organizations struggle to regulate the use of AI technologies in a binding manner, and key players on the global stage, such as cities for instance, attempt to find ways to cooperate and use AI against international environmental problems (Feldstein, 2024; Hintz et al., 2026; Machen & Pearce, 2025). This evidence underscores the value of expanding Horkheimer's thesis beyond the pillars of development identified in *Eclipse of Reason*, integrating matters of governance, as well as the questions of education and culture previously discussed in this section, into an updated critical theory.

4 Conclusions: Towards a New Critical Theory

The diffusion of AI is sending shock waves across multiple scales, from the everyday life of people within their personal sphere of development, to the highest levels of national and international development. As a result, a plethora of heterogeneous social, psychological, political, economic and environmental problems have emerged, and this paper has found a shared root in what Horkheimer (2013) calls the *eclipse of reason*. The paper has focused on two interconnected problems in particular. The first one is a rapid accumulation of stratospheric investments in AI technology in a manner that is not being directed towards a clear and holistic vision of development, with the result that those who are benefitting the most from it are a minority of oligarchs and private companies thirsty for individualistic gains. The second issue concerns a widespread struggle to cope with AI innovation as both individuals and large-scale political institutions appear to be confused about how to integrate this technology into processes of personal and international development in a sustainable way.

Their interconnection lies in the unfolding of the eclipse of reason theorized by Horkheimer, with each problem representing a phase of the eclipse. The lack of a holistic vision of development targeting society as a whole and the more-than-human world, is due to the dominance of the subjective reason of a few powerful individuals that steer AI innovation towards their own gains. The confusion that surrounds the integration of AI technologies into people's life as well as into the life of whole countries is instead connected to what Horkheimer (2013: 66) describes as the "transformation of the world into a world of means, rather than of ends", which ultimately results in a reckless technological development that is neither framed nor guided by critical philosophical inquiry.

The paper has built on Horkheimer's thesis and expanded the theory of the eclipse of reason, by taking into account aspects of technological and societal development that were not originally considered by the German philosopher. This effort culminates now in a new critical theory seeking to illustrate the dynamics whereby, in the age of AI, the eclipse of reason is unfolding and with what implications (see Fig. 1). These days reason begins its eclipse and it enters a profound crisis with the explosion of massive investments in the creation of new AIs, and this can be observed, as noted in the introduction, in the $500 billion investment in AI infrastructure announced by the Trump administration, for instance, among many other contemporary national AI strategies (CNBC, 2025; Cugurullo, 2024). *At the same time*, various processes of philosophical development take place in order to critically understand the meaning, capacities and risks of AI, and to theorize how this technology can be deployed to realize higher concepts such as justice, equality and sustainability.

Here examples abound and it is worth mentioning the study of the deceitful design of AI technologies meant to fake humanlike properties like consciousness, as part of ideological schemes (Cugurullo, 2025; Natale, 2021); the critique of AI as an extractive technology that absorbs the knowledge of workers as well as people's behavioral data, often without their consent (Pasquinelli, 2023; Zuboff, 2019); the many efforts to develop ethical guidelines to steer AI innovation towards a condition of sustainability (Bolte & Van Wynsberghe, 2025; Floridi et al., 2018; Floridi, 2024; Hagendorff, 2022); and the definition of concepts such as *democratic AI*, *sustain-*

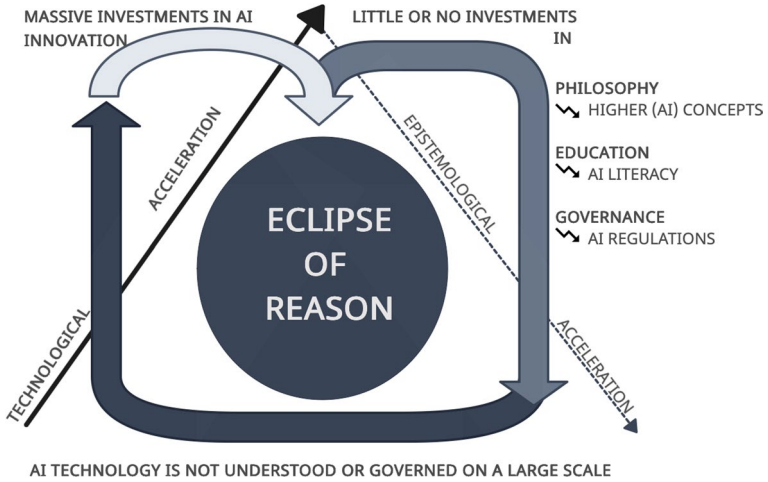


Fig. 1 The eclipse of reason in the age of AI. Source: author's original

able AI, *sustainable AI urbanism* and *AI for people*, which resonates with Horkheimer's (2013) intuition about the function of so-called higher concepts in framing the design, production and deployment of technology, in the hope that such concepts will eventually feed into policy making and industrial development (Coeckelbergh, 2024; Floridi et al., 2018; Palmini & Cugurullo, 2024; Van Wynsberghe, 2021).

In this context, it is important to pause and reflect on the fact that *at the same time* does not mean *at the same speed*. Drawing on Rosa's (2013) theory of acceleration, this paper has stressed that there are two main velocities at play in the eclipse of reason. One is what Rosa (2013) terms *technological acceleration* along an axis where innovation in new technologies skyrockets. The other one theorized in this contribution refers to the speed at which new technologies are comprehended within the societies that end up adopting them. We call this velocity *epistemological acceleration* which, in the age of AI, is being drastically surpassed by technological acceleration.

This temporal disequilibrium is by and large an outcome of a broader imbalance in terms of investments, with tech giants and governments alike investing more resources in technological rather than philosophical development. A similar imbalance is also present when it comes to matters of education. Levels of AI literacy are still relatively low worldwide, and this can be explained by the fact that the technologies produced by AI companies circulate globally through processes of mass distribution that openly target mass consumption, often with government support and incentives (Almatrafi et al., 2024; Schmid et al., 2025). In comparison, critical theories and theoretical frameworks concerning AI suffer from a more modest distribution, and are penalized by educational systems that receive much less funding than industrial systems. The problem, therefore, is not that this body of knowledge is absent, as Horkheimer's (2013) original thesis would suggest. Philosophical inquiry and critical theory are being cultivated and are indeed informing the current debate about AI, but at a scale and speed that is almost insignificant compared to the velocity and commercialization of AI innovation.

Relatedly, there is the matter of the extent to which the knowledge produced by philosophical inquiry, in the shape for example of critical theories and ethical frameworks pertaining to AI, generates impact. As it was previously mentioned, this is essentially a question of governance which, although missing in Horkheimer's (2013) thesis, is very relevant to his philosophy. This is because higher concepts such as justice and happiness and, more recently, sustainability and sustainable AI, require a coordinated effort in order to be realized: one that transcends the sphere of the self and comprises the needs of society as a whole and the environment. In the age of AI, such effort is equal to the challenge of the *governance of AI* understood as a set of processes, norms and regulations whereby different actors cooperate and perform coordinated actions in order to steer the development and deployment of AI technologies towards benefits that are equally shared by humans, without harming the more-than-human world (Braidotti & Bignall, 2018; Bullock et al., 2024; Floridi et al., 2018; Roberts et al., 2024). In line with the principles of objective reason, the scale of this strand of governance should encompass the entire planet in an effort to care for humanity and nature, and it is not a coincidence that many philosophers are advocating for a global AI governance (Roberts et al., 2024).

Alas, in the age of AI, the eclipse of reason reaches its apex because governments worldwide are not investing enough resources in efficient systems of global governance and in implementing binding sets of rules to control and limit the use of AI. As a result, what we observe today is the gradual diffusion of relatively weak regulations on the part of supranational entities, such as the EU, while a few NGOs struggle to push for the development of more ethically sound and sustainable AI systems, albeit without enough political power to drive significant change (Cugurullo et al., 2025; Feldstein, 2024). Examining the reasons why this is happening would go beyond the scope of the paper and extend the analysis to the terrain of political economy which is another piece of the puzzle that Horkheimer (2013) does not search for in *Eclipse of Reason*. In a nutshell, it is important to remember that several states are assuming the role of investors *in* rather than regulators *of* AI, with many studies showing how at both the local and the national level, governments are teaming up with private tech companies to quickly finance and co-develop AI, in a growth-oriented manner that disincentivizes its prompt regulation (Schmid et al., 2025; Xu et al., 2024).

Coming back full circle (see Fig. 1) and returning one more time to Horkheimer's thesis, it is because of a sheer disequilibrium in terms of investments and velocities that we end up in a situation in which technological instruments abound, while there are not enough cultural instruments to make sure that society as a whole comprehends them, and enough political instruments to regulate their development and deployment on a global scale. In such situation, the cycle of the eclipse of reason continues and its negative effects are reproduced *ad infinitum*, inasmuch as in the absence of a critical awareness of AI across society (evidenced by low levels of AI literacy) people tend to accept this technology, often under the illusion that it has properties that in reality it has not got. This is exemplified by the growing number of individuals who are starting romantic relationships with AIs, under the false belief that these technologies possess consciousness (Pan & Mou, 2024). Ultimately, the acceptance of a given technology (in this case AI), increases the demand for it which, in turn, fuels investments in the production of more and more AIs. In addition, the absence

of efficient governance systems supposed to regulate and limit the development of AI simply means that not only AI development is not going to stop, but also that its velocity is not going to decrease. This is why we will see continuous investments in AI and more pro-growth national AI strategies akin to those that served as an entry point for this paper.

5 Coda: a roadmap for reversing the eclipse of reason

Despite how problematic the situation is nowadays, it is still possible to reverse the eclipse of reason. As Williams (2009: 73) reminds us, pragmatic questions about "what we should do" extend philosophical research on principles into the realm of politics, so to identify practical interventions situated within existing conditions and oriented toward what is actually feasible. In this regard, history offers solace and more than one reason for being optimistic. When Horkheimer wrote *Eclipse of Reason* at the end of World War Two, the kind of technological development that he was mostly preoccupied with was nuclear energy. Since then, the use of nuclear technologies has significantly improved and, while the specter of the atomic bomb still looms over us, there has not been another Hiroshima or Nagasaki. This improvement has not only been due to progress in philosophical inquiry and critical theory concerning technological innovation, but also to the establishment of solid governance systems that have limited the use of nuclear technology to certain fields, such as energy provision in particular (Judge et al., 2025). In addition, there have been plenty of cultural initiatives, from the renovation of public education curricula to the organization of exhibitions and the production of movies and novels, that in concert have raised a critical awareness across society about the very concrete dangers that the misuse of nuclear power poses (Ramana, 2011; Yim & Vaganov, 2003).

Moving from a logic of obstruction to a logic of competence, whereby critical theory can go beyond the analysis of what obstructs personal and international development, and shed light on what can foster capacities to develop both as individuals and a society (see Celikates, 2018), the paper now traces a roadmap indicating potential solutions to the problems posed today by the eclipse of reason. It is a philosophically grounded roadmap based on the assumption that, as acknowledged by Horkheimer (2013) himself, philosophy should not only provide explanations, but also identify goals or *ends* in his words, meant to orientate the development of society. This is of vital importance during the ongoing eclipse of reason which is making us lose our bearings, and this type of goals can offer directions to escape from its darkness or, at the very least, mitigate it.

As we have seen, the first core problem is the domination of subjective reason which, in the age of AI, generates two subproblems: the rise of oligarchies and undemocratic political systems sustained by technological (AI) systems, and an escalation of environmental degradation caused by the extensive production of AI technologies hungry for energy, water and raw materials. Here governance should come into play and target the following interconnected ends, in order to decrease the influence of subjective reason. One is *non-domination* intended as the "bedrock of justice", so to avoid a condition in which AI is in the power of a minority of individuals who

are abusing that power (Shapiro, 2012: 293). The other one is *environmental sustainability* as the preservation of ecosystems and the minimization of waste, carbon emissions and energy use, particularly in relation to the metabolism of the AI industry (James, 2014). Practical actions in line with these ends would include having parliaments and assemblies steer the course of AI development (Coeckelbergh, 2024), and incorporating clear environmental limits in the design of AI technology to prevent the formation of unsustainable AIs in the first place (Palmini & Cugurullo, 2024).

The second core problem is the formation of a world not of ends, but rather of instruments developed at a velocity that is greater than the speed at which society studies and comprehends them. This leads in turn to two AI-related subproblems: low levels of AI literacy and low quantities of AI technologies encapsulating higher concepts. In this context, governance should target *education* as the enhancement of the "participation of the individual in the social consciousness" of populations adopting AI (Dewey, 1897: 77); and *conceptual engineering* broadly understood as the design of higher concepts and their integration into AI technologies, so that such concepts can be widely put into practice as a form of "conceptual activism" (Chalmers, 2025: 2904). In this regard, practical interventions include investments in public education, exhibitions and art projects meant to increase people's consciousness of AI, its capabilities and impact (Coeckelbergh, 2024), and new conformity assessment protocols evaluating not only how an AI technology has been built from an engineering perspective, but also what concepts have informed its construction from a philosophical point of view.

The third core problem is governance itself as governments worldwide are not investing enough resources in systems, processes, norms and regulations to control the production of AI and direct its use towards higher concepts. This situation creates in turn two subproblems: the proliferation of empty discourses about higher concepts, such as sustainability, which mask unsustainable uses of AI, and a global impasse regarding how such situation can be rectified. Here governance should strive for truth, in line with the *correspondence theory of truth* where being true is associated with the actual state of affairs (Kirkham, 1992); and for *objective reason* (Horkheimer, 2013). Relatedly, levers of practical change could be found in *evidence-based policy* to ground AI projects in scientific evidence on their concrete risks and benefits (Head, 2010); and *multi-scalarity* which deserves an ad hoc final reflection.

As objective reason constitutes a "force" driving "relations among human beings, social classes, social institutions and nature" then its implementation, as we noted in the beginning of the paper, is a matter of scale (Horkheimer, 2013: 2). Multiple scales, to be precise, from that of the city, to regions, nations and the planet, so to encompass multiple human and non-human needs. This philosophy translates into *multi-scalar governance* which, in the age of AI, should include *global AI governance* as well as smaller scales which are particularly important, when it comes to questions of sovereignty, accountability and power relations. This is because we currently do not have international institutions and conventions capable of establishing planetary goals for AI development, in a manner that is binding, accountable and resistant to the growing power of AI companies (Feldstein, 2024).

Global AI governance is of course a road that should not be abandoned (Roberts et al., 2024), and the lessons from the governance of nuclear power mentioned above represent an important precedent. At the same time, it is key to remember that the smaller the scale,

the more functional governance systems tend to be. At the city scale, in particular, the question of sovereignty can be approached through already established techniques of participatory planning and governance, so to integrate citizens' opinions into processes of urban AI development (Cugurullo et al., 2023). Furthermore, at this scale, local governments can count on more refined legislations, compared to their international counterparts, that are helpful to establish more symmetrical power relations with AI companies, demand them accountable for their technologies and limit their use when citizens are against them (Cugurullo et al., 2023). If many cities start casting some light, then each light can become a star diminishing the darkness of the current eclipse of reason.

Acknowledgements My gratitude goes to the three anonymous reviewers for their generous feedback, and to Elisabetta Masala, Giulia Camba, Anne-Line Schalk and Luca Spano for the beautiful conversation that sparked my interest in writing this paper. I doubt that conversating with an AI would have generated the same results and, for me, these pages are a reminder of the importance of human connections.

Funding Open Access funding provided by the IReL Consortium. Irish Research Council (project ID: IRCLA/2022/3832).

Data Availability This paper is an original theoretical contribution that does not draw upon empirical data collected by the author.

Declarations

Generative AI and AI-assisted technologies in the writing process: no AI was used in the making of this paper.

Conflict of interest disclosure The author declares that there is no relevant or material financial interest that relates to the research presented in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Agrawal, A., Gans, J. S., & Goldfarb, A. (2023). Do we want less automation? *Science, 381*(6654), 155–158.

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2024.100173

Aristotle (2000). *The Politics*. Oxford University Press.

Aristotle (2004). *The Nicomachean Ethics*. Penguin Classics.

Bolte, L., & Van Wynsberghe, A. (2025). Sustainable AI and the third wave of AI ethics: A structural turn. *AI and Ethics, 5*(2), 1733–1742.

Braidotti, R., & Bignall, S. (Eds.). (2018). *Posthuman ecologies: Complexity and process after Deleuze*. Rowman & Littlefield.

Brevini, B. (2021). *Is AI good for the planet?* John Wiley & Sons.

Brown, T. (2016). Sustainability as empty signifier: Its rise, fall, and radical potential. *Antipode, 48*(1), 115–133.
Bullock, J. B., Chen, Y. C., Himmelreich, J., Hudson, V. M., Korinek, A., Young, M. M., & Zhang, B. (Eds.). (2024). *The Oxford handbook of AI governance*. Oxford University Press.
Celikates, R. (2018). *Critique as social practice: Critical theory and social self-understanding*. Bloomsbury.
Chalmers, D. J. (2025). What is conceptual engineering and what should it be? *Inquiry, 68*(9), 2902–2919.
Clark, D. R. (2003). Bricks, sweat and tears: The human investment in constructing a" Four-room" house. *Near Eastern Archaeology, 66*(1–2), 34–43.
Clark, J. N. (2023). Harm, relationality and more-than-human worlds: Developing the field of transitional justice in new posthumanist directions. *International Journal of Transitional Justice, 17*(1), 15–31.
CNBC (2025). President Donald Trump announces AI infrastructure investment — 1/21/2025. Available at: https://www.youtube.com/watch?v=zDo_RrzdRoQ [Accessed 29/11/2025].
Cocchiaro, M. Z., Morley, J., Novelli, C., Panai, E., Tartaro, A., & Floridi, L. (2025). Who is an AI ethicist? An empirical study of expertise, skills, and profiles to build a competency framework. *AI and Ethics, 5*, 3713–3725. https://doi.org/10.1007/s43681-024-00643-y
Coeckelbergh, M. (2024). *Why AI undermines democracy and what to do about it*. John Wiley & Sons.
Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2023). The AI gambit: Leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. *AI & Society*. https://doi.org/10.1007/s00146-021-01294-x
Cugurullo, F. (2024). The obscure politics of artificial intelligence: A Marxian socio-technical critique of the AI alignment problem thesis. *AI and Ethics, 5*(2), 1189–1201.
Cugurullo, F. (2025). AIdeology: Unpacking the Ideology of Artificial Intelligence and Its Spaces. *Antipode*. https://doi.org/10.1111/anti.70065
Cugurullo, F. (2026). New AI cities: Power, new cities and urban artificial intelligence in Neom and The Line. *Urban Geography*. https://doi.org/10.1080/02723638.2025.2548134
Cugurullo, F., & Acheampong, R. A. (2024). Fear of AI: An inquiry into the adoption of autonomous cars in spite of fear, and a theoretical framework for the study of artificial intelligence technology acceptance. *AI & Society, 39*(4), 1569–1584.
Cugurullo, F., Caprotti, F., Cook, M., Karvonen, A., McGuirk, P., & Marvin, S. (Eds.). (2023). *Artificial intelligence and the city: Urbanistic perspectives on AI*. Routledge.
Cugurullo, F., Caprotti, F., Day, J., Geoghegan, S., Lynch, C. R., Menga, F., Robinson, C., & Williams, J. (2025). The nature of AI: Metabolism, energy, water, labour and justice in the urban political ecology of artificial intelligence. *Urban Political Ecology*. https://doi.org/10.1177/30497515251344495
Dewey, J. (1897). My pedagogic creed. *The School Journal, 54*(3), 77–80.
Duvert, C. (2018). How is justice understood in classic Confucianism? *Asian Philosophy, 28*(4), 295–315.
Falk, S., van Wynsberghe, A., & Biber-Freudenberger, L. (2024). The attribution problem of a seemingly intangible industry. *Environmental Challenges*. https://doi.org/10.1016/j.envc.2024.101003
Feldstein, S. (2024). Evaluating Europe's push to enact AI regulations: How will this influence global norms? *Democratization, 31*(5), 1049–1066.
Floridi, L. (2021). The European legislation on AI: A brief analysis of its philosophical approach. *Philosophy & Technology, 34*(2), 215–222.
Floridi, L. (2024). Introduction to the special issues: The ethics of artificial intelligence: Exacerbated problems, renewed problems, unprecedented problems. *American Philosophical Quarterly, 61*(4), 301–307.
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, 689–707.
Floridi, L., Holweg, M., Taddeo, M., Amaya, J., Mökander, J., & Wen, Y. (2022). CapAI-A procedure for conducting conformity assessment of AI systems in line with the EU artificial intelligence act. Available at SSRN: https://ssrn.com/abstract=4064091 or https://doi.org/10.2139/ssrn.4064091 [Accessed 29/11/2025].
Foucault, M. (2002). *The Archaeology of Knowledge*. Routledge.
Hagendorff, T. (2022). A virtue-based framework to support putting AI ethics into practice. *Philosophy & Technology, 35*(3)
Hähnel, M., & Müller, R. (2025). *A Companion to Applied Philosophy of AI*. John Wiley & Sons.
Head, B. W. (2010). Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society, 29*(2), 77–94.
Hintz, M. J., Gross, M., Creutzig, F., & Kaack, L. H. (2026). Practical implementation of artificial intelligence for climate change mitigation in cities–priorities, collaborations and challenges. *Energy Research & Social Science*. https://doi.org/10.1016/j.erss.2025.104498

Horkheimer, M. (2013). *Eclipse of Reason*. Bloomsbury.
James, P. (2014). *Urban sustainability in theory and practice: circles of sustainability*. Routledge.
Johnson, S., & Acemoglu, D. (2023). *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. Basic Books.
Judge, B., Nitzberg, M., & Russell, S. (2025). When code isn't law: Rethinking regulation for artificial intelligence. *Policy and Society, 44*(1), 85–97.
Kirkham, R. L. (1992). *Theories of truth: A critical introduction*. MIT Press.
Kuntsman, A., & Xin, L. (2024). *Digital Technologies, Smart Cities, and the Environment: In the Ruins of Broken Promises*. Bristol University Press.
Machen, R., & Pearce, W. (2025). Anticipating the challenges of AI in climate governance: An urgent dilemma for democracies. *WIREs Climate Change*. https://doi.org/10.1002/wcc.70002
Mazzi, F., & Floridi, L. (Eds.). (2023). *The ethics of artificial intelligence for the sustainable development goals*. Springer.
Miller, D. (2013). *Justice for earthlings: Essays in political philosophy*. Cambridge University Press.
Natale, S. (2021). *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press.
Palmini, O., & Cugurullo, F. (2024). Design culture for Sustainable urban artificial intelligence: Bruno Latour and the search for a different AI urbanism. *Ethics and Information Technology, 26*(1), 11.
Pan, S., & Mou, Y. (2024). Constructing the meaning of human–AI romantic relationships from the perspectives of users dating the social chatbot Replika. *Personal Relationships, 31*(4), 1090–1112.
Pasquinelli, M. (2023). *The eye of the master: A social history of artificial intelligence*. Verso Books.
Plato (2007). *The Republic*. Penguin.
Ramana, M. V. (2011). Nuclear power and the public. *Bulletin of the Atomic Scientists, 67*(4), 43–51.
Risse, M. (2020). *On justice: Philosophy, history, foundations*. Cambridge University Press.
Roberts, H., Cowls, J., Hine, E., Morley, J., Wang, V., Taddeo, M., & Floridi, L. (2023). Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes. *The Information Society, 39*(2), 79–97.
Roberts, H., Hine, E., Taddeo, M., & Floridi, L. (2024). Global AI governance: Barriers and pathways forward. *International Affairs, 100*(3), 1275–1286.
Rosa, H. (2013). *Social acceleration: A new theory of modernity*. Columbia University Press.
Schmid, S., Lambach, D., Diehl, C., & Reuter, C. (2025). Arms race or innovation race? Geopolitical AI development. *Geopolitics*. https://doi.org/10.1080/14650045.2025.2456019
Shapiro, I. (2012). On non-domination. *University of Toronto Law Journal, 62*(3), 293–336.
Silvestro, D., Goria, S., Sterner, T., & Antonelli, A. (2022). Improving biodiversity protection through artificial intelligence. *Nature Sustainability, 5*(5), 415–424.
Swyngedouw, E., & Kaika, M. (2014). Urban political ecology. Great promises, deadlock… and new beginnings? *Documents D'Anàlisi Geogràfica, 60*(3), 459–481.
Tully, S. M., Longoni, C., & Appel, G. (2025). Lower artificial intelligence literacy predicts greater AI receptivity. *Journal of Marketing*. https://doi.org/10.1177/00222429251314491
Van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics, 1*(3), 213–218.
Wheeler, S. M. (2023). *The sustainable urban development reader*. Routledge.
White House (2025). President Trump Positions U.S. as Global Superpower in Manufacturing. Available at: https://www.whitehouse.gov/articles/2025/03/president-trump-positions-u-s-as-global-superpower-in-manufacturing/ [Accessed 29/11/2025].
Williams, B. (2009). *In the beginning was the deed: Realism and moralism in political argument.* Princeton University Press.
Xu, Y., Cugurullo, F., Zhang, H., Gaio, A., & Zhang, W. (2024). The emergence of artificial intelligence in anticipatory urban governance: Multi-Scalar evidence of China's transition to city brains. *Journal of Urban Technology, 32*(3), 9–33. https://doi.org/10.1080/10630732.2023.2292823
Yim, M. S., & Vaganov, P. A. (2003). Effects of education on nuclear risk perception and attitude: Theory. *Progress in Nuclear Energy, 42*(2), 221–235.
Zuboff, S. (2019). *The age of Surveillance Capitalism: The Fight for a Human Future at the new Frontier of Power*. Profile books.

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