AI and Ethics
https://doi.org/10.1007/s43681-024-00556-w

ORIGINAL RESEARCH

A critical inquiry into the personal and societal perils of Artificial Intelligence

Prokopis A. Christou[1]

Received: 8 May 2024 / Accepted: 20 August 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract
Artificial Intelligence (AI) has introduced unparalleled dynamics and possibilities for individuals, organizations, and society as a whole, though its rapid adoption has also sparked skepticism. This study examines the personal and societal perils associated with the widespread integration of AI, employing a methodological critical inquiry framed within a specific theoretical and allegorical context. The theoretical framework, which includes the triad of technological determinism, social construction of technology, and the information society is augmented by the theory of forms presented in Plato's allegory of the cave. The study highlights the distortion of personal and societal perceptions driven by AI, drawing parallels between the allegory's shadows and the manipulated realities AI can create. The study enhances understanding and prompts critical reflection on the implications and potential future personal and societal perils of AI, while advocating for ethical, responsible, evaluative, and human-centric development, application, and use of AI to ensure it benefits society.

Keywords AI impacts · AI and ethics · AI and society

1 Introduction

Artificial Intelligence (AI) has emerged as a transformative force in modern society, profoundly influencing diverse sectors ranging from healthcare to finance, and reshaping the contours of human interaction and capability. As a technological advance, AI's impact is multifaceted, characterized by both opportunities and challenges, fundamentally altering how societies function, businesses and organizations operate, and individuals interact, [1, 28, 33, 50]. However, the proliferation of AI also presents significant challenges, embracing various considerations, such as privacy threats, AI-driven cyber-attacks, bias, and offensive AI, which have become central concerns [20, 39, 40]. More specifically, AI systems can perpetuate existing biases if not properly designed and monitored. Privacy concerns arise with the extensive data collection required to train AI models. Furthermore, the increasing automation of jobs poses socioeconomic challenges, including job displacement and the need for workforce reskilling. On a global level, AI technology is a key factor in national security and international competitiveness, leading to an ongoing race for AI supremacy. Thus, AI as a technological tool is reshaping the world and society in profound ways with its influence spanning across multiple sectors, offering immense potential for progress and innovation. Yet, it simultaneously presents significant challenges that require careful consideration, As AI continues to evolve and advance, it is imperative for researchers, policymakers, technologists, and society at large to navigate these challenges, ensuring that AI development is ethical, equitable, and beneficial to society. By taking into account the theoretical perspectives on the evolution and impact of AI, it becomes imperative to conduct a critical inquiry study to deeply understand AI's influence on society and the potential threats it poses.

The pivotal aim of this study is to examine the societal perils associated with AI, through a methodological critical inquiry supported by a specific theoretical framework along with a philosophical allegory. The need of such research is particularly pronounced given the diverse ways in which AI can shape societal norms, values, and structures. Critical inquiry may lead to reasoned judgements on complex issues [3], while allegories, as time-honored narrative tools, hold significant value in research inquiry, for explaining current

✉ Prokopis A. Christou
Prokopis.christou@cut.ac.cy

[1] Cyprus University of Technology, Archiepiskopou Kyprianou 30, Limassol 3036, Cyprus

Published online: 16 September 2024

and complex phenomena [8, 21, 45]. The word allegory is derived from ancient Greek, from “allos,” meaning other, different, and “agoreuein,” meaning to speak in the agora (assembly). This etymology reflects the core concept of an allegory, such as in the form of a metaphoric story that can be interpreted to reveal a hidden meaning, typically a moral or political one. By distilling intricate and multifaceted issues into more familiar and relatable narratives, allegories facilitate a deeper understanding and engagement with the subject matter. They serve as bridges between abstract concepts and the lived experiences of individuals and provide a framework for understanding modern societal issues, offering a time-tested lens through which contemporary challenges can be viewed and analyzed. This approach enhances comprehension among diverse audiences and perhaps most importantly it stimulates critical thinking and reflection by drawing parallels between past wisdom and current realities. In a world where complex phenomena, such as AI technology, are increasingly prevalent, the use of allegories in research becomes a powerful tool to elucidate, connect, and convey profound insights in a more impactful and enduring manner. A critical inquiry study, utilizing this scientific theoretical and philosophical allegorical framework, can interrogate the extent to which AI shapes society. It allows for the exploration of how AI-driven ‘realities’ influence decision-making processes, both at individual and institutional levels, and the potential consequences of these influences at a societal level. Such a study is important for several reasons. Firstly, it encourages a deeper understanding of the societal implications of AI, moving beyond technical and efficiency-oriented perspectives. Secondly, it fosters critical and reflective thinking about AI and its role in society, prompting questions about autonomy, agency, and overreliance on AI-means. Thirdly, it provides an alternative, novel and profound understanding of how AI is reshaping reality, the societal threats it poses, and the ethical considerations it necessitates. It enables hence a more nuanced and holistic view of AI, one that recognizes its potential benefits while critically examining its broader impacts on society. The paper proceeds by justifying the rationale of using critical inquiry through the lens of an allegory, supported by a solid theoretical framework, to address the overall aim of this study.

1.1 Critical inquiry through the lens of allegory

Employing critical inquiry as a methodological approach can yield profound insights into a study area or a phenomenon [12, 46, 47], particularly when placed within a theoretical framework and intertwine with an allegorical perspective, like Plato’s allegory of the cave [54]. Such methodological approach may be regarded as adept at examining AI’s multifaceted impacts and perils at a societal level. In more detail, critical inquiry, by its nature, challenges existing assumptions and power structures. Through its argumentation and critical thinking dynamics it leads to reasoned judgements on complex issues [2, 3], and the generation of new knowledge [17]. When applied to the study of AI, it may scrutinize not only the technological dimensions but also the ethical, and societal implications. Such approach is particularly useful for delivering insightful information and facilitating the emergence of new knowledge. Besides, critical inquiry encourages researchers to delve deeper into an idiosyncratic study area, such as the societal implications of AI, embracing issues such as of privacy, autonomy, and the potential for AI to perpetuate societal biases. By questioning the status quo, this approach illuminates areas often overlooked in traditional analyses, such as the socio-political dimensions of AI technology. Moreover, critical inquiry fosters a reflective and reflexive stance in research. This reflexivity is vital in AI research, given the technology’s rapid evolution and its pervasive impact across diverse societal spheres.

The use of allegories like Plato’s cave is particularly effective in offering novel perspectives within the context of critical inquiry. Allegories allow complex and abstract concepts to be conveyed in a more tangible and relatable manner. As Berek [6] p.119) positions, ‘allegory says one thing and means another.’ Such metaphorical framing can lead to a deeper understanding of how AI led technology may shape human understanding and societal norms. Even so, they move beyond myths, since they provide a rational account in which reality in represented in a more abstract and universal way than myth allows [36]. Furthermore, allegories can bridge the gap between technical AI research and broader societal discourse. They provide a narrative that is accessible to a non-specialist audience, facilitating broader engagement with the ethical and social implications of AI, potentially leading to a more informed and nuanced academic but also public discourse about the role and impact of AI in society. Thus, critical inquiry, when combined with allegorical approaches, offers a powerful framework for studying AI perils. Such approach goes beyond technical and functional analyses to explore the deeper societal implications of AI. It challenges prevailing assumptions, uncovers hidden power dynamics, and fosters a more nuanced understanding of the technology’s (in the form of AI) impact. By drawing on allegories like Plato’s, researchers can communicate complex ideas in a more accessible and engaging manner, facilitating a richer and more inclusive discourse about the future of AI and its role in shaping human reality.

In this study, AI implications and impacts were examined, by incorporating a critical inquiry methodological approach, based on qualitative principles [31], aiming to advance knowledge on the topic under examination [17]. The first phase of the study involved a careful consideration of current studies linked to AI evolution and impacts. This was placed within a specific theoretical framework, to allow a critical evaluation of existing literature, and deep understandings of the phenomenon under study. The theoretical framework combined the crescendos of three main theories; the theory of technological determinism, social construction of technology theory, and the theory of information society. These are concerned with the adaptation, use, implications and impacts of technology (such as in the form of AI), and the interrelationship with key factors and domains, such as, society, and psychology. It may be argued that the triad of these specific theories, instead of only one, would have allowed a more holistic understanding of AI, its evolution and associated societal perils. During this stage, the author critically engaged with existing knowledge, identifying dominant narratives in the discourse surrounding AI and its implications. The second phase involved the careful examination of Plato's allegory of the cave, with an emphasis placed on its allegorical meaning, within a contemporary context. The third phase was concerned with an evaluation of the impacts and perils of AI with a critical lens [13], supported by a scientific theoretical framework combined with the dynamics of the proposed allegory, leading to certain conclusions, the delivery of further research avenues, and implications.

1.2 AI implications and impacts within the framework of technological determinism, social construction of technology, and information society

AI currently stands at the forefront of technological innovation, offering a number of benefits while simultaneously posing significant challenges, in various fields [7, 32]. This juxtaposition of AI's advantages and disadvantages has become a focal point for both technological development and ethical considerations in scholarly work [25, 49]. Among the most prominent benefits of AI is its capacity to process and analyze vast amounts of data at speeds and accuracies unattainable by human capability. This ability has profound implications across various sectors, such as in healthcare, environmental science, manufacturing, business, marketing, and services. Another significant advantage of AI is its role in advancing research. AI's capacity to identify patterns and correlations, also within large datasets, can lead to new scientific discoveries and technological innovations. Academics, have also argued the usefulness of AI as a useful and efficient research and analytical tool, if used in conjunction with human evaluative and critical skills [10].

Even so, the advancement of AI has brought a spectrum of challenges, drawbacks, and ethical dilemmas. One of the primary concerns is the issue of job displacement [19]. As AI and automation technologies become more prevalent, there is growing apprehension about the potential loss of employment, particularly in sectors reliant on routine, manual tasks. This shift necessitates a re-evaluation of job roles and a focus on workforce reskilling and upskilling to prepare for an increasingly automated future. Another significant drawback is the risk of bias and discrimination in AI algorithms [18], given that if not properly designed, AI systems can perpetuate existing societal biases. Data privacy and security are further challenges in the AI landscape. The vast amounts of data required to train and operate AI systems raise concerns about user privacy and data protection. Furthermore, the ethical implications of AI decision-making processes, especially in life-critical domains, are areas of intense debate [4, 26, 42]. Decisions traditionally made by humans are increasingly being delegated to AI, raising questions about moral and ethical judgment in AI systems and the accountability in the event of failures or mishaps. Thus, it is well acknowledged that AI presents a dichotomy of tremendous benefits and significant challenges.

As AI continues to evolve and permeate various aspects of life, it is imperative to address these challenges. The evolution and expansion of AI and its potential to create societal perils, such as overreliance on AI, can be elucidated through the lens of several relevant theories. Among them, the triad of technological determinism, social construction of technology, and the theory of the information society provide insightful perspectives. More specifically, technological determinism posits that technological developments drive or facilitate societal changes [51, 55]. In the context of AI, this theory suggests that the advancements in AI technology will inevitably shape and dictate future societal structures, behaviors, and norms. As AI becomes more sophisticated, it may result to an increased dependency on AI systems. However, this overreliance on AI can lead to a range of issues and risks [34, 38], such as diminished human decision-making and evaluative skills, loss of jobs, and a potential erosion of critical thinking abilities. As such, technological determinism warns of a future where human agency is significantly reduced, with AI dictating many aspects of life, from mundane daily activities to critical decisions in various fields.

Likewise, the theory of information society provides another useful theoretical perspective, focusing on the central role of information and technology in shaping contemporary society [5, 15]. Within the context of AI the theory may highlight the increasing availability and reliance on AI-generated information and data analytics. This eventually can lead to an 'information society' where AI becomes the primary source of knowledge and insight. This shift could create a societal threat in the form of 'information overload' or 'algorithmic governance,' where individuals and institutions rely heavily on AI for information processing and decision-making. Such overreliance might lead to vulnerabilities in critical thinking, privacy infringements, and the potential manipulation of information. In contrast, social construction of technology [14, 30], argues that society also shapes technological development. This theory is useful in understanding how societal choices, values, and power structures influence the direction and nature of AI development. For instance, if the societal emphasis is on efficiency and productivity, AI systems may be developed primarily to automate tasks, potentially leading to a workforce overly reliant on AI, reducing the value of human labor and skills across many sectors. Such overreliance could also manifest in decision-making processes, where the preference for AI-driven analytics and predictions could undermine human judgment, skills and expertise. As AI continues to evolve and expand, the theories above collectively betray a future where the impacts of AI on society will be profound and transformative. Technological determinism indicates a path where AI will shape societal behaviors, potentially leading to overreliance and reduced human agency. The information society theory points to the transformation into an AI-driven world, with significant reliance on AI for information and decision-making processes. All the same, social construction of technology emphasizes the role of societal values and decisions in steering AI development, potentially fostering or mitigating overreliance.

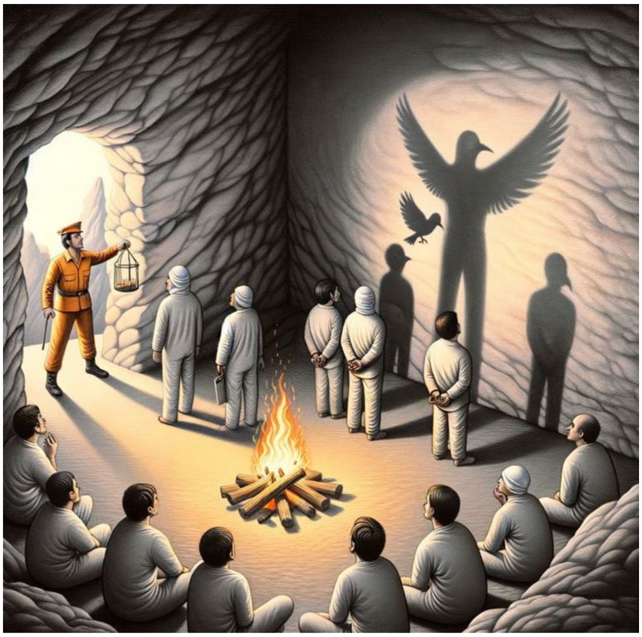

Fig. 1 Plato's allegory of the cave (Image created by ChatGPT4 through instructions provided by the author)

Plato's allegory of the cave offers an additional and profound prism, complementing and deepening the insights provided by the scientific theories of technological determinism, the information society, and the social construction of technology. As explained in more detail in the subsequent sections, the specific allegory, delivers a useful philosophical framework, supplementing the preceding scientific one, to explore the complex interplay between AI and society. While doing so, it highlights the risks of AI overreliance, the importance of critical awareness in a future seemingly AI-dominated world, and the potential for societal transformation through informed and ethical engagement with AI. This allegorical perspective, combined with the theoretical approaches mentioned, underscores the need for thoughtful consideration and proactive management and reflection upon AI and its societal impacts.

1.3 Plato's theory of forms and allegory of the cave

Plato, an ancient Greek philosopher born around 427 BC, is one of the most influential figures in philosophical thought, and intellectual history. His writings, mostly in the form of dialogues where characters discuss philosophical issues, cover a range of topics, such as epistemology, metaphysics, politics, and ethics. His theory of Forms or Ideas, which posits that non-physical forms represent the most accurate reality, has been a central theme throughout his work. The theory asserts that that the physical world is not really the 'real' world; instead, ultimate reality exists beyond our physical world. Plato's allegory of the cave, remains one of the most profound and enduring metaphors in Western philosophy [16, 29, 52]. The allegory, deeply embedded in the field of epistemology, provides a framework for understanding the nature of reality and human perception, positing a distinction between the world of appearances and the realm of true knowledge. The allegory depicts prisoners who have been confined in a cave since birth, bound in such a manner that they can only face forward, observing a wall. Behind them lies a fire, and between the prisoners and the fire is a walkway. Along this walkway, people carry objects that cast shadows on the wall. The prisoners, having never seen the actual objects, perceive these shadows as the most real and accurate representations of reality (refer to Fig. 1).

In this specific allegory, the cave symbolizes the sensory world, while the shadows on the wall represent the perceptions and beliefs based on sensory experiences. This state reflects the first stage in Plato's line of thought, where belief is based solely on what is seen and heard in the physical world. The prisoners' inability to see the objects casting the shadows illustrates the limitations of sensory experience and the superficial understanding (such as, of the real objects-world), that results from relying solely on such perceptions.

The philosopher then proceeds by introducing the concept of a journey out of the cave, towards the light (sun), representing the philosopher's path towards enlightenment, and true knowledge. One prisoner, freed from the chains, gradually ascends from the cave into the world above, while this transition is not without difficulty, which essentially represents that the path to knowledge can be challenging and painful. This ascent symbolizes the intellectual journey from 'ignorance' to knowledge, requiring the willingness to question previously held beliefs and assumptions. Upon reaching the outside world, the freed prisoner comes to understand that the sun, not the fire, is the true source of light and that what was experienced in the cave was mere illusion. This moment epitomizes the realization of the Theory of Forms - the understanding that the physical world is only a shadow of the true, and that ultimate reality exists beyond our physical world. The return of the freed prisoner to the cave signifies the philosopher's role in society, which faces the challenge of sharing this insight with those still imprisoned by their sensory perceptions. The resistance and disbelief met by the philosopher reflect the inherent difficulties in enlightening those who are accustomed to the shadows and skeptical of a reality they have never experienced. The allegory emphasizes the distinction between the perceived world and the world of true knowledge, explicates the journey from ignorance to knowledge, and underscores an obligation to impart knowledge to others. It furthermore highlights the transformative power of education and the need to question the apparent realities of the world.

2 Discussion

A seminal concept such as Plato's allegory of the cave, offers a profound framework for understanding the contemporary implications of AI and its societal perils. As explained in the preceding section, misinterpreting the shadows for reality, the prisoners live in a world of illusion until one is freed and comes to understand the true nature of the world. This allegory is remarkably applicable in the context of AI, offering insightful perspectives on how AI may shape perceptions and influence society. The power dynamics depicted in the allegory are reflective of those in the AI domain. Just as the prisoners' understanding of reality is controlled by those who project the shadows, in today's world, a small number of powerful tech companies and governments hold significant control over AI technologies. This concentration of power raises concerns regarding a number of issues, including amongst others what information is channelled to people (including students), and whether that information is accurate, reliable, and trustworthy. The allegory serves as a cautionary tale about the potential for AI to be used as a tool for manipulation and control, echoing the controlled reality experienced by the cave's prisoners. Of course, one may question, are we to be regarded and labelled as 'prisoners?'. Well, in its metaphoric sense, one may argue that people may be trapped in technological webs. While technology offers unprecedented conveniences and connectivity, it also ensnares users in a web of constant engagement and exposure, with concerns raising about mental health and personal autonomy. In today's digital era, people may be seen as 'prisoners' of technology, metaphorically shackled by their dependence on technology, with this captivity manifesting in an almost inseparable bond with smartphones, computers, AI tools and AI assistants, engendering hence a form of addiction [9, 35, 41].

Furthermore, AI, in its essence, functions as a creator of digital 'shadows,' curating the information and stimuli that individuals receive through various platforms. Like the shadows on the cave wall, the information provided by AI systems is often a distorted representation of reality. Social media algorithms, for example, tailor content based on user preferences and behaviors, creating echo chambers that reinforce existing beliefs and perceptions. This curated reality can lead to a skewed understanding of the world, much like the prisoners who perceive shadows as the entirety of existence and what is true. The allegory illustrates the potential of AI to limit and shape human understanding, perceptions, thoughts, and behaviour. As a vivid example of this situation within a contemporary AI context, De Vynck [11] referred to an 'AI deepfake apocalypse' being present. The author made reference to images that are created by AI which have become ubiquitous, often employed in creating unauthorized explicit content, distorting facts in political campaigns, and utilizing celebrity look-alikes for product promotion on social platforms. For instance, a video released by Princess Catherine revealing her cancer diagnosis in March 2024, sparked widespread discussion. Rumours circulated on social media, suggesting that the video was altered using AI. Although both BBC Studios, the creators of the video, and Kensington Palace refuted the involvement of AI technology, the rumours continued to circulate. Nonetheless, within a scientific context, researchers from various disciplines, while acknowledging possibilities of AI-assisted tools, increasingly make reference of how AI-driven programs and applications may perpetuate bias (Hagendorf et al., 2023; Timmons et al., [43, 53]. Despite this, as AI systems become more sophisticated and embedded in daily activities of people, they increasingly influence economic, social, and political arenas. It may be argued, that as society grows more dependent on AI, these programs could evolve to become key players and shapers of society by shaping perceptions, beliefs, norms, values and behaviours, through their further integration into communication, research, media, and education. This central role could shift power dynamics, where those who control AI technology wield significant influence, potentially leading to new forms of governance and societal structures centered around AI. In fact, in 2023 the former CEO and chairman of Google, expressed the following concerns, amongst others, at the JFK Forum:

Most concerning are the "extreme risks" of AI being used to enable massive loss of life if the four firms at the forefront of this innovation, OpenAI, Google, Microsoft, and Anthropic, are not constrained by guardrails and their financial incentives are "not aligned with human values" [44].

Another crucial aspect of the allegory is the journey of the freed prisoner who comes to understand the truth. This journey symbolizes the critical awakening needed to comprehend and respond to AI's societal impacts and possible threats. It underscores the need for critical thinking and awareness in the age of AI. Individuals, organizations, and institutes, much like the freed prisoner, should be able to understand, evaluate and if necessary question the algorithmically constructed realities (e.g., content, results, and information) presented to them. Obviously, this does not imply that they should learn from a technical viewpoint to understand 'algorithms' and the design process of AI and AI-assisted tools and programs. Instead, they must be able to identify the potential pitfalls, and threats of AI technology. Conceivably, this 'journey' is not without challenge, as the allegory suggests; understanding the broader impacts of AI requires effort and a willingness by individuals and organizations to confront possible uncomfortable truths about how these technologies influence society. For example, along with the potential benefits, researchers should be able to pinpoint and address the drawbacks of AI systems, while marketers must be aware of ethical considerations when dealing with data derived and analyzed through AI technological tools.

The ethical implications and societal risks associated with AI, as elucidated by the allegory, are manifold. The risks of AI (such as, bias), mirror the distorted reality of the cave's shadows. Indeed, AI systems, if not carefully designed and monitored, can perpetuate existing societal biases, and lead to unfair outcomes, for humans [22, 37], and animals [24]. This parallels the distorted perceptions based on the shadows, highlighting the importance of understanding and addressing the inherent biases in AI systems. In the context of AI's rapid advancement, certain individuals, like educators and academics, hold a crucial role akin to the escaped prisoner in the philosopher's allegory of the cave, tasked with the moral obligation to enlighten society about the potential perils of AI. Alike the escaped prisoner returns to the cave to inform others of the world beyond their limited perceptions, these individuals must educate the public about the implications of AI, which extend beyond the superficial allure of technological convenience, leisure, and efficiency. They have a duty to illuminate the ethical, social, and psychological impacts of AI, including issues of privacy, autonomy, and the potential for AI to perpetuate biases, create false content, or manipulate information. By raising awareness and fostering a nuanced understanding of AI, they could guide society towards informed and responsible usage and regulation of these technologies. In fact, the academic community commenced examining the multifaceted impacts of AI on society, and exploring how AI affects various aspects, such as privacy, healthcare, markets, social equity, and ethical boundaries [23, 27, 48]. It is important that they delve deeper into the consequences of AI in various field and contexts. Academics and other researchers play a crucial role in shaping public policy and industry practices by providing evidence-based analyses, specific frameworks, and practical recommendations. A closer collaboration with policymakers and stakeholders is hence stressed, to ensure that AI development and use aligns with anthropocentric, societal values and ethical standards. Furthermore, the allegory prompts reflection on the societal changes brought about by AI. Just as the freed prisoner experiences a profound transformation upon leaving the cave, society as a whole is undergoing a transformative process with the integration of AI. The allegory serves as a reminder of the need for continuous re-evaluation of how AI is integrated into various fields, such as, research, health, education, marketing, and business, how it impacts towards society, managed and addressed. Certain research avenues can help thus in the continuous re-evaluation of AI's role in society, ensuring that its integration is aligned with societal and anthropocentric values, and ethics. For instance, research can focus on developing and re-evaluating ethical frameworks specific to AI usage. Also, continuous studies are needed to explore how AI technologies are reshaping social dynamics, including work, education, and personal relationships. Further research of how AI technologies affect personal privacy and data security is also needed. Exploring the psychological effects of AI on individuals, such as dependency, changes in cognitive functions, and the impact on mental health, are also vital.

Overall, Plato's allegory of the cave reflects on the power dynamics of AI, highlights the potential of AI to create a controlled perception of reality, and emphasizes the importance of critical awareness and questioning in the age of AI. The depiction of prisoners confined to a cave, mistaking shadows of objects for reality, resonates powerfully with the potential societal impacts of AI. It serves as a metaphor for how AI, much like the shadows on the cave wall, can create a specific, seemingly broad but arguably limited or distorted version of reality. This parallels the concerns of the theory of technological determinism, where technology in the form of AI may shape societal norms and behaviors, potentially leading to overreliance on AI means, and a reduction in human agency. As individuals increasingly rely on AI for decision-making, from personal choices to critical decisions, there is a risk that this reliance mirrors the prisoners' acceptance of shadows as the complete truth. The ancient-perceived allegory warns of the contemporary dangers of society not questioning the AI-generated realities it delivers, potentially leading to a lack of critical thinking and appraisal of outcomes, and an unquestioned acceptance of algorithmically determined outcomes. Furthermore, in alignment with the information society theory, the allegory illuminates concerns about an AI-driven world where information processing and decision-making are predominantly or entirely entrusted to AI. Conversely, the social construction of technology, which emphasizes the role of societal values and decisions in shaping technology, aligns with the allegory's suggestion of enlightenment and the possibility of understanding and changing one's circumstances. That is, as the freed prisoner gains knowledge and seeks to enlighten others, AI designers, researchers, academics, and educators, have an obligation to investigate and evaluate the impacts of AI, shape AI's design, development and application according to rigorous ethical standards, equitable practices, and human-centric values. All the same, the role of academics and educators is highlighted in this regard, in terms of taking an active role in informing society of the potential perils of AI. This aspect of the allegory encourages active engagement, critical discourse, careful evaluation and rigorous reflection in AI development, use and channelling of information to society (e.g., children, students, institutes and businesses), about AI potential threats and perils.

3 Conclusion

In the profound landscape of AI, Plato's allegory of the cave offers a critical lens for scrutinizing the societal ramifications and impacts of AI. The allegory portrays prisoners mistaking shadows for reality, a scenario uncannily reminiscent of the modern human's increasing engagement (or should we state addiction?) with technology, and more specifically AI. The essence of the allegory, with its focus on illusion versus reality, directly parallels the controlled perceptions AI can engender in society, echoing the concerns of technological determinism. AI, akin to the cave's shadows, holds the potential to shape people's perceptions, societal norms and behaviors, riskily edging towards an overreliance that diminishes human agency. AI's power to curate and filter reality, much like the echo chambers of social media, warns of a possibly narrowed, biased comprehension of the world and its realities, a distortion akin to the prisoners' flawed perception. Moreover, AI's role in perpetuating existing biases and potentially unfair outcomes underscores the urgency to address ethical implications. In this context, the allegory implores a societal awakening akin to the freed prisoner's journey, demanding critical thinking, awareness, and a challenging of AI-induced realities, and AI-generated content. This awakening is not solely a technical understanding of AI algorithms but a broader, ethically grounded engagement with AI technology. It entails recognizing AI's societal impacts, balancing its conveniences against threats to 'true knowledge', mental health, psychology, relationships, and privacy. Key contributors to AI design and implementation (such as, computer scientists and engineers, machine learning engineers, software developers, data scientists, and user experience designers), along with ethicists, researchers, educators and academics emerge as modern-day counterparts to the allegory's enlightened prisoner, bearing the moral duty to illuminate AI's psychological, and social impacts. They play a crucial role in safeguarding society of the potential threats of AI, and guiding society towards informed usage and regulation of AI technologies. As a concluding remark, imagine a world that cannot distinguish between real and AI-manipulated videos depicting violent incidents, leading to a non-empathetic society that argues over even established and real facts. The long-honoured allegory highlights the necessity of safeguarding true facts and knowledge, freedom, integrity, fairness, trust, and human-centric values in a world that is becoming strongly influenced and driven by AI.

Funding No funding.

Data availability No data are available.

Declarations

Ethics approval and consent to participate No participants were involved in this study.

Informed consent Thus informed consent was not sought.

Conflict interests No conflict of interest.

References

1. Al-Surmi, A., Bashiri, M., Koliousis, I.: AI based decision making: Combining strategies to improve operational performance. Int. J. Prod. Res. **60**(14), 4464–4486 (2022)
2. Battersby, M., Bailin, S.: Inquiry: A new Paradigm for Critical Thinking. University of Windsor, Windsor, Ontario (2018)
3. Battersby, M., Bailin, S.: Critical inquiry: Considering the context. Argumentation. **25**, 243–253 (2011)
4. Belk, R.: Ethical issues in service robotics and artificial intelligence. Serv. Ind. J. **41**(13–14), 860–876 (2021)

5. Beniger, J.: The Control Revolution: Technological and Economic Origins of the Information Society. Harvard University Press (2009)
6. Berek, P.: Interpretation, allegory, and Allegoresis. Coll. Engl. **40**(2), 117–132 (1978)
7. Blanco-Gonzalez, A., Cabezon, A., Seco-Gonzalez, A., Conde-Torres, D., Antelo-Riveiro, P., Pineiro, A., Garcia-Fandino, R.: The role of Ai in drug discovery: Challenges, opportunities, and strategies. Pharmaceuticals. **16**(6), 891 (2023)
8. Brown, S., Stevens, L., Maclaran, P.: What's the story, allegory? Consum. Markets Cult. **25**(1), 34–51 (2022)
9. Chang, F.C., Chiu, C.H., Chen, P.H., Chiang, J.T., Miao, N.F., Chuang, H.Y., Tseng, C.C.: Smartphone addiction and victimization predicts sleep problems and depression among children. J. Pediatr. Nurs. **64**, e24–e31 (2022)
10. Christou, P.A.: The Use of Artificial Intelligence (AI) in qualitative research for Theory Development. Qualitative Rep. **28**(9), 2739–2755 (2023)
11. De Vynck, G.: The AI deepfake apocalypse is here. These are the ideas for fighting it. *The Washington Post.* Retrieved from: (2024). https://www.washingtonpost.com/technology/2024/04/05/ai-deepfakes-detection/. Accessed: 08/04/2024
12. Denzin, N.K.: Critical qualitative inquiry. Qualitative Inq. **23**(1), 8–16 (2017)
13. Denzin, N.K., Giardina, M.D. (eds.): Qualitative Inquiry through a Critical lens, pp. 1–9. Routledge, London (2016)
14. Douglas, D.G.: The Social Construction of Technological Systems, Anniversary Edition: New Directions in the Sociology and History of Technology. MIT Press (2012)
15. Duff, A.S.: Daniel Bell's theory of the information society. J. Inform. Sci. **24**(6), 373–393 (1998)
16. Ferguson, A.S.: Plato's simile of light. Part II. The allegory of the Cave (continued). Classical Q. **16**(1), 15–28 (1922)
17. Fielding, N.G.: Challenging others' challenges: Critical qualitative inquiry and the production of knowledge. Qualitative Inq. **23**(1), 17–26 (2017)
18. Gichoya, J.W., Thomas, K., Celi, L.A., Safdar, N., Banerjee, I., Banja, J.D., Purkayastha, S.: AI pitfalls and what not to do: Mitigating bias in AI. Br. J. Radiol. **96**(1150), 20230023 (2023)
19. Gruetzemacher, R., Paradice, D., Lee, K.B.: Forecasting extreme labor displacement: A survey of AI practitioners. Technol. Forecast. Soc. Chang. **161**, 120323 (2020)
20. Guembe, B., Azeta, A., Misra, S., Osamor, V.C., Fernandez-Sanz, L., Pospelova, V.: The emerging threat of Ai-driven cyber attacks: A review. Appl. Artif. Intell. **36**(1), 2037254 (2022)
21. Guenther, L.P.: Allegory analysis: A methodological framework for a tool for psychology. In: Re-Inventing Organic Metaphors for the Social Sciences, pp. 89–103. Springer International Publishing, Cham (2023)
22. Gupta, M., Parra, C.M., Dennehy, D.: Questioning racial and gender bias in AI-based recommendations: Do espoused national cultural values matter? Inform. Syst. Front. **24**(5), 1465–1481 (2022)
23. Hagendorff, T.: The ethics of AI ethics: An evaluation of guidelines. Mind. Mach. **30**(1), 99–120 (2020)
24. Hagendorff, T., Bossert, L.N., Tse, Y.F., Singer, P.: Speciesist bias in AI: How AI applications perpetuate discrimination and unfair outcomes against animals. AI Ethics. **3**(3), 717–734 (2023)
25. Hancock, J.T., Naaman, M., Levy, K.: AI-mediated communication: Definition, research agenda, and ethical considerations. J. Computer-Mediated Communication. **25**(1), 89–100 (2020)
26. Heinrichs, B.: Discrimination in the age of artificial intelligence. AI Soc. **37**(1), 143–154 (2022)
27. Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S.B., Koedinger, K.R.: Ethics of AI in education: Towards a community-wide framework. Int. J. Artif. Intell. Educ., 1–23. (2022)
28. Holmes, W., Tuomi, I.: State of the art and practice in AI in education. Eur. J. Educ. **57**(4), 542–570 (2022)
29. Huard, R.L.: Plato's Political Philosophy: The cave. Algora Publishing, New York (2007)
30. Klein, H.K., Kleinman, D.L.: The social construction of technology: Structural considerations. Sci. Technol. Hum. Values. **27**(1), 28–52 (2002)
31. Koro-Ljungberg, M., Cannella, G.S.: Critical qualitative inquiry: Histories, methodologies, and possibilities. Int. Rev. Qualitative Res. **10**(4), 327–339 (2017)
32. Liang, W., Tadesse, G.A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., Zou, J.: Advances, challenges and opportunities in creating data for trustworthy AI. Nat. Mach. Intell. **4**(8), 669–677 (2022)
33. Mariani, M.M., Perez-Vega, R., Wirtz, J.: AI in marketing, consumer research and psychology: A systematic literature review and research agenda. Psychol. Mark. **39**(4), 755–776 (2022)
34. Marr, B.: *The 15 biggest risks of Artificial Intelligence.* Forbes. Retrieved from: (2023). https://www.forbes.com/sites/bernardmarr/2023/06/02/the-15-biggest-risks-of-artificial-intelligence/?sh=2e5fa9e52706. Accessed: 11/04/2024
35. Marriott, H.R., Pitardi, V.: One is the loneliest number… two can be as bad as one. The influence of AI friendship apps on users' well-being and addiction. Psychol. Mark. **41**(1), 86–101 (2024)
36. Meeks, J.L.: Allegory in early Greek philosophy. In: Gungov, A., Verene, D.P. (eds.) Studies in Historical Philosophy, vol. 3. Ibidem (2020)
37. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. ACM Comput. Surv. (CSUR). **54**(6), 1–35 (2021)
38. Miller, K.: *AI overreliance is a problem. Are explanations a solution? Stanford* University, Human-Centered Artificial Intelligence. Retrieved from: (2023). https://hai.stanford.edu/news/ai-overreliance-problem-are-explanations-solution. Accessed: 11/04/2024
39. Mirbabaie, M., Brünker, F., Möllmann, N.R., Stieglitz, S.: The rise of artificial intelligence–understanding the AI identity threat at the workplace. Electron. Markets, 1–27. (2022)
40. Mirsky, Y., Demontis, A., Kotak, J., Shankar, R., Gelei, D., Yang, L., Biggio, B.: The threat of offensive ai to organizations. Computers Secur. **124**, 103006 (2023)
41. Murimi, R.: What Makes AI Addictive? The Role of Discounting, Risk Aversion and Self-regulation. In *International Conference on Big Data Intelligence and Computing* (pp. 451–462). Singapore: Springer Nature Singapore. (2022)
42. Nassar, A., Kamal, M.: Ethical dilemmas in AI-powered decision-making: A deep dive into big data-driven ethical considerations. Int. J. Responsible Artif. Intell. **11**(8), 1–11 (2021)
43. Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.E., Staab, S.: Bias in data-driven artificial intelligence systems—An introductory survey. Wiley Interdisciplinary Reviews: Data Min. Knowl. Discovery, **10**(3), e1356. (2020)
44. Pazzanese, C.: Former Google chairman details disruptions, dangers technology will bring to economy, national security, and other aspects of American life. *The Harvard Gazette.* Retrieved from: (2023). https://news.harvard.edu/gazette/story/2023/10/a-tech-warning-ai-is-coming-fast-and-its-going-to-be-rough-ride/. Accessed: 12/04/2024
45. Pongeluppe, L.S.: The allegory of the Favela: The Multifaceted effects of Socioeconomic mobility. Adm. Sci. Q., 00018392241240469. (2024)
46. Ren, C., Pritchard, A., Morgan, N.: Constructing tourism research: A critical inquiry. Annals Tourism Res. **37**(4), 885–904 (2010)
47. Ropers-Huilman, B.: Witnessing: Critical inquiry in a poststructural world. Int. J. Qualitative Stud. Educ. **12**(1), 21–35 (1999)

48. Quinn, T.P., Jacobs, S., Senadeera, M., Le, V., Coghlan, S.: The three ghosts of medical AI: Can the black-box present deliver? Artif. Intell. Med. **124**, 102158 (2022)
49. Safdar, N.M., Banja, J.D., Meltzer, C.C.: Ethical considerations in artificial intelligence. Eur. J. Radiol. **122**, 108768 (2020)
50. Salah, M., Halbusi, A., H., Abdelfattah, F.: May the force of text data analysis be with you: Unleashing the power of generative AI for social psychology research. Computers Hum. Behavior: Artif. Hum., 100006. (2023)
51. Smith, M.R., Marx, L. (eds.): Does Technology Drive History? The Dilemma of Technological Determinism. MIT Press (1994)
52. Świercz, P.: The allegory of the Cave and Plato's epistemology of politics. Folia Philosophica. **42**, 115–139 (2019)
53. Timmons, A.C., Duong, J.B., Fiallo, S., Lee, N., Vo, T., Ahle, H.P.Q., Chaspari, M.W., T: A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. Perspect. Psychol. Sci. **18**(5), 1062–1096 (2023)
54. Wright, J.H.: The origin of Plato's Cave. Harv. Stud. Classical Philology. **17**, 131–142 (1906)
55. Wyatt, S.: Technological determinism is dead; long live technological determinism. Handb. Sci. Technol. Stud. **3**, 165–180 (2008)

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.