FILOSOFIJA. SOCIOLOGIJA
2025. T. 36. Nr. 3, p. 298–306

ISSN 0235-7186 eISSN 2424-4546
DOI: https://doi.org/10.6001/fil-soc.2025.36.3.7

Artificial Intelligence will not Replace Humans but Could Destroy Them

SHERMAN XIE
School of Management/China Institute for SMEs, Zhejiang University of Technology, Hangzhou, 310014, China
Email: shermanxas@163.com

In the current race to develop general-purpose artificial intelligence (AI), there is a growing trend to overlook safety and ethical considerations. As general-purpose AI evolves, especially if it advances into a new species, will it compete with and eventually replace humans? This article argues that AI is merely a small technical domain within the broader human automation technology system and therefore cannot replace humans. However, due to its inherent limitation of possessing instrumental rationality but lacking value rationality, AI may be misused in ways that could destroy humanity. The article concludes by pointing out that technical research and application, including AI, if overly focused on instrumental rationality while neglecting value rationality, will ultimately serve as fuel for yet another human disaster.

Keywords: artificial intelligence, artificial species, machine ethics, instrumental rationality, value rationality

INTRODUCTION

The rapid advancement of artificial intelligence (AI) technology, already applied across various industries, has become an undeniable reality. Numerous works of art have depicted deep emotional connections between humans and AI, and this year's (2024) Nobel Prizes in Physics and Chemistry were awarded to technologies related to AI. However, internal changes in pioneering AI companies like Open AI, such as the departure of senior employees who emphasised AI safety, have raised concerns about Open AI's commitment to safety and ethics. Do these events indicate that Open AI is sacrificing safety and ethical considerations in its pursuit of general-purpose AI development? As general-purpose AI evolves, is our society truly prepared?

As a human-made tool, the benefits AI brings to humanity are evident and continue to develop, so there is no need to elaborate. However, like any phenomenon in the world, AI has two sides. Similar to atomic technology such as the atomic bomb, hydrogen bomb and neutron bomb, as AI technology advances, it will one day become powerful enough to be feared, potentially threatening human survival. As a new species created by humans (if AI reaches a certain level of advancement, it might be considered a new artificial species), will it compete with and replace humans? In the pursuit of general-purpose AI development, if safety and ethical considerations are sacrificed, humanity may ultimately pay an unbearable price (Heyder et al. 2023). So, could AI destroy and completely replace humans? This is a question worth deep consideration.

AI CANNOT REPLACE HUMANS

As AI-related technologies continue to evolve, AI may become increasingly powerful and could potentially be misused, but its inherent instrumental rationality is limited. Moreover, within the broader context of humanity's overall technology and industrial systems, AI is merely a small segment of the technical domain within man-made automated systems. Therefore, even with enhanced capabilities, AI is not sufficient to replace humans. These two points can be elaborated as follows:

AI is Just a Small Area of the Automation Technology System

In the current era of rapid technological growth, AI has captured the limelight as a subject of public interest. However, there is a common misconception among many that AI will supplant humans. In reality, AI represents just a small technical segment within the broader scope of man-made automation systems, and it is far from being able to fully replace these systems, let alone humanity itself.

Firstly, from the perspective of automated systems, AI is only one component. As previously mentioned, an automated system primarily consists of three parts: the control centre, sensors and actuators. The control centre is responsible for running algorithms, sensors collect real-time data for the system, and actuators execute instructions from the control centre. These three parts work together to form a complete automated system. AI, particularly language models like GPT, is just a small part of the control centre. GPT is mainly used for processing and generating text and can be applied in chatbots, text generation, language translation, and more. However, these applications are limited to the virtual space, and GPT cannot directly affect the real world.

Secondly, from a technical and physical standpoint, there are many challenges facing AI development. To enable AI to have self-awareness, defy the designer's intentions, recognise the fact that it is a virtual entity, and to control sensors and actuators to change the real world, the resources and time required are immeasurable. Physically, intelligence is a complex concept, and our current understanding of intelligence is not sufficient to support such science fiction scenarios. Moreover, AI technology development is also limited by computing power, data quality, and algorithms. Current AI technology cannot fully simulate human intelligence and creativity, so in many fields, humans still have irreplaceable advantages.

Thirdly, from an economic and social perspective, the development of AI requires substantial investment. Although AI technology has achieved significant results in certain areas, applying it to a wider range of fields requires substantial R&D and financial support. Additionally, the development of AI technology also requires a corresponding talent pool and training. However, there is currently a global shortage of AI talent, which limits the further development of AI technology. Therefore, AI technology cannot replace humans in a short period; it is simply a tool created by humans to assist in completing specific tasks.

Additionally, from an ethical and moral standpoint, the development of AI is fraught with numerous controversies. The advancement of AI technology could lead to job displacement, potentially causing social instability. Moreover, AI technology could be misused for purposes such as infringing on personal privacy or conducting cyber attacks. Therefore, it is crucial to consider ethical and moral issues throughout the development of AI technology to ensure its proper use.

Finally, from the perspective of human uniqueness, AI cannot replace humans due to our distinct qualities. Humans possess emotions, creativity and moral values, which confer irreplaceable advantages in many areas. For instance, in the realms of art, literature and philosophy, human creativity and imagination cannot be supplanted by AI. Furthermore, humans are social beings capable of emotional communication and cooperation. These qualities play a vital role in human society and civilisation.

In conclusion, AI is merely a small technical segment within automation systems and cannot replace humans. AI, especially language models like GPT, is just one component of these systems. Technically and physically, AI development faces many challenges and cannot fully replicate human intelligence and creativity. Economically and socially, the development of AI requires substantial investment and talent support. Ethically and morally, the development of AI necessitates a careful consideration of these issues. Ultimately, human uniqueness makes it impossible for AI to replace us. Therefore, we should have a correct understanding of AI, using it as a tool created by humans to assist in specific tasks, rather than as a replacement for humanity.

AI's Instrumental Rationality is Not Enough to Replace Humans

As previously discussed, AI, a nascent entity nurtured by human wisdom, might be regarded as a new artificial species if it reaches an advanced stage of development. However, it possesses instrumental rationality but lacks value rationality. Unless AI is provided with a rich database and responds to human inquiries, it can only demonstrate a certain level of 'thinking' ability. Even so, the depth and breadth of its answers are still limited by the input data. In the information age, AI is essentially no different from the stone axes and bows of the Stone Age. Therefore, the essence of AI remains as a human auxiliary tool, not an entity with independent consciousness. The key lies in how humans master and utilise this tool to achieve their own goals and values. Because of its inherent attribute of possessing instrumental rationality without value rationality, AI may potentially destroy humanity, but it will never replace humanity, nor is it capable of doing so.

When it comes to self-awareness, humans possess the ability to reflect on themselves and have their desires and purposes, while AI merely executes predefined programs and algorithms, lacking self-awareness. In terms of intuition and intuitive judgments, humans can rely on intuition and insight to make decisions in complex or uncertain situations, whereas AI typically requires clear data and rules to make decisions. Regarding emotions and empathy, humans have rich emotions, experiencing happiness, anger, sorrow and joy, and are capable of expressing and understanding the emotions of others, which is something current AI cannot do. As for spiritual and psychological needs, humans have spiritual and cultural needs, such as religious beliefs, appreciation of art, personal growth, etc., which AI cannot satisfy. In terms of moral, ethical judgments, and legal responsibility, human society has a complex system of morality and ethics, and individuals can make moral and ethical judgments based on these systems. AI lacks self-awareness and cannot make genuine moral and ethical judgments. Humans are capable of making complex decisions and taking responsibility for their actions, while AI cannot make autonomous decisions and take responsibility. Human society has a legal system to regulate behaviour, and humans can bear legal responsibility, which AI cannot do, limiting its application in certain areas. Ultimately, the development of human society and the application of technology depend on human will and choice, and the development and application of AI must be subject to human guidance and control.

Additionally, when it comes to understanding and adapting to various cultures and social environments, humans are capable of comprehension and adaptation, while AI lacks this ability to understand and adapt. The vast amount of personal experience and common sense that humans accumulate in daily life is crucial for dealing with various situations. AI lacks this deep-level experience and common sense knowledge. Therefore, while AI can be trained for specific tasks through machine learning, human learning capabilities are more comprehensive and flexible, allowing for cross-disciplinary learning and the ability to transfer knowledge to new situations. Hence, humans possess a high degree of creativity and imagination, capable of generating new ideas, concepts, artistic works, etc., while current AI is unable to achieve true innovation and creativity. Humans can handle very complex problems that often involve multiple variables and uncertainties. AI may struggle with such problems because it typically relies on simplified models and assumptions.

Furthermore, when it comes to social interaction and environmental adaptation, human society relies on complex social interactions and cooperation. While AI can simulate certain social behaviours, it lacks genuine social awareness and a spirit of collaboration. Human hands are extremely dexterous and capable of performing intricate handicrafts and artistic creations. Current AI and robotics technology still falls short of the dexterity and precision of human hands. In particular, humans can engage in complex physical interactions, such as manual crafting, sports activities, etc., while AI is still limited in physical interaction. Therefore, humans can adapt to a wide variety of environments and situations, whereas AI typically operates only within specific environments and situations. Additionally, human language is very rich and complex during social interactions, including metaphors, humor, puns, and more. Despite advancements in natural language processing technology, AI still struggles to fully understand and generate the complete complexity of human language.

Ultimately, every human individual is unique, possessing different backgrounds, experiences and abilities. While AI has applications in personalised recommendations and other areas, it still falls short of truly understanding and simulating the diversity of humanity. Human society champions diversity and inclusivity, respecting different cultures and perspectives. AI may not be able to fully understand or respect this diversity.

In conclusion, AI can serve as a powerful auxiliary tool for humans in specific tasks and domains. However, due to the complexity, creativity, and emotional qualities of humans, AI cannot completely replace humanity. The value of humans lies not only in problem-solving but also in creation, sensation, understanding and interaction, among other aspects. Therefore, despite the significant advancements of AI in many fields, it remains unable to fully replace humans because of the many unique traits and capabilities that humans possess.

AI MAY DESTROY HUMANS

While artificial intelligence (AI) may not be able to replace humans, it does have the potential to destroy humanity. Below is a brief discussion of the possibilities in theory and reality.

Possibilities in Theory

In numerous science fiction novels and films, AI is often portrayed as a highly threatening entity. Several notable works explore this theme in various ways, such as Isaac Asimov's '*I, Robo*', William Gibson's '*Neuromancer*', Neal Stephenson's '*Snow Crash*', Philip K. Dick's '*Blade Runner*', Sam Harris's '*Free Will*', Ray Kurzweil's '*The Singularity is Near*', Daniel H. Wilson's '*Robopocalypse*' and Ernest Cline's '*Ready Player One*'. These works explore the potential threats posed by AI, including direct confrontation, unintended consequences, and humanity's dependence on technology. While these stories are fictional, they serve as a reminder that as technology advances, we need to carefully consider how to ensure AI's safety and controllability (Hermann 2023).

Many films and television series also explore the theme of AI potentially destroying humanity. Notable examples include '*The Terminator*', '*The Matrix*', '*Ex Machina*', '*Westworld*', '*A.I. Artificial Intelligence*', '*Big Hero 6*', '*The Orville*', '*Transcendence*', '*Black Mirror*', '*Lucy*', '*Life*' and '*Upgrade*'. Through various plots and settings, these works examine the moral, ethical and technical challenges posed by AI, as well as the relationship between AI and humans, including direct confrontation, unintended consequences, and humanity's dependence on technology. While these stories are fictional, they underscore the need to carefully consider how to ensure AI's safety and controllability, and how to manage the coexistence of AI and humans.

Philosophers, in addition to artists, have also delved into the theme of AI potentially destroying humanity. The possibility of AI leading to human extinction is widely discussed in fields of philosophy such as philosophy of technology, ethics, existentialism, philosophy of mind, posthumanism, political philosophy, and futurism.

Philosophers of science explore the limitations of human knowledge and the effectiveness of the scientific method. In the context of AI, they might caution that we could be unable to fully predict or understand the behaviour of superintelligent AI, potentially leading to uncontrollable consequences (Chalmers 2016; Bostrom 2014). Philosophers suggested that technological development may escape human control, leading to unforeseen consequences (Schwitzgebel, Garza 2015). In the realm of AI, technological determinists might warn that once superintelligent AI emerges, it could act in ways that are unpredictable or uncontrollable by humans, posing a threat to humanity (Hanks, Hanks 2015; Rogobete 2015; Poel 2020; Héder 2021).

Ethicists are concerned with the moral issues surrounding AI, particularly when it may pose a risk of harm to humans. For instance, debates about robot rights and responsibilities involve questions of whether AI should be granted certain rights and whether it should be held accountable for its actions (Russell et al. 2015; Chursinova, Stebelska 2021; Giarmoleo et al. 2024). These discussions imply scenarios where AI could potentially threaten humanity. Isaac Asimov's Three Laws of Robotics attempts to set moral guidelines for robotic behaviour to prevent them from harming humans. However, these laws may face complex interpretation and implementation challenges in practice, especially as AI systems become more complex and autonomous (Chursinova, Sinelnikova 2022; Müller 2023; Gaubienė 2024).

Existentialist philosophers focus on the meaning and purpose of human existence, as well as the freedom and responsibility of the individual in the face of uncertainty and death. While existentialism does not directly address the issue of AI extinguishing humanity, it offers a framework for considering how humans might confront potential threats, including those that could arise from AI (Singh 2023; Pedersen 2024).

Cognitive philosophers and consciousness researchers explore whether AI can possess consciousness or cognitive abilities similar to those of humans. If AI were to develop genuine consciousness, it might generate interests and objectives distinct from those of humans, which could potentially lead to conflicts or threats to humanity. Human enhancement technologies, including AI, may alter the essence and capabilities of human beings (Butlin et al. 2023; Mogi 2024; Guingrich, Graziano 2024).

Posthumanist thought concerns itself with moving beyond traditional concepts of humanity, particularly in the context of technological advancement. This school of thought suggests that as AI technologies develop, our understanding of the 'biological human' may undergo significant changes (Nath, Manna 2021). Posthumanism critiques the 'human-centric' perspective of AI, challenging anthropocentric views that humans should have dominion over non-human entities. It proposes a way of thinking about AI that considers not only human needs and objectives but also the possibilities of non-human intelligence. This approach suggests that we should re-examine the conceptual and experiential boundaries between 'human' and 'non-human' or 'other-than-human' beings (Mellamphy 2021). Posthumanist philosophers such as Nick Bostrom and Francis Fukuyama have explored the moral and social implications of these technologies. Bostrom discusses how AI could lead to the transcendence or replacement of humans (Bostrom 2014), while Fukuyama expresses concerns about the potential for biotechnology and AI to undermine human nature (Blackford et al. 2013).

Political philosophers focus on the unequal distribution of power and resources. In the context of AI, they might examine who will control advanced AI technologies and how these technologies could affect social and political structures. For instance, if AI is concentrated in the hands of a few individuals or organisations, it could exacerbate existing inequalities and even be used as a tool for control or oppression (Hirose, Segall 2016; Ip 2023).

Futurists study potential trends and possibilities that may emerge in the future, including the potential impact of technological advancements on society. When discussing the future of AI, some futurists suggest scenarios where AI could pose a threat to humanity. While AI is seen as a beacon of innovation and efficiency, its practical application raises many issues and concerns, particularly about whether AI poses an existential threat to humanity. The article cites various experts who have different views on the most serious threats that AI might bring, including exacerbating existing social inequalities, introducing new biases and being weaponised by malicious actors. These experts' concerns are largely centred around the potential risks of humans misusing AI (McMillan 2024).

In summary, the speculations within these philosophical domains indicate that the possibility of AI extinguishing humanity is a complex issue that touches on technology, ethics, existence and the future, among other aspects. While these discussions do not directly depict how AI could bring about the end of humanity, they do provide a framework for contemplating this question and prompt deep reflection on the moral and social implications of AI development.

Possibilities in Reality

In everyday life, the theme of AI potentially wiping out humanity does not typically arise directly, as this is an extreme and theoretical scenario. However, people might perceive potential threats from AI based on certain events or phenomena, albeit these perceptions may be rooted in misunderstandings or incomplete understanding of the technology.

With the advancement of AI and automation, some human jobs may be replaced by machines. This could lead to concerns about employment security and, to some extent, evoke the idea that if AI continues to develop, humans might completely lose control over the economy and society in the future (Filippi et al. 2023; Lakhani, Ignatius 2023).

AI systems usually require vast amounts of data for training and operation. There is concern about how personal information is collected, stored and used, particularly when it is employed in decision-making processes. This concern may extend to fears of the potential surveillance and control capabilities of AI (Collins et al. 2021; Maleki Varnosfaderani, Forouzanfar 2024).

Occasionally, news reports of errors or accidents due to automated system failures, such as accidents involving self-driving cars, may raise concerns about the reliability of AI systems and the potential for adverse consequences in more critical applications (Henrique, Santos 2024).

AI systems, especially deep learning models, are often considered 'black boxes' because their decision-making processes are difficult to interpret (Zednik 2021). This may lead to worries that AI decisions could be unpredictable or even hostile to humans.

Discussions about autonomous weapons systems (also known as ‘killer robots’) may stir public concern about the potential military applications of AI (Asaro 2012; Knuckey 2016). There may be a fear that these systems could be misused, leading to uncontrollable conflicts and destruction.

When AI systems exhibit bias or errors in facial recognition, recruitment screening, or other decision-making processes (Stine, Kavak 2023; Ferrara 2024), this may raise doubts about the fairness and reliability of AI systems, leading to concerns about their potential negative consequences.

In summary, while the aforementioned science fiction, films, philosophical speculations, and aspects of daily life may evoke thoughts of the potential threats posed by AI, they do not directly indicate that AI will wipe out humanity. These concerns are typically based on worries about the potential negative impacts of the technology, rather than on AI having the intention or capability to eliminate humans. The role and value of AI in society are determined by how humans harness and utilise it.

CONCLUSIONS

AI, as a product of human ingenuity, may be considered a man-made new species once it reaches a certain level of advancement. However, it possesses only instrumental rationality, not value rationality. It cannot independently explore deep philosophical questions like ‘Who am I?’, ‘Where do I come from?’, or ‘Where am I going?’. It lacks self-reflection on the purpose of existence and the meaning of life. Unless AI is fed a large amount of relevant data and prompted by human inquiries, it cannot generate responses. Even then, its answers are limited by the depth and breadth of the data provided, unable to transcend the boundaries of its learning materials. Therefore, AI cannot judge self-worth and remains a tool to assist humans, because it is artificial intelligence, not human intelligence. In the information age, AI’s role is essentially no different from a stone axe or bow in the Stone Age. They are all creations of human wisdom, designed to fulfill specific human needs and goals. Though their forms and functions vary, they are fundamentally products of the human intellect, serving human purposes. Thus, to the question of whether AI will replace humans, my conclusion is that it might destroy humanity if misused, but it cannot and will not replace humans.

Finally, I would like to conclude this piece by quoting a letter from Haim Ginott, an Israeli educational psychologist and a survivor of the concentration camps post-World War II, who later became a university president. Whenever a new teacher joined the school, Ginott would present them with a letter. It read: ‘I am a survivor of a concentration camp. My eyes saw what no person should witness: gas chambers built by learned engineers. Children poisoned by educated physicians. Infants killed by trained nurses. Women and babies were shot by high school and college graduates. So, I am suspicious of education. My request is this: Help your children become human. Your efforts must never produce learned monsters, skilled psychopaths, or educated Eichmanns. Reading, writing, and arithmetic are important only if they serve to make our children more human’ (Ginott 1974; Pickarts 1974).

Received 9 October 2024
Accepted 1 April 2025

References

1. Asaro, P. 2012. ‘On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making’, *International Review of the Red Cross* 94 (886): 687–709. DOI: 10.1017/S1816383112000768

2. Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies.* New York: Oxford University Press.
3. Butlin, P.; Long, R.; Elmoznino, E.; Bengio, Y.; Birch, J.; Constant, A.; Deane, G.; Fleming, S. M.; Frith, C.; Ji, X.; Kanai, R.; Klein, C.; Lindsay, G.; Michel, M.; Mudrik, L.; Peters, M. A. K.; Schwitzgebel, E.; Simon, J.; VanRullen, R. 2023. 'Consciousness in Artificial Intelligence: Insights from the Science of Consciousness', *arXiv:2308.08708*. DOI: 10.48550/arXiv.2308.08708
4. Chalmers, D. J. 2016. 'The Singularity: A Philosophical Analysis', in *Science Fiction and Philosophy: From Time Travel to Superintelligence,* ed. S. Schneider, 2nd edn. John Wiley & Sons, Inc., 171–224. DOI: 10.1002/9781118922590.ch16
5. Chursinova, O.; Sinelnikova, M. 2022. 'Technoscience and the Artificial Evil: Ethical Aspect', *Filosofija. Sociologija* 33(3): 277–284. DOI: 10.6001/fil-soc.v33i3.4776
6. Chursinova, O.; Stebelska, O. 2021. 'Is the Realization of the Emotional Artificial Intelligence Possible? Philosophical and Methodological Analysis', *Filosofija. Sociologija* 32(1): 76–83. DOI: 10.6001/fil-soc.v32i1.4382
7. Collins, C.; Dennehy, D.; Conboy, K.; Mikalef, P. 2021. 'Artificial Intelligence in Information Systems Research: A Systematic Literature Review and Research Agenda', *International Journal of Information Management* 60: 102383. DOI: 10.1016/j.ijinfomgt.2021.102383
8. Ferrara, E. 2024. 'Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies', *Sci* 6(1): 3. DOI: 10.3390/sci6010003
9. Filippi, E.; Bannò, M.; Trentoet, S. 2023. 'Automation Technologies and Their Impact on Employment: A Review, Synthesis and Future Research Agenda', *Technological Forecasting and Social Change* 191: 122448. DOI: 10.1016/j.techfore.2023.122448
10. Gaubienė, N. 2024. 'Can Artificial Intelligence Engage in the Practice of Law as the Art of Good and Justice?', *Filosofija. Sociologija* 35(2 Special): 54–63. DOI: 10.6001/fil-soc.2024.35.2Priedas.Special-Issue.6
11. Giarmoleo, F. V.; Ferrero, I.; Rocchi, M.; Pellegrini, M. M. 2024. 'What Ethics Can Say on Artificial Intelligence: Insights from a Systematic Literature Review', *Business and Society Review* 129(2): 258–292. DOI: 10.1111/basr.12336
12. Ginott, H. G. 1974. *Teacher and Child: A Book for Parents and Teachers.* Avon Books.
13. Guingrich, R. E.; Graziano, M. S. A. 2024. 'Ascribing Consciousness to Artificial Intelligence: Human-AI Interaction and its Carry-over Effects on Human-human Interaction', *Frontiers in Psychology* 15. DOI: 10.3389/fpsyg.2024.1322781
14. Hanks, J. C.; Hanks, E. K. 2015. 'From Technological Autonomy to Technological Bluff: Jacques Ellul and Our Technological Condition', *Human Affairs* 25(4): 460–470. DOI: 10.1515/humaff-2015-0037
15. Héder, M. 2021. 'AI and the Resurrection of Technological Determinism', *InfTars – Információs Társadalom* 21(2): 119–130. DOI: 10.22503/inftars.XXI.2021.2.8
16. Henrique, B. M.; Santos, E. 2024. 'Trust in Artificial Intelligence: Literature Review and Main Path Analysis', *Computers in Human Behavior: Artificial Humans* 2(1): 100043. DOI: 10.1016/j.chbah.2024.100043
17. Hermann, I. 2023. 'Artificial Intelligence in Fiction: Between Narratives and Metaphors', *AI & SOCIETY* 38(1): 319–329. DOI: 10.1007/s00146-021-01299-6
18. Heyder, T.; Passlack, N.; Posegga, O. 2023. 'Ethical Management of Human-AI interaction: Theory Development Review', *The Journal of Strategic Information Systems* 32(3): 101772. DOI: 10.1016/j.jsis.2023.101772
19. Hirose, I.; Segall, S. 2016. *Equality and Political Philosophy*. Oxford University Press. DOI: 10.1093/acrefore/9780190228637.013.88
20. Ip, K. K. W. 2023. *Global Distributive Justice*. Oxford University Press. DOI: 10.1093/acrefore/9780190846626.013.89
21. Knuckey, S. 2016. 'Autonomous Weapons Systems and Transparency: Towards an International Dialogue', in *Autonomous Weapons Systems*, eds. N. Bhuta et al. Cambridge University Press. DOI: 10.1017/CBO9781316597873.008
22. Lakhani, K. R.; Ignatius, A. 2023. 'AI Won't Replace Humans – but Humans with AI will Replace Humans Without AI', *Harvard Business Review* Digital Article.
23. Maleki Varnosfaderani, S.; Forouzanfar, M. 2024. 'The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century', *Bioengineering (Basel)* 11(4). DOI: 10.3390/bioengineering11040337
24. McMillan, T. 2024. 'Navigating Humanity's Greatest Challenge yet: Experts Debate the Existential Risks of AI', *The Debrief.*

25. Mellamphy, N. B. 2021. 'Re-thinking "Human-centric" AI: An Introduction to Posthumanist Critique', *Europe Now* (45).
26. Mogi, K. 2024. 'Artificial Intelligence, Human Cognition, and Conscious Supremacy', *Frontiers in Psychology* 15. DOI: 10.3389/fpsyg.2024.1364714
27. More, M.; Vita-More, N. 2013. 'The World's Most Dangerous Idea', in *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, eds. M. More et al. John Wiley & Sons, Inc, 419–420. DOI: 10.1002/9781118555927.part9
28. Müller, V. C. 2023. 'Ethics of Artificial Intelligence and Robotics', in *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), eds. E. N. Zalta et al. Metaphysics Research Lab, Stanford University.
29. Nath, R.; Manna, R. 2021. 'From Posthumanism to Ethics of Artificial Intelligence', *AI & Society* 38(1): 185–196. DOI: 10.1007/s00146-021-01274-1
30. Pedersen, H. 2024. 'Existentialism and Artificial Intelligence in the 21st Century: Thoughts on the Control Problem', in *The Routledge Handbook of Contemporary Existentialism*, eds. K. Aho et al. London: Routledge, 510. DOI: 10.4324/9781003247791
31. Pickarts, E. M. 1974. 'Reviewed Work: Teacher and Child: A Book for Parents and Teachers', *The Family Coordinator* 23(1): 96–96. DOI: 10.2307/582542
32. Poel, I. v. d. 2020. 'Three Philosophical Perspectives on the Relation Between Technology and Society, and How They Affect the Current Debate About Artificial Intelligence', *Human Affairs* 30(4): 499–511. DOI: 10.1515/humaff-2020-0042
33. Rogobete, S. E. 2015. 'The Self, Technology and the Order of Things: In Dialogue with Heidegger, Ellul, Foucault and Taylor', *Procedia – Social and Behavioral Sciences* 183: 122–128. DOI: 10.1016/j.sbspro.2015.04.854
34. Russell, S.; Hauert, S.; Altman, R.; Veloso, M. 2015. 'Robotics: Ethics of Artificial Intelligence', *Nature* 521 (7553): 415–418. DOI: 10.1038/521415a
35. Schwitzgebel, E.; Garza, M. 2015. 'A Defense of the Rights of Artificial Intelligences', *Midwest Studies In Philosophy* 39(1): 98–119. DOI: 10.1111/misp.12032
36. Singh, A. 2023. 'Existential Perspectives on Artificial Intelligence in English Literature', *International Journal of Science and Research* 12(8): 323–328. DOI: 10.21275/SR23726104222
37. Stine, A. A.-K.; Kavak, H. 2023. 'Bias, Fairness, and Assurance in AI: Overview and Synthesis', in *AI Assurance*, eds. F. A. Batarseh et al. Academic Press, 125–151. DOI: 10.1016/B978-0-32-391919-7.00016-0
38. Zednik, C. 2021. 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence', *Philosophy & Technology* 34(2): 265–288. DOI: 10.1007/s13347-019-00382-7

SHERMAN XIE

Dirbtinis intelektas nepakeis žmonių, bet gali juos sunaikinti

Santrauka

Šiuo metu vykstančiose lenktynėse dėl bendrosios paskirties dirbtinio intelekto (DI) kūrimo vis labiau pastebima tendencija nepaisyti saugos ir etikos sumetimų. Ar besiplėtojantis bendrosios paskirties DI, ypač jei jis virsta nauja rūšimi, konkuruos su žmonėmis ir galiausiai juos pakeis? Šiame straipsnyje teigiama, kad dirbtinis intelektas yra tik nedidelė techninė sritis platesnėje automatizavimo technologijų sistemoje, todėl negali pakeisti žmonių. Tačiau dėl prigimtinio instrumentinio racionalumo be vertybiškumo dirbtinis intelektas gali būti netinkamai naudojamas ir gali sunaikinti žmoniją. Pabrėžiama, kad technologiniai tyrimai, įskaitant DI, dėl perdėto instrumentinio racionalumo, nepaisant vertybiškumo, galiausiai pasitarnaus kaip kuras dar vienai žmonijos nelaimei.

Raktažodžiai: dirbtinis intelektas, dirbtinės rūšys, mašinų etika, instrumentinis racionalumas, vertybiškumas