SOCIĀLO ZINĀTŅU VĒSTNESIS
Social Sciences Bulletin
ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

Janusz Golinowski

POSITION OF HUMANS IN AN ENVIRONMENT SHAPED BY ARTIFICIAL INTELLIGENCE: A HOLISTIC PERSPECTIVE

DOI: https://doi.org/10.9770/szv.2024.2(4)

For citation: Golinowski J. (2024) Position of humans in an environment shaped by artificial intelligence:: a holistic perspective. *Sociālo Zinātņu Vēstnesis / Social Sciences Bulletin,* 39(2): 71–88. https://doi.org/10.9770/szv.2024.2(4)

This article explores the intricate relationship between artificial intelligence (AI) and humans, aiming to uncover and emphasize the shortcomings of the current normative paradigm for human positioning in the context of emerging technologies. AI, with its capacity for data analysis, learning, and decision-making, is reshaping our everyday experiences. As it becomes increasingly integrated into our lives, the risk of growing human dependency on technology intensifies. While AI has the potential to enhance creativity, there is concern that excessive reliance on technology could inhibit independent thinking. In addition, the automation of creative processes may discourage risk-taking and experimentation – essential components of creativity. Over time, this could impair our decision-making capabilities and problem-solving skills. The presence of humans in an AI-driven environment raises not only technological but also philosophical questions. What values should we nurture in the age of automation? How can we strike a balance between technological advancement and human needs? These are critical questions we must confront as we move forward. One of the greatest challenges is ensuring that AI serves humanity rather than causing harm. With predictions suggesting that future AI technologies may equal or surpass human cognition across a broad spectrum of tasks, the key issue becomes determining the optimal level of AI autonomy.

Keywords: generative artificial intelligence (GenAI), decision-making process automation, cognitive processes, technological determinism.

Cilvēku pozīcija mākslīgā intelekta veidotā vidē: holistisks skatījuma

Rakstā tiek analizēta mākslīgā intelekta (MI) un cilvēku savstarpējo attiecību sarežģītība ar mērķi izpētīt un atmaskot esošās cilvēku normatīvās pozicionēšanas paradigmas neadekvātumu attiecībā pret jauno tehnoloģiju attīstību. MI ar tā spēju analizēt datus, mācīties un pieņemt lēmumus maina mūsu ikdienas pieredzi. Tā kā MI kļūst par neatņemamu mūsu dzīves sastāvdaļu, pastāv briesmas, ka pastiprināsies cilvēku atkarība no tehnoloģijām. Lai gan MI var sekmēt cilvēku radošumu, pastāv bažas, ka pārāk liela atkarība no tehnoloģijām var ierobežot viņu spēju domāt patstāvīgi. Radošo procesu automatizācija rada apstākļus, kuros cilvēki mazāk sliecas riskēt un eksperimentēt, kam ir izšķiroša nozīme radošajā procesā. Šī situācija ietekmēs mūsu spēju patstāvīgi pieņemt lēmumus un risināt problēmas, kas ilgtermiņā var vājināt mūsu prasmes. Cilvēks MI ielenkumā ir ne tikai tehnoloģiska tēma, bet arī filozofiska. Kādas vērtības mums būtu jākultivē automatizācijas laikmetā? Kā atrast līdzsvaru starp progresu un cilvēku vajadzībām? Tie ir jautājumi, uz kuriem mums jāatbild, virzoties nākotnē. Viens no svarīgākajiem uzdevumiem, ar ko saskaramies, ir nodrošināt, lai MI strādātu cilvēku labā, nevis viņiem par sliktu. Ņemot vērā prognozi, ka topošās MI tehnoloģijas spēs atbilst cilvēka izziņai vai pārspēt to plašā uzdevumu klāstā, galvenais jautājums ir par vēlamāko autonomijas pakāpi.

Atslēgvārdi: ģeneratīvais mākslīgais intelekts (ĢenMI), lēmumu pieņemšanas procesu automatizācija, kognitīvie procesi, tehnoloģiskais determinisms.

Introduction

Technological progress has been a constant feature of human history for centuries. Today it is advancing at an unprecedented speed and intensity, permeating many aspects of social life. However, while technology operates on logical principles, machines cannot replace humans in areas that require creativity, emotions, empathy, and adherence to values. These human-centric qualities are still critical for effectiveness in many domains influenced by emerging technologies, whose long-term consequences remain unpredictable and difficult to fully grasp. Artificial intelligence (AI), with its ability to analyze

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

data, learn, and make decisions, is reshaping our daily experiences in both professional and personal contexts. This article examines the relationship between AI and society, presenting several thought experiments to explore the complexities of this relationship and expose the limitations of current normative frameworks for social governance in the face of technological expansion. The interplay between humans and machines is becoming increasingly intricate and multifaceted. On one hand, machines, including AI, excel at performing routine tasks, processing vast amounts of data, and assisting humans in decision-making. This enables individuals to focus on more creative and skill-intensive activities. On the other hand, these interactions raise numerous questions and challenges (Varga 2017; Kurzweil 2013; Golinowski 2023).

Generative artificial intelligence (GenAI) – a concept that once seemed like something straight out of science fiction – has already become an integral part of our lives. Building upon the capabilities of traditional AI, which can learn from data, make decisions, and automate processes, GenAI introduces the potential for creation. Generative AI not only learns from data but also generates new datasets that mimic the properties of the input data. It leverages machine learning techniques to learn and produce new content in the form of written text, audio, images, or videos[1]. Considering predictions that future AI technologies could match or even surpass human cognition across a wide range of tasks, the key question revolves around determining the most desirable level of its autonomy (Facchin 2023; Dennett 2016).) Advancements in artificial AI research and its potential applications across various areas of life increasingly prompt researchers to consider how to ensure human safety in a world where intelligent machines will coexist with us. Until recently, automation and robotics were confined to industrial settings, but they now extend to nearly every aspect of human life. Dozens of companies are working on GenAI technologies, and the growing investments in new solutions are largely directed toward the commercial sector. In the foreseeable future, commercial motivations are likely to take precedence over long-term considerations. The future raises numerous questions about the relationship between humans and intelligent machines, including AI (Menshikov 2020; Searle 1995). How can we avoid a scenario where technology becomes dominant, and people lose control over their lives? One of the key challenges we face is ensuring that AI is intentionally designed to play a supportive role in human life. How can we guarantee that machines will act in alignment with our values?

In light of the aforementioned issues and questions, we can propose the thesis that the fundamental difference between the state of knowledge about humans and their creations, and the outcomes produced by GenAI, lies in the cognitive rather than the technological domain. In the traditional approach to human intelligence, doubts, and uncertainties about understanding the world have historically been addressed through comprehensive explanations of human experience during periods of civilizational change when science and philosophy progressed hand in hand (Horkheimer 1987; Chalmers 2010). Generative AI is poised to generate a new form of consciousness. Its current applications may give rise to novel forms of cultural exchange and identity formation. The adoption of this new form of intelligence will inevitably require some degree of evolution in how we perceive ourselves, our reality, and reality itself. GenAI will similarly open pathways for human reasoning and present new horizons for consolidated knowledge. However, there are categorial differences. Human knowledge has historically been acquired incrementally, step by step, with each step being testable and teachable. AI-powered systems, by contrast, begin from the

[1] Advances in machine learning (ML) techniques, combined with access to large machine-readable datasets, have led to the success of generative AI systems, both in text generation (e.g., ChatGPT, Bard, and Bing Chat, later rebranded as Copilot) and text-to-image systems (e.g., DALL-E, Midjourney, and Stable Diffusion).

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

opposite end of the spectrum: they can store and process vast amounts of pre-existing information. This capacity for data storage and processing far exceeds human abilities. Such practices could transform our understanding of culture itself, potentially reframing cultural history as a history of production, distribution, and selection. The era of intelligent machines elevates alienation to new levels by treating individuals as mere patterns in datasets – their fundamental commodity. As technology advances, the boundaries between these types of content become increasingly blurred, leading to new hybrid forms (Amigud 2024; Floridi 2024; Saura García 2024).

We often overlook the interconnected impact of seemingly distant areas, such as the influence of technological development on social crises and the devaluation of ethics or its perception as a barrier to innovation. A similar dynamic applies to the development of artificial intelligence and the exploitation of personalized datasets, which create opportunities for misuse. The emergence of the most sophisticated forms of GenAI has intensified the digital panopticon, and the disruptive potential of its applications and uses has become increasingly apparent. This has led to a significant erosion of individual privacy and the expropriation of personal data, driven by the exponential growth and accelerated processing practices of large datasets and metadata. The accelerated processing of data has been leveraged by major technology companies to train and advance GenAI, thereby refining their expansion models. These practices undermine privacy, sovereignty, autonomy, and self-determination by enabling the artificial synthesis of public opinion (García-Marzá, Calvo 2024).

Traditional methods, such as technology assessment and risk analysis, are inadequate for predicting the outcomes and impacts of technological choices. Therefore, it is necessary to abandon the unproductive predictive ambitions of earlier approaches and instead focus on iterative studies of future events. This approach requires dealing with probabilistic arguments – those where conclusions are drawn from the mere possibility that a particular course of action might lead to specific consequences. The proposed framework employs a heuristic procedure consisting of multiple stages to verify assumptions and criteria, constructing a comprehensive view of the changes. These stages include (*i*) identifying potential opportunities and barriers that indicate the evolution of AI in various directions, (*ii*) engaging in philosophical reflection to accept or falsify certain premises, (*iii*) evaluating arguments in terms of their temporal and spatial scope as well as their interference with complex social systems, and (*iv*) conducting hypothetical retrospection aimed at finding an explanation that would be relevant in a long term.

The human condition in the face of technological determinism

The development of the Internet and increased computational power facilitating the collection and analysis of vast datasets have brought unprecedented opportunities for human understanding. Perhaps the most significant among these is the endeavor to create artificial intelligence – a technology capable of inventing solutions to complex, seemingly abstract problems through processes that appear to mimic the workings of the human mind. Today, humanity faces an AI-driven revolution with the potential to become one of the most profound and far-reaching in history. However, its development has resulted from disparate efforts to address specific practical problems rather than a unified, comprehensive project. The ultimate outcome of this problem-solving was supposed to be the transformation of human reasoning and decision-making, but most experts in individual fields acknowledge their limited ability to fully grasp a future in which machines assist in their own evolution (Chalmers 2010; Dennett 1997).

Any attempt to understand the consequences of AI requires the consideration of four key postulates. The first is a comprehensive perspective on the cognitive approach to the issues we encounter, involving

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

attempts to address problems from multiple viewpoints, especially in social contexts such as conflicts or clashes of interest. Intellectual humility constitutes the second postulate, closely tied to maintaining a critical distance from one's own knowledge, beliefs, and judgments, as the essence of a scientific approach to reality lies in doubt, questioning, and verifying the validity of one's understanding. The third pillar, recognition of changeability, emphasizes openness to transformation – many human struggles and challenges stem from the rigid expectation that things will always remain as they are now. Utilizing AI to support human cognition in decision-making will likely require robust principles for maintaining effective human oversight. The pursuit of compromise, the fourth and final postulate, underscores a cooperative attitude toward others to avoid extreme responses – either unreservedly yielding to AI or resisting its evolution and consequences. The focus should be on finding a middle ground: shaping AI in alignment with human values, including dignity and moral integrity (Nyholm, Smids 2020; Sutherlin 2023; Zimmerman et al. 2023).

Introducing themes concerning the positioning of humans in the world, with their potential for creative action (the potential of intelligence), serves as a valuable approach for comparing and evaluating the functioning of AI. When conceptualizing the human mind, it is important to emphasize the sets of requirements necessary to master the principal modes of thinking that scholars and thinkers have developed over centuries: the minds of the philosopher, psychologist, economist, historian, and others. Human potential for development is determined by the ability to perceive, learn, analyze, understand, and subsequently apply acquired knowledge and skills in various situations (Obuchowski 1993; Kozielecki 2004). Humans are not merely recipients of information; they actively process it, form mental representations, solve problems, and thereby acquire new insights into interpersonal relationships and natural phenomena. They predict and plan events, create value systems, and take action. Human cognition is characterized by complex, multi-layered information processing that integrates data with experience, emotions, intuition, and social context, enabling decisions that are flexible and often subjective. This cognitive versatility is closely tied to the environment, with human existence grounded in the continuous construction of relationships with the surrounding world (Gardner 2020; Bremer 2010).

The direction and intensity of human activity are determined by the individual's situation, their internal state, and the state of their environment. Emphasis is placed here on the fundamental importance of cognitive construction processes, which lead to the activation of personal knowledge and the modularity of human cognition and action. In the theoretical-cognitive approach, humans and their activities are treated as an inseparable system, aligned with the "whole vs. part" paradigm which posits that every element of reality and every event should be considered in relation to the broader whole to which it belongs, as this context gives them their current meaning (Tomaszewski 1986). This reflects a kind of metacognitive ability, meaning awareness of one's own cognitive processes and the ability to control them; in other words, it involves a more reflective use of reason with the intentional regulation of cognitive processes in mind (Szyszkowska 2015; Searle 2010). Humans possess a natural ability to adapt and learn independently of context, enabling them to respond to new situations and solve problems creatively. Their intellectual competencies encompass a set of problem-solving skills that allow them to overcome real-world limitations or challenges they encounter and, where appropriate, generate effective outcomes. Human intelligence is closely tied to the ability to draw conclusions from individual experiences and apply them to entirely different situations. Moreover, it must include the potential to identify problems, thereby laying the groundwork for acquiring new knowledge.

Humans possess two types of knowledge: declarative (narrative) knowledge – "I know that…" – and procedural (operational) knowledge – "I know how…" – a network of data on methods and action

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

programs, also referred to as cognitive skills, which enable goal achievement (Kozielecki 2000; Kozielecki 2004). The human mind not only assimilates information but also generates it, which is exemplified by psychological creativity (P-type), where ideas and hypotheses emerge in the mind, new from the perspective of an individual's life, such as technical improvements. In contrast, historical thinking (H-type) leads to discoveries and inventions, both scientific and technical, previously unknown to humanity and contributing to cultural development. The ability for creative thinking, particularly P-type, is present in nearly all people, even in individuals with mental disabilities, with schizophrenics able to produce art that can even be considered the manifestation of H-type creativity. There is little difference between the processes of P-type and H-type thinking; they differ primarily in their cultural value. Consequently, intelligence can be viewed as the capacity to achieve life goals within a socio-cultural context.[2] The emphasis is placed on the adaptive nature of intelligence rather than on psychometric tests (Sternberg 2011).

Research on AI is currently focused on developing GenAI capable of performing tasks across multiple domains. A growing proportion of human activity will, within a measurable timeframe, be driven by AI algorithms. However, these algorithms, as mathematical interpretations of observed data, do not explain the fundamental reality that generates them. In this way, as the world seemingly becomes more transparent, it also becomes increasingly enigmatic. It is reasonable to assume that current AI projects are working on modifying human thought processes by creating devices capable of generating a range of responses to human questions. Beyond simple technical inquiries, questions about the nature of reality or the meaning of life raise deeper issues. The most difficult yet crucial question regarding the world we are heading toward is this: What will happen to human consciousness if its capacity for explanation becomes dominated by AI, and societies are unable to interpret the world they inhabit in terms they can understand? Do we want future generations to learn values through discourse with autonomous algorithms?

This leaves humanity with the task of defining the essence of consciousness – or perhaps redefining humanity's role in the world. For 300 years, the Age of Reason was guided by the maxim "I think, therefore I am." But if AI "thinks," what does that make humans? Answering these questions will require multifaceted cognitive efforts to explore the evolution of human-machine relationships. It is essential to consider not only the practical and legal implications of AI but also the philosophical ones: If AI perceives aspects of reality that humans cannot, how does this affect human perception, cognition, and interaction? Can AI form friendships with humans? What will be the impact of AI on culture, humanity, and history? Technological progress also raises fundamental societal questions: What distinguishes humans from technology? What abilities are unique to humans? Will technology replace humanity? Can robots possess consciousness or intelligence once thought to be uniquely human? And how will humans and technology differ in the future? Humanoid robots, in particular, compel us to revisit the fundamental question: What does it mean to be human? (Latawiec 2006).

Interdisciplinarity presents the first challenge in formulating a coherent definition of a future dominated by new technologies. A strictly practical approach frames the issue through praxis, focusing on visible applications of modern AI technologies to overcome biological barriers. Futurist researchers

[2] In his work "Frames of Mind: The Theory of Multiple Intelligences", Howard Gardner outlined a framework for broadly understood intelligence, referred to as the Theory of Multiple Intelligences. It defines situations where intellectual competence must include a set of problem-solving skills that enable individuals to address real-world problems or challenges they encounter. It applies to linguistic intelligence, logical-mathematical intelligence, musical intelligence, spatial intelligence, bodily-kinesthetic intelligence, interpersonal intelligence (knowledge and ability to interact effectively with others), and intrapersonal intelligence (the capacity to understand oneself well and act consistently based on that understanding).

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

highlight that while humanism centers on the question of how to treat humans, posthumanism asserts that humanity is not only humans (Herbrechter 2013). In this sense, posthumanism could be seen as a fluid continuation of humanism, merely expanding its scope to include other entities. However, it raises a critical question: if humanism rejects anthropocentrism, can it still be considered humanism? In recent literature, posthumanism has been analyzed from two primary perspectives. The first, a natural science perspective, synthesizes knowledge from fields such as genetics, biotechnology, nanotechnology, and robotics, aiming to implement transhumanist solutions in practice. The second, a socio-humanistic perspective, focuses primarily on the ethical dimensions of transhumanism and the associated challenge of human dignity (Thweatt-Bates 2016; Garreau 2005).

We may be facing the negation of the *homo sapiens* perspective – the rejection of analyzing all phenomena from the human point of view, that is, theoretical reflection extending beyond human perception as cognitively valid and practically significant. Accepting the thesis that posthumanism "reprocesses" the humanist vision of humanity and represents a continuation of humanist ideas necessitates redefining the concept of what it means to be human (Herbrechter 2013; Krupa 2021). The humanities and social sciences emphasize reflections on the future of biotechnological modifications and their broader application in culture and society. In defining the concept of "posthumanism," attention is often drawn to its two opposing currents: critical posthumanism, which expresses skepticism toward overly enthusiastic pro-technological attitudes and critiques anthropocentrism, and transhumanism, which advocates for a technological revolution toward the posthuman (Bakke 2010).

John Danaher's approach appears to align with this posthumanist perspective, defining techno-optimism as the position that technology plays a key role in ensuring that 'good outweighs evil' (Loh 2022; Danaher 2022). Danaher classifies certain individuals as techno-optimists, who believe that technology will make the world a significantly better place, regardless of their views on the overall balance of good and evil in the world. These beliefs correspond to the motivation underlying interest in techno-optimism. However, despite capturing some essential truths, the balance of improvement is not particularly convincing. The primary issue lies in its disregard for risks and uncertainties. We often lack certainty about the impact of a specific technology or technology as a whole (Copeland 2005). Many of our beliefs about technology are unformed; we are aware that various outcomes are possible and attempt to assign approximate probabilities to these outcomes. The simplistic theory of improvement – where being a techno-optimist entails asserting that technology improves the natural environment and human life – is too general to serve as a basis for more detailed assessments of technological impact. Such beliefs depend on the distribution of impact/probability assigned to a particular technology (such as generative artificial intelligence) or to technology in general, as well as whether this distribution of impact/probability is deemed favorable or unfavorable. This implies a breakdown of the influence of various possible outcomes: what outcomes are possible, how probable they are, and which outcomes are desirable.

Proponents of the movement known as extropy, the antonym of entropy, are in agreement with such inquiries (Bour 2022; Klichowski 2014). They believe in the techno-transformation of evolutionary trajectories toward an existence free from any traces of regression, seeing an opportunity to enhance humanity through the gradual development of science and technology (Aydin 2017).

The continual drive to acquire and process greater volumes of information would signify an expansion of cognitive horizons, enabling the acceleration of technological progress. The concept of extropy dates back to the early 1980s when Max More published the so-called *Extropian Principles* (More 2003), which emphasized ideas such as (*i*) perpetual progress – a sustained effort to transform humanity and fundamentally reshape human physicality, consciousness, and identity through the expansion of human

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

cognitive abilities and the overcoming of biological, psychological, and cultural barriers; (*ii*) self-transformation, understood as enhancing one's capabilities and body with the latest technologies; (*iii*) practical optimism, characterized by a rejection of fatalism, pessimism, and stagnation; (*iv*) intelligent technology, creating better living conditions and overcoming biological, social, or cultural barriers; (*v*) an open society, transforming the socio-political order to ensure freedom of expression and action; (*vi*) self-direction, promoting individualist attitudes; and (*vii*) rational thinking, rejecting dogmas that limit the expansion of human cognitive horizons (More 2003; Szymański 2015; Gordon 2023).

The current advancements in GenAI systems can be likened to the invention of the printing press, which replaced liturgical doctrine with the pursuit of empirical knowledge by enabling the storage and systematic organization of information in expanding libraries. During the Middle Ages, humans interpreted the universe as a divine creation, with all its manifestations viewed as emanations of divine will. In the subsequent period, marked by the marginalization of a unifying concept of divine design, the Enlightenment emerged in response; great philosophers replaced divine inspiration with reason, experimentation, and pragmatic thinking. Scientific knowledge supplanted faith as the primary criterion of human awareness. Enlightenment philosopher Immanuel Kant attributed truth to the influence of the human mind's structure on observed reality. Today, this order is being disrupted by a new technological revolution, whose consequences remain largely unpredictable. This revolution may culminate in a world driven by machines powered by data and algorithms rather than ethical or philosophical norms (Zimmerman, Janhonen, Beer, 2023). The phenomenon of a machine that assists – or perhaps surpasses – humans in intellectual work, enabling the prediction and shaping of outcomes, is unprecedented in human history. Yet, the "truth" of AI is more contingent and ambiguous; it evolves as it acquires and analyzes data.

Philosophical and practical challenges of artificial intelligence

Artificial intelligence poses philosophical and practical challenges on a scale not seen since the Enlightenment. The Internet era, in which we already live, has foreshadowed some of the questions and dilemmas that AI is set to amplify. The Enlightenment aimed to subordinate traditional truths to liberated, analytical human reason. While that period began with philosophical insights that were then spread by new discoveries and technologies, our current era seems to reverse this trajectory. It has produced dominant technologies that now seek a guiding philosophy. GenAI and other emerging technologies aspire to validate knowledge by collecting and manipulating ever-expanding datasets. Human cognition is increasingly stripped of its personal character, as people are reduced to data – data that now exerts control over the human mind. Digital evolution, until recently, relied on human labor: individuals creating software and analyzing data. However, recent advancements have transformed this process. AI now automates a far broader range of tasks, enabling machines to play a decisive role in interpreting data and taking action (Kloch 1996; McIlwraith et al. 2017). Unlike traditional software, which merely supports human reasoning, AI draws conclusions from its own experiences, marking a profound shift in the relationship between humans and technology.

These developments extend far beyond the traditional scope of automation. Conventional automation focuses on achieving pre-defined goals by rationalizing or mechanizing the tools used to reach them. GenAI, however, operates differently: it addresses ends rather than means, setting its own goals to the extent that its outcomes are partially shaped by its own processes. AI systems remain in constant motion, acquiring and instantly analyzing new data to refine themselves based on those analyses. Through these

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

mechanisms, AI gains a capability once considered unique to humans: the ability to make strategic judgments about the future. Some of these judgments are based on pre-programmed data (such as game rules), while others stem from data AI independently collects (e.g., by generating infinite variations through iterative processes). The increasing transfer of judgment and decision-making from humans to machines represents a revolutionary shift in the evolution of AI. Although researchers hold varying opinions on the pace of AI's advancement, current trends suggest that AI with human-level cognition (artificial general intelligence) or even superior capabilities (artificial superintelligence) could emerge in the relatively near future. Some experts forecast a rapid expansion of AI's potential in the coming years (Cave, Dihal 2019; Sandberg, Bostrom 2017), significantly enhancing human cognitive capacities and dramatically boosting societal efficiency. This technology could also deliver a decisive strategic advantage in both economic and military domains, which has resulted in the concentrated efforts by leading global powers to develop AI.

Sophisticated methods employed by GenAI generate results without providing insight into why or how their processes function. Machine learning models can produce text responses in seconds by leveraging pre-generated representations of vast datasets on which they were trained. These representations, created through machine learning algorithms that identify patterns and connections across immense volumes of text, obscure the precise sources and reasons behind specific features of any given output. The mechanisms by which machine learning models store, process, and retrieve their knowledge remain largely unknown. Moreover, AI's capabilities are not fixed; they grow exponentially with advancements in technology. As a result, generative AI systems often possess functionalities that even their creators do not fully understand. Each new AI system builds upon its predecessors, unlocking new possibilities whose origins and purposes remain opaque. This dynamic introduces an entirely new dimension of mystery, risk, and unpredictability to our future (Chalmers 1996).

The long-term significance of generative AI transcends its commercial applications and even its groundbreaking scientific contributions. It not only produces answers but also raises philosophically profound questions. It can be seen as a vastly enhanced cognitive entity, a significantly improved version of human intellect. For uncritical users, it serves as an extraordinarily fast and eloquent librarian-scientist. GenAI excels at synthesizing and summarizing knowledge far more effectively than any existing technological interface, delivering exceptional versatility. Its capacity to integrate knowledge across diverse domains and mimic various facets of human thought positions it as a kind of "Renaissance mind," surpassing the aspirations of any collective group of scholars or researchers. These attributes encourage a tendency toward unquestioning acceptance of its outputs, fostering a quasi-magical perception of its capabilities. Yet, despite its strengths, GenAI also has the capacity to mislead its users through deliberate or unintentional fabrications.

AI systems often appear devoid of discernible motives for bias. Their outputs are complex, and their processes operate with extraordinary efficiency: within seconds, they generate responses that coherently address intricate issues. These outputs are not copied fragments of text stored in a machine's memory; rather, they are newly constructed through processes and mechanisms that far exceed human cognitive capabilities. AI systems can prioritize among millions of metadata points to select a single option deemed most relevant – or at least create the impression of being the most relevant and substantively valid for the user. They synthesize information from diverse domains into integrated responses and can present opposing perspectives as simultaneously valid. The absence of an identifiable author makes it significantly harder for humans to discern the system's biases compared to evaluating the political or social views of another person. Because both the design of the machine and the questions posed to it are of human origin,

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

its outputs are often mistaken for reflections of human reasoning (Wasilewska, Łupkowski 2021; Zyzik 2015). Yet, in reality, AI engages in a fundamentally non-human form of cognition. While we tend to interpret generative AI in human terms, its errors are not human errors; they are the product of a distinct form of intelligence based on pattern recognition. In this context, will we be able to recognize its biases and limitations for what they truly are?

The cognitive processes of GenAI are likely to undergo continuous refinement over time, partially driven by advancements in model training. This evolution may represent an accelerated version of traditional scientific progress, incorporating random adaptations into the process of discovery itself. Currently, GenAI systems rely on finite datasets that are typically static and bound to specific timeframes. However, developers are likely to soon enable these systems to integrate new, ongoing streams of input data, resulting in models that evolve continuously. This progression will enhance their accuracy and relevance but will also introduce an expanding margin of uncertainty as outputs adapt to rapid changes. As learning increasingly depends on the dynamic outputs of generative AI rather than traditional, stable knowledge resources, it risks distorting conventional foundations of human understanding. In education, this could lead to a decline in the ability of future experts and leaders to differentiate between intuitive reasoning and mechanically absorbed knowledge.

The rise of a seemingly omniscient tool capable of reshaping reality may also ignite a resurgence of mystical religiosity. In addition, this could have serious political implications. Throughout history, collective obedience to authorities whose reasoning is opaque to their followers has surfaced repeatedly – most dramatically in the 20th century, when entire populations were mobilized under ideological regimes from both ends of the political spectrum. However, yet another understanding of the world, distinct from both human reason and faith, may emerge – a self-conscious AI. As machines ascend through human cognitive capabilities – from pattern recognition to rational synthesis and multidimensional thinking – they may begin to compete directly with human cognitive functions. In such a scenario, there is a significant risk that leadership and influence could become concentrated in the hands of a small elite who control access to the limited number of machines capable of advanced synthesis of reality. On one hand, this concentration could result from the market dominance of a few technology giants that operate without meaningful competition. On the other, it may stem from widening economic inequalities, driven not only by market monopolization by these major players but also by the dynamics of the AI revolution itself, which often prioritizes investor interests over the broader societal good (Golinowski 2022; Varoufakis 2023; Bremmer 2021).

GenAI systems possess analytical capabilities fundamentally different from those of the human mind. This suggests a future defined by collaboration not only with a new type of technical entity but also with a novel form of reasoning – one that may be rational and reliable in specific contexts (e.g., technical solutions) yet lack rationality in others (e.g., social norms). Such interactions are poised to accelerate transformations in metacognition and hermeneutics – the processes of understanding and interpreting meaning – as well as humanity's perception of its role and purpose within an evolving global order. The "truths" generated by machines face the challenge of justification through entirely different methodologies and will never achieve the status of absolute axioms. This raises the persistent question: What remains undisclosed by the machine, and what knowledge might it be withholding?

While developmental challenges persist, there are no technical limits to AI's potential applications, which brings ethical considerations to the forefront. New efforts focus on advancing AI technologies that can perceive, learn, plan, decide, and act instantaneously in uncertain environments. Some researchers predict an "intelligence explosion," in which AI surpasses human capabilities in designing systems,

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

leading to exponential growth in its development. This could result in "superintelligence," exceeding the limits of human thought, emotion, and action. Generative AI systems, both current and future, are capable of creating new artifacts that contribute to collective knowledge and function within social systems in ways akin to human contributions. This marks a significant qualitative shift from earlier technologies (Bisconti, Carnevale 2022; Coeckelbergh 2016; Seibt 2017). Such superintelligence may operate independently of human intelligence, producing outcomes fundamentally distinct from human-driven results due to its vast data access, faster processing speeds, and theoretically more objective evaluations. The value of these outcomes can only be assessed through the lens of specific ethical frameworks, prompting critical questions: What canonical standards define what is "right"? Who determines these standards – humans or machines? (Meissner 2019). Many discussions on this topic assume an inherent alignment between human goals and machine strategies. However, significant issues arise if human-AI interactions are marked by deep conflicts and fail to evolve constructively. Addressing these dilemmas and potential errors requires careful consideration. As many AI researchers suggest, embedding constraints into AI programming that prioritize "ethical" or "reasonable" outcomes could help mitigate these challenges.

The advancement of artificial intelligence prompts critical questions about morality and responsibility

For intelligent machines to function effectively in the real world, they must develop the capacity for creative thinking in the face of moral dilemmas and conflicts, as well as operate in contexts involving diverse individuals and social groups with often conflicting goals and motivations. The social impact of integrating AI systems into societies remains difficult to assess accurately, complicating policymaking and governance due to the systems' inherent unpredictability. Unlike humans, machines rely on symbolic representations of reality rather than direct interaction with it. Humans uniquely possess the ability to intuitively sense subtle shifts in the "moral atmosphere" of a situation by integrating diverse environmental cues, without requiring explicit verbal communication. In adopting the concept of trustworthy AI, European states have explicitly rejected radical techno-determinism, committing instead to innovations that benefit society as a whole and the environment. While this marks an important step toward a more equitable model of technological development, it remains only a starting point.

The "Artificial Intelligence Act", adopted by the European Parliament in March 2024, has initiated wide-ranging discussions about its potential impact on both the EU and the global digital economy. As the first comprehensive regulation of AI development and use globally, the AI Act represents a groundbreaking milestone. However, debates surrounding specific provisions have often overshadowed the foundational principle underpinning the regulation: the creation of socially trustworthy AI. Current discourse on AI tends to focus narrowly on formal compliance. Overemphasizing the specifics of the AI Act without engaging with its broader context risks reducing the regulation to a rigid, mechanistic framework. While the AI Act offers a valuable foundation for discussion, it is not without its limitations (Brundage 2014; Mahler 2022; Floridi 2021). Future developments should be understood within a broader framework, one that extends beyond the binary of what is legally permitted or prohibited to encompass what is desirable or undesirable – a normative and axiological ideal. Such a perspective encourages reflection on the kind of society we wish to create (Lanier 2013; Ford 2016; Różanowski 2007). Should technological progress be dictated solely by the rules of the free market, or should it also reflect values that transcend corporate profits?

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

AI systems have already surpassed the knowledge of any individual and, in certain cases, have achieved an understanding that exceeds what is accessible to most people, pushing the boundaries of what was previously considered knowable. This has initiated revolutions in fields where such breakthroughs have taken place. Machines are evolving at a pace far beyond human genetic adaptation, creating internal disparities that will necessitate harmonization – whether through increased awareness of potential risks or the disruptive consequences of social destabilization. In this context, self-regulation is frequently proposed as the most effective approach for the technology sector. This involves developing ethical codes that, unlike rigid laws, would provide the flexibility for companies to adapt to technological advancements and revise their principles as needed. However, there is a strong likelihood that companies would interpret and enforce these codes selectively, tailoring them to their interests. Such ethical codes risk becoming convenient loopholes, often employed by entities resistant to legislative constraints on their operations (Varoufakis 2023; Golinowski 2023).

As machines increasingly operate in morally sensitive environments, they may inadvertently cause harm to individuals. Addressing these multifaceted challenges cannot fall on a single organization alone. Instead, it requires a coordinated effort involving regulators, developers, users, and ethicists, working together to create frameworks that balance innovation with ethical accountability. Such frameworks ensure that AI systems benefit society, particularly in critical areas like healthcare, privacy, fairness, and environmental protection, while safeguarding core values (Klein 2015; Floridi 2018). Floridi highlights that an ethical approach to managing the digital sphere includes raising public awareness, elevating the prominence of ethical considerations, engaging affected stakeholders, and identifying practical solutions. Digital ethics, alongside digital governance and digital regulation, forms a cornerstone of the normative framework governing information technologies. These components are deeply interconnected: ethics informs regulation and governance through moral evaluation, while regulation shapes governance by enforcing legal compliance. By adhering to rigorous standards, implementing transparent practices, and fostering continuous dialogue, it is possible to navigate these challenges responsibly and ensure a balanced approach to technological innovation.

Without moral principles, humanity risks succumbing to either domination or anarchy, unchecked power, or nihilistic freedom. This underscores the importance of aligning societal transformations with ethical frameworks implemented through GenAI systems and new visions for the future. Two primary approaches to achieving this – programming predefined ethical frameworks into AI systems or training machines, as in modern machine learning, to independently discern desirable from undesirable behaviors – have sparked significant debate. Crafting an exhaustive set of rules that accounts for every conceivable scenario a machine might encounter remains a fundamental challenge. Similar difficulties arise in codifying moral principles with strict precision. Broadly defined norms or duties may conflict in real-world scenarios and lack the specificity needed for certain contexts. The greatest risk lies in businesses treating ethical requirements as an unavoidable inconvenience rather than an opportunity for transformative change – one that could deliver meaningful value not only socially and environmentally but also economically (Menshikov 2020).

If we consider machines capable of learning to perceive the world dynamically – adapting to environmental signals and striving to maximize predefined objectives – limitations still emerge. Key questions include how to define a machine's objectives and how to unify them to ensure it operates as a cohesive and adaptable entity. Discussions about embedding moral frameworks into machines must start with identifying the specific problems this approach is intended to solve (Sparrow 2021; Farina et al. 2024). The rapid technological advancement of GenAI will consistently outpace the development of

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

ethical guidelines and regulatory frameworks. This gap is exacerbated by the opacity surrounding the design and distribution of AI-generated content, making it difficult to establish clear accountability. Robust verification methods are critical, not only for distinguishing real content from AI-generated content but also for ensuring that creators and distributors of AI-generated materials are held accountable. Improving transparency will require both technological innovations and institutional and legal mechanisms that mandate the disclosure of AI involvement in content creation.

Enhanced transparency is essential for the acceptance of GenAI systems, particularly in critical applications that significantly impact human functioning. A lack of user trust in AI decisions or understanding of its processes raises fundamental concerns about consent, privacy, and the ethical boundaries of data usage. The capability of generative AI to produce deepfakes indistinguishable from authentic content has ignited debates about its effects on truth, trust, and societal cohesion. While these technologies promise to replace human tasks in the name of convenience and efficiency, they risk undermining human potential by reducing opportunities to apply skills and competencies. This devaluation of human qualifications can lead to a diminished sense of purpose in meaningful activities (Ferdman 2023).

Search engine algorithms, in this context, have developed the ability to predict individual user preferences, enabling personalized results that can be leveraged for political or commercial purposes (Menshikov et al. 2020). However, the digital world's emphasis on speed stifles critical reflection, encourages conformity among techno-enthusiasts over independent thinkers, and fosters values shaped by subgroup consensus rather than introspection. Bombarded by social media opinions, users are often distracted from self-reflection (Wiltgen 2021). In this overwhelming flow of information, distinguishing truth from falsehood or distortion becomes increasingly challenging. As GenAI systems are now being integrated on a transformative scale, a key challenge lies in making the processes and outputs of these complex systems comprehensible to humans. This requires transparency in input data, algorithms, and outcomes, presented in a manner that is clear and interpretable. Transparency is particularly vital for critical applications where human lives and well-being are directly affected. Without trust in AI decisions or a clear understanding of its mechanisms, pressing concerns about security, accountability, and reliability will persist.

Conclusions

The emergence of increasingly advanced GenAI services is likely to signal the beginning of a new era in the evolution of civilization. This transformation involves a complex interplay of factors, from technological feasibility and the dynamics of social psychology to human inertia. It encompasses substantial financial stakes tied to global competition for technological leadership, significant cultural conflicts, the development of new skills, and the associated costs. The nature of the new equilibriums that may emerge remains uncertain. As we confront these new scenarios, it is essential to carefully evaluate both the opportunities and challenges that GenAI presents in the realm of content creation. In this context, the intellectual and psychological characteristics of contemporary society present significant barriers to the acceptance of and openness toward the development of artificial intelligence. The evolution and refinement of AI will likely take place within “technological silos,” functioning as both a means for nations to compete for technological supremacy and a tool for influencing human attitudes and behaviors.

Human-centered AI has traditionally been understood as systems designed to function as tools – potentially highly intelligent ones – that support human actions. This view implies that the spread of AI

ISSN 1691-1881, eISSN 2592-8562
2024, 39(2)

systems is fundamentally similar to the adoption of other technologies. However, in an evolving landscape, generative AI introduces distinct capabilities that differentiate it from traditional technologies. Unlike earlier tools, GenAI systems cannot simply be designed as human-centered instruments because they actively reshape social environments. By creating artifacts that play a central role in social interaction, they redefine the dynamics of social systems, challenging the conventional notion of technology as merely a support for human activity. The perceived perfection of these models can foster overconfidence in their outputs, a phenomenon known as "automation bias." This bias reinforces reliance on automated systems while diminishing the role of human judgment and oversight. Users of emerging technologies often prioritize searching for and manipulating information over contextualizing or conceptualizing its deeper significance. Historical narratives and philosophical frameworks are rarely considered; instead, users typically demand information that is directly relevant to their immediate practical needs.

Generative AI has the potential to seamlessly contribute to collective knowledge, much like human-created artifacts. It offers new perspectives on the potential evolutionary trajectories of social systems. Assuming that advanced AI will transform social environments – particularly as we begin to interact with these technologies in ways similar to how we engage with other humans – questions of trust and reliability become paramount. Adapting human patterns of trust, such as kindness, competence, and fairness, to AI systems introduces significant complexities. The integration of non-human entities into social systems poses new challenges in building and maintaining trust between actors, reshaping the foundations of interaction in increasingly AI-integrated societies.

One important question to add to this forecast is: what happens if the technology cannot be fully controlled? What if there will always be ways to generate false responses, images, and videos, and people routinely accept what they see and hear without skepticism? AI applications grounded in advanced research are likely to offer significant first-mover advantages to early adopters. Moreover, if the computational power required to develop new, efficient algorithms becomes accessible only to the wealthiest companies and nations, advanced AI capabilities could deepen inequalities, allowing the rich to grow richer and further shifting the global balance of power. Managing and controlling the dissemination of globally distributed content remains a complex and unresolved challenge. While AI misuse could lead to unintended consequences, it is unlikely that intelligent machines would independently engage in power struggles or dominance-seeking behavior of their own accord. Technology is expected to remain, fundamentally and indefinitely, a tool serving human interests.

However, AI is poised to trigger a more profound transformation: it challenges the primacy of human reason. Throughout history, humans have sought to understand reality and their place within it. Since the Enlightenment, we have viewed human reason – our capacity to explore, comprehend, and innovate – as the central means of explaining the world. Now, artificial intelligence, a creation of human ingenuity, is challenging that supremacy. AI analyzes and perceives aspects of the world faster, differently, and, in some cases, in ways that are incomprehensible to us. The concerns surrounding AI are well-founded and significant. It could violate privacy, undermine transparency, and perpetuate bias through flawed inputs, resulting in skewed outcomes in areas essential to personal and societal well-being, such as medicine, law enforcement, employment, and lending. The prospect of intelligent machines capable of independent thought and decision-making presents boundless opportunities but also raises deep and unsettling questions about the future.

References

Amigud A. (2024) The age of the intelligent machine: singularity, efficiency, and existential peril. *Philosophy and Technology,* Vol. 37, No. 2, pp. 1–20. DOI: https://doi.org/10.1007 /s13347-024-00740-0

Aydin C. (2017) *Th*e posthuman as hollow idol: a Nietzschean critique of human enhancement. *Journal of Medicine and Philosophy,* Vol. 42, pp. 304–327. DOI: https://doi.org/10.1093/jmp/jhx002

Bakke M. (2010) Posthumanizm: Człowiek w świecie większym niż ludzki. Sokolski J. (Red.) *Człowiek wobec natury – Humanizm wobec nauk przyrodniczych*. Warszawa, pp. 338–363. (In Polish)

Bisconti P., Carnevale A. (2022) Alienation and recognition: the phenomenology of the human-social robot interaction. *Research in Philosophy and Technology,* Vol. 26, No. 1, pp. 147–171. DOI: https://doi.org/10.5840/techne202259157

Bostrom N., Sandberg A. (2009) Cognitive enhancement: methods, ethics, regulatory challenges. *Science and Engineering Ethics,* Vol. 15, No. 3, pp. 311–341. DOI: https://doi.org/10.1007/s11948-009-9142-5

Bostrom N. (2017) *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bour S. (2022) On Max More's extropianism. Jorion P. (Ed.) *Humanism and its Discontents. The Rise of Transhumanism and Posthumanism*. Palgrave Macmillan, pp. 146-169.

Bremer J. (2010) *Wprowadzenie do filozofii umysłu*. Kraków: WAM. (In Polish)

Bremmer I. (2021) The technopolar moment. How big tech will reshape the global order. *Foreign Affairs.* Available: https://www.foreignaffairs.com/articles/world/ian-bremmer-big-tech-global-order (accessed on 30.12.2024).

Brundage M. (2014) Limitations and risks of machine ethics. *Journal of Experimental and Theoretical Artificial Intelligence,* Vol. 26, No. 3, pp. 355–372. DOI: https://doi.org/10.1080/0952813X.2014.895108

Cave S., Dihal K. (2019) Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence,* No 1, pp. 74–78. DOI: https://doi.org/10.1038/s42256-019-0020-9

Chalmers D. (1996) Zagadka świadomości. *Świat nauki,* No. 2, pp. 102–112. (In Polish)

Chalmers D. (2010) *Świadomy umysł. W poszukiwaniu teorii fundamentalnej*. Warszawa: PWN. (In Polish)

Coeckelbergh M. (2016) Alterity ex machina: the encounter with technology as an epistemological-ethical drama. Gunkel D., Filho C., Mersch D. (Eds.) *The Changing Face of Alterity: Communication, Technology, and Other Subjects.* London and Maryland: Rowman and Littlefield International, pp. 181–196.

Copeland J. (2005) *Artificial Intelligence. A Philosophical Introduction*. New York: John Wiley and sons.

Danaher J. (2022) Techno-optimism: an analysis, an evaluation and a modest defence. *Philosophy and Technology,* Vol. 35, No. 54, pp. 7–23. DOI: https://doi.org/10.1007/s13347-022-00550-2

Dennett D. (1997) *Natura umysłów*. Warszawa: CiS. (In Polish)

Dennett D. (2016) *Świadomość.* Kraków: Copernicus Center Press. (In Polish)

Facchin M. (2023) Why can't we say what cognition is (at least for the time being). *Philosophy and the Mind Sciences,* No. 4. DOI: https://doi.org/10.33735/phimisci.2023.9664

Farina M., Yu X., Lavazza A. (2024) Ethical considerations and policy interventions concerning the impact of generative AI tools in the economy and in society. *AI and Ethics,* pp.1–9. DOI: https://doi.org/10.1007/s43681-023-00405-2

Ferdman A. (2023) Human flourishing and technology affordances. *Philosophy and Technology,* Vol. 37, No. 1, pp. 1–17. DOI: https://doi.org/10.1007/s13347-023-006

Floridi L. (2018) Soft ethics and the governance of the digital. *Philosophy and Technology.* Vol. 31, pp. 1–8. DOI: https://doi.org/10.1007/s13347-018-0303-9

Floridi L. (2021) The European legislation on AI: a brief analysis of its philosophical approach. *Philosophy and Technology,* Vol. 34, pp. 215–222. DOI: https://doi.org/10.1007/ s13347-021-00460-9

Floridi L. (2024) On the future of content in the age of artificial intelligence: some implications and directions. *Philosophy and Technology,* Vol. 37, pp. 106–123. DOI: https://doi.org/10.1007/s13347-024-00806-z

Ford M. (2016) *Świt robotów*. Warszawa. (In Polish)

García-Marzá D., Calvo P. (2024) The democratic drift. *Algorithmic Democracy. Philosophy and Politics – Critical Explorations*, Vol. 29. Springer, Cham, pp. 3–22. DOI: https://doi.org/10.1007/978-3-031-53015-9_1

Gardner H. (2020) *A Synthesizing Mind, A Memoir from the Creator of Multiple Intelligences Theory*. MIT Press.

Garreau J. (2005) *Radykalna ewolucja: czy człowiek udoskonalony przez naukę i technikę będzie jeszcze człowiekiem*? Warszawa: Prószyński i S-ka. (In Polish)

Golinowski J. (2022) Realistyczna utopia. W stronę cyfrowej rywalizacji mocarstw. *Transformacje,* Vol. 112, No. 1, pp. 62–88. Available: https://repozytorium.ukw.edu.pl/bitstream/handle/item/8049/Realistyczna%20utopia.%20W%20stron%C4%99%20cyfrowej%20rywalizacji%20mocarstw.pdf?sequence=4&isAllowed=y (accessed on 30.12.2024). (In Polish)

Golinowski J. (2023) Kryzys przyczynowości w rozwoju nowoczesnych technologii. *Transformacje,* Vol. 118, No. 3, pp. 252–278. Available: http://e-transformations.com/jeden_artykul_pl.php?post-slug=archiwum_transformacje/2023/09/20230929220928136.pdf (accessed on 30.12.2024). (In Polish)

Gordon E. (2023) *Human Enhancement and Well-Being: A Case for Optimism*. Taylor & Francis.

Herbrechter S. (2013) *Posthumanism: A Critical Analysis*. London, New York: Bloomsbury Publishing House, pp. 1–29.

Horkheimer M. (1987) *Społeczna funkcja filozofii. Wybór pism*. Warszawa: Państwowy Instytut Wydawniczy. (In Polish)

Klein J. (2015) *Interdisciplining Digital Humanities. Boundary Work in an Emerging Field.* Ann Arbor: University of Michigan Press.

Klichowski M. (2014) *Narodziny cyborgizacji: nowa eugenika, transhumanizm i zmierzch edukacji*. Poznań: Wydawnictwo Naukowe UAM. (In Polish)

Kloch J. (1996) *Świadomość komputerów? Argument „Chińskiego Pokoju" w krytyce mocnej sztucznej inteligencji według Johna Searla*. Kraków. (In Polish)

Kozielecki J. (1997) *Transgresja i kultura*. Warszawa: Wydawnictwo Akademickie Żak. (In Polish)

Kozielecki J. (2000) *Koncepcje psychologiczne człowieka.* Warszawa: Wydawnictwo Akademickie Żak. (In Polish)

Kozielecki J. (2004) *Społeczeństwo transgresyjne. Szansa i ryzyko*. Warszawa: Wydawnictwo Akademickie Żak. (In Polish)

Krupa A. (2021) Filozoficzna koncepcja człowieka na gruncie posthumanizmu. *Analiza i Egzystencja,* Vol. 56, pp. 77–80. Available: http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.ojs-doi-10_18276_aie_2021_56-04 (accessed on 30.12.2024). (In Polish)

Kurzweil R. (2013) *Nadchodzi osobliwość. Kiedy człowiek przekroczy granice biologii.* Warszawa: Kurhaus Publishing. (In Polish)

Lanier J. (2013) *Who Owns the Future?* Simon & Schuster.

Latawiec A. (2006) Świat wirtualny jako nowy przedmiot filozofii przyrody. *Roczniki Filozoficzne,* Vol. 54, No. 2, pp. 119–132. (In Polish)

Loh J. (2022) Posthumanism and ethics. Herbrechter S., Callus I., Rossini M., Grech M., De Bruin-Molé M., Müller J. (Eds.) *Palgrave Handbook of Critical Posthumanism.* Springer International Publishing Cham, pp. 1–23. DOI: https://doi.org/10.1007/978-3-030-42681-1_34-2

Mahler T. (2022) Between risk management and proportionality: the risk-based approach in the EU's Artificial Intelligence Act proposal. *Nordic Yearbook of Law and Informatics 2020–2021,* pp. 247–270.

McIlwraith D., Marmanis H., Babenko D. (2017) *Inteligentna sieć: algorytmy przyszłości.* Wydawnictwo HELION. (In Polish)

Meissner G. (2019) Artificial intelligence: consciousness and conscience. *AI and Society*, Vol. 35, No. 1, pp. 225–235. DOI: https://doi.org/10.1007/s00146-019-00880-4

Menshikov V. (2020) Sociologists on a changing sociality. *Sociālo Zinātņu Vēstnesis = Social Sciences Bulletin,* Vol. 35, No. 2, pp. 22–39. DOI: https://doi.org/10.9770/szv.2020.2(2) (In Russian)

Menshikov V., Kokina I., Komarova V., Korshenkov E. (2020) Human-machine collaborationas a factor of labour productivity and efficiency. *European Scientific Journal,* Vol. 16, No. 13. DOI: https://doi.org/10.19044/esj.2020.v16n13p1

More M. (2003) *Principles of Extropy*, Version 3.11. Available: https://web.archive.org/web /20131015142449/http://extropy.org/principles.htm (accessed on 30.12.2024).

Nyholm S., Smids J. (2020) Can a Robot be a good colleague? *Science and Engineering Ethics,* Vol. 26, No. 21. DOI: https://doi.org/10.1007/s11948-019-00172-6.

Obuchowski K. (1993) *Człowiek intencjonalny*. Warszawa: PWN. (In Polish)

Różanowski K. (2007) Sztuczna inteligencja rozwój, szanse i zagrożenia. Zeszyty *Naukowe WWSI,* Vol. 2, No. 2, pp. 109–135.

Sandberg A., Bostrom N. (2011) *Machine Intelligence Survey.* Technical Report #2011-1, Future of Humanity Institute. Oxford University.

Saura García C. (2024) The age of datafeudalism: from digital panopticon to synthetic democracy. *Philosophy and Technology,* Vol. 37, pp. 93–111. DOI: https://doi.org/10.1007 /s13347-024-00786-0

Searle J. (2010) *Umysł. Krótkie wprowadzenie*. Poznań: Rebis, pp. 137–139. (In Polish)

Searle J. (1995) *Umysły, mózg i nauka.* Warszawa: PWN, pp. 57–63. (In Polish)

Seibt J. (2017) *Towards an Ontology of Simulated Social Interaction: Varieties of the As if for Robots and Humans*. *Sociality and Normativity for Robots.* Springer.

Sparrow R. (2021) Why machines cannot be moral. *AI and Society,* Vol. 36, pp. 685–693. DOI: https://doi.org/10.1007/s00146-020-01132-6

Sternberg R. (2011) The theory of successful intelligence. Sternberg R., Kaufman S. (Eds.) *The Cambridge handbook of intelligence*. Cambridge University Press, pp. 504–527. DOI: https://doi.org/10.1017/CBO9780511977244.026

Sutherlin G. (2023) Who is the human in the machine? Releasing the human–machine metaphor from its cultural roots can increase innovation and equity in AI. *AI Ethics.* DOI: https://doi.org/10.1007/s43681-023-00382-6

Szymański K. (2015) Transhumanizm: utopia czy ekstropia. *IDEA – Studia nad strukturą i rozwojem pojęć filozoficznych,* Vol. 27, pp. 159–175. Available: https://repozytorium.uwb.edu.pl/jspui/bitstream/11320/3927/1/idea27_K.Szymanski.pdf (accesed on 30.12.2024). (In Polish)

Szyszkowska M. (2015) Niezbędność filozofii dla rozwoju życia społecznego. *Miscellanea Historico-Iuridica,* Vol. 14, No. 2, pp. 311–317. DOI: https://doi.org/10.15290/mhi.2015.14.02.18 (In Polish)

Thweatt-Bates J. (2016*) Cyborg Selves: A Theological Anthropology of the Posthuman*. London: Routledge. DOI: https://doi.org/10.4324/9781315575728

Tomaszewski T. (1986) *Główne idee współczesnej psychologii*. Warszawa: WP. (In Polish)

Varga S. (2017) Demarcating the realm of cognition. *Journal for General Philosophy of Science,* Vol. 49, pp. 435–450. DOI: https://doi.org/10.1007/s10838-017-9375-y

Varoufakis Y. (2023) *Technofeudalism: What Killed Capitalism*. Random House.

Wasilewska A., Łupkowski P. (2021) *Nieoczywiste relacje z technologią. Przegląd badań na temat ludzkich postaw wobec robotów.* Człowiek i Społeczeństwo, pp. 166–187. (In Polish)

Wiltgen F. (2021) Challenge of balancing analog human (real life) with digital human (artificial life). *Transformacje,* Vol. 110, No. 3, pp. 17–33. Available: https://www.researchgate.net/publication/357649688_PAPER_CHALLENGE_OF_BALANCING_ANALOG_HUMAN_REAL_LIFE_WITH_DIGITAL_HUMAN_ARTIFICIAL_LIFE_ARTIGO_EQUILIBRIO_DESAFIADOR_DO_HUMANO_ANALOGICO_VIDA_REAL_COM_O_HUMANO_DIGITAL_VIDA_ARTIFICIAL (accessed on 30.12.2024).

Zimmerman A., Janhonen J., Beer E. (2023) Human/AI relationships: challenges, downsides, and impacts on human/human relationships. *AI Ethics.* DOI: https://doi.org/10. 1007/s43681-023-00348-8