

**CRITICISM OF THE AI PHENOMENON IN THE CONTEXT OF INDUSTRY 4.0
DEVELOPMENT (EXPERTISE AND FORECASTING)**

Volodymyr Fisanov

Yuriy Fedkovych Chernivtsi National University,
Faculty of History, Political Science and International Relations, Ukraine

v.fisanov@chnu.edu.ua

<https://orcid.org/0000-0001-6550-8903>

Ivan Osadtsa

Yuriy Fedkovych Chernivtsi National University,
Faculty of History, Political Science and International Relations, Ukraine

i.osadtsa@chnu.edu.ua

<https://orcid.org/0000-0001-5593-5944>

Vasyl Baranets

Yuriy Fedkovych Chernivtsi National University,
Faculty of History, Political Science and International Relations, Ukraine

baranets.vasyl@chnu.edu.ua

<https://orcid.org/0009-0003-8096-5821>

Abstract

The article presents a critical analysis of the phenomenon of artificial intelligence (AI) in the context of Industry 4.0 development, with a focus on expertise and forecasting. It examines the conceptual foundations, technological achievements, and risks associated with the implementation of AI in various fields - education, business, security, and geostrategy. Particular attention is given to the instrumentalization of human consciousness, the problems of algorithmic bias, and the consequences of AI use for democratic governance and global security. The article analyzes ongoing debates on the prospects of forming superintelligence and the potential emergence of a singleton as a model of global AI governance, assessing its philosophical, ethical, and geopolitical dimensions. The authors argue that despite the transformational potential of AI in improving local decision-making and optimizing human activity, there are significant risks associated with autonomy, control, and the preservation of humanistic values in the digital age. Effective AI development requires a balance between innovation and ethical considerations to ensure its role as a tool for positive change rather than a threat to autonomy and social stability.

Key words: artificial intelligence, instrumentalization of the mind, politics and AI, super intelligence, Singleton, AI criticism, new educational technologies, geostrategic confrontation, USA, China, digital citizen, Industry 4.0.

Introduction

Every conscious citizen of Ukraine can feel the profound contradictions of today's international environment - one that is increasingly defined by complexity, unpredictability, and illogical dynamics. In our digital age, the trajectories of global development are becoming exceedingly difficult to anticipate. Undoubtedly, since the beginning of the 21st century - and especially in the past decade - amidst the ongoing technological revolution and the rise of Industry 4.0, a new and indispensable factor has emerged. No serious discussion in the technosphere, political sphere, or media landscape is complete without reference to this phenomenon: artificial intelligence, in all its multifaceted dimensions - from large-scale manufacturing to nanotechnology, from medicine, forensics, and creativity to education, military affairs, and foreign policy.

Given this context, there is a growing necessity to critically examine (as far as the available evidence allows) the most controversial aspects of AI use - particularly its instrumentalization and potential for uncontrollability, especially in the military domain at the geostrategic level, where inter-state competition now plays out in the form of a new arms race.

The *central scientific hypothesis* of this article is that the implementation of artificial intelligence across multiple domains of life in the early 21st century not only facilitates innovation but also introduces serious risks, especially by contributing to the technologization of human consciousness. AI is becoming a powerful - yet overly formalized - instrument of influence on key dimensions of human existence. In this regard, scholars, experts, and policymakers must urgently seek mechanisms and levers capable of mitigating, if not eliminating, the negative impacts of AI's rapid global expansion. This article aims to examine the current state of AI deployment and outline preliminary steps for its containment, particularly within the framework of Industry 4.0. It should be noted that any proposed responses to AI-

related challenges are unlikely to be exhaustive, and even seemingly viable solutions must remain open to revision over time.

In a broader philosophical sense, the *article's ultimate purpose*, as envisioned by the authors, is to explore whether the significant financial investments and localized successes in AI can, in the end, reduce (and to what extent) the prevailing existential discomfort of the human condition - bringing about a more harmonious relationship between the individual, society, and the world.

In a narrower analytical sense, the central question - approached through a critical lens - may be hypothetically formulated as follows: Will the risks and challenges associated with the gradual "conquest" of the world by the AI phenomenon ultimately outweigh the localized achievements of contemporary instrumental reason - both those already realized and those likely to emerge in the near future as a result of rapid technological leaps within Industry 4.0?

The methodological framework employed by the authors aligns with the dynamic system through which artificial intelligence is primarily developing - as one of the driving instruments of Industry 4.0. The topic is addressed using an interdisciplinary approach, which introduces the reader to diverse perspectives on the subject under analysis. A critical method is also employed, especially when AI is viewed as an overly formalized instrument with potentially far-reaching effects on modern society. This method is fully justified given that the issue of resisting manipulative dangers associated with AI deployment is likely to become a major component of public interest in both the medium- and long-term.

The authors further attempt to understand the AI phenomenon through the lens of a nonlinear, multi-variant evolution of complex systems. Emphasis is also placed on the need to preserve and strengthen the ethical dimension of AI development - an aspect intrinsic to contemporary European practices. This is essential to ensure the continued humanitarian evolution of complex sociotechnical systems, increasing their manageability and transparency under the supervision of well-trained technical and intellectual experts.

By adopting these methods and perspectives, the authors seek to emphasize the centrality of the human factor in systems of AI control, while also contributing to the awareness of researchers, civil servants, and NGO practitioners, regarding the safe and predictable use of innovative information technologies.

Literature Review

Already in the 20th century, representatives of the renowned Frankfurt School warned that since the dawn of the modern era, human reason has been increasingly formalized - gradually losing its humanistic orientation and, in some cases, even its profound human content. Max Horkheimer, in particular, cautioned against the instrumentalization of reason, emphasizing that the rationalization driving modern civilization tends “to destroy the very substance of reason in whose name the process claims to advance” (M. Horkheimer, 2006, p. 23).

Another domain where both the achievements and risks of AI tools are becoming increasingly evident is education, which may face an erosion of core values and a growing loss of intrinsic motivation among learners. Indeed, when AI “searches, thinks, and writes for you”, as it often does today, the process of education risks becoming superficial, reducing the learner’s engagement with meaning-making. In a world marked by information pluralism, digital intelligence seemingly “assists” in identifying reliable informational anchors - yet often substitutes its own selections for those of human critical thinking. As Niklas Luhmann once noted, in the context of overwhelming informational polyphony, educational priorities must shift toward teaching young people how to choose and think independently. Yet in the present moment, it is often AI that “chooses”, leaving the human mind sidelined from meaningful inquiry and motivated cognition.

European policymakers are striving to preserve democratic values within education, particularly amid AI-driven transformations that dissolve the physical barriers to communication

and consumption. This necessitates the formation of a responsible digital citizenry. Of particular importance are the Council of Europe's 2019 Recommendations, which define a Digital Citizen as a person who "possesses the competences of democratic culture to engage competently and positively with digital technologies", participate in social and civic life, pursue lifelong learning, and commit to upholding human rights and dignity (Recommendation CM/Rec(2019)10).

The competitive global dynamics surrounding AI development - especially among the United States, China, and the European Union - have been extensively studied. Anu Bradford, a Finnish-American scholar specializing in international law, has explored this topic through three key lenses: (1) the concept of digital empires, (2) different types of regulatory battles (both vertical and horizontal), and (3) three forms of power within the digital sphere - private, infrastructural, and regulatory (Bradford, 2023).

Within the Ukrainian academic context, Nataliya Vinnykova (Kharkiv University) has analyzed the geostrategic competition between the U.S., China, and the EU in AI development. She emphasizes that European countries largely adhere to an ethics-centered approach in AI production and implementation. However, they lag behind the U.S.-China duopoly in terms of investment and infrastructure, which constrains their global leadership potential in the AI field (Vinnykova, 2021, pp. 24-25).

Meanwhile, Shadi Banitaan (University of Detroit) and colleagues from various institutions have focused on the role of AI within modern industry. They highlight AI as "the most promising achievement of Industry 4.0 for business," discussing its advantages, challenges, and practical applications in this area of innovation. According to the authors, IT infrastructure investment and talent acquisition are crucial to unlocking AI's full potential (Banitaan et al., 2023).

In this context, Alejandro Chacón, Cecilio Angulo, and Pere Ponsa from the Polytechnic University of Catalonia (Spain) explore the formation of a new type of workforce. They emphasize strategies for establishing effective "synergistic interaction between operators and

machines”, which is seen as essential for the next phase of industrial-technological evolution (Chacón et al., 2020).

Among the first Ukrainian scholars to analyze the challenges and prospects of integrating “intelligent machines” - along with the complexities of further technologization and even the self-organization of AI within the well-known paradigm of “order emerging from chaos” - were Mykola Ozhevan and Dmytro Dubov from Kyiv. In their 2017 work, the authors distinguish two foundational methodological approaches originally advanced by the founding fathers of cybernetics and computer science: Norbert Wiener and John von Neumann. According to followers of Wiener, intelligent machines should remain human assistants - mere instrumental extensions of our cognitive capacities. However, adherents of von Neumann’s school argue that these machines may eventually reach the stage of autonomous existence and self-organization, evolving into autopoietic systems (Ozhevan & Dubov, 2017, p. 70).

Notably, one of the co-authors of that study has also examined the evolution of modern media through the lens of Marshall McLuhan’s methodological framework, emphasizing the growing potential of media technologies as instruments for extending and augmenting the human body and mind (Fisanov, 2020).

The legal foundations and EU regulatory frameworks aimed at countering manipulative technologies - particularly those using AI to distort public opinion and erode trust in democratic institutions - are explored by Eglė Kavoliūnaitė-Ragauskienė of the Law Institute of Lithuania (Kavoliūnaitė-Ragauskienė, 2024). Her research underscores that no purely technical solution can fully address AI-enabled manipulation; rather, the response must be anthropological. This means enhancing human digital literacy, cultivating critical thinking, and fostering personal responsibility - particularly in resisting disinformation and manipulative influence campaigns via social media and digital platforms.

A cornerstone in the field of artificial intelligence remains the comprehensive textbook

by Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, often described as a reference compendium in AI studies. Their work centers on the concept of the “intelligent agent”, detailing its operational models, including production systems, neural networks, reactive agents, real-time planners, and decision-theoretic frameworks (Russell & Norvig, 2003).

The philosophical and technological implications of superintelligence have been critically explored by Oxford philosopher Nick Bostrom, who maps potential trajectories for the future of AI and its transformative impact on the human condition in the 21st century (Bostrom, 2020).

A recurring theme in contemporary AI discourse is the issue of value alignment. As highlighted by Shengnan Hu and colleagues, a highly optimized superintelligence must be oriented toward serving human needs, fulfilling human desires, and maximizing the realization of human values (Shengnan et al., 2021). This normative framework is central to guiding AI development amid the rapid changes ushered in by Industry 4.0 and the global digital transformation.

In this context, the authors of this article hope that their focus on AI within the framework of Industry 4.0 will contribute to academic and policy-oriented discussions surrounding Ukraine’s digital sovereignty. Their intention is to inform scholars and public sector professionals working to strengthen national resilience against emerging global risks and to foster the development of a responsible digital citizenry for the 21st century.

Results and Discussion

When analyzing the fundamental trends in the development of artificial intelligence, one must acknowledge the undeniable achievements of this emerging virtual technology across many local spheres of society - from medicine to business, from media and management systems to educational practices. Let us focus, for example, on the field of education, which is particularly familiar to the authors. Experts emphasize that innovative development based on AI technologies has led to the rise of expert systems, chatbots, intelligent tutors, personalized learning systems,

virtual learning environments, and machine learning technologies (Baranets & Osadtsa, 2024, pp. 162-163).

When it comes to personalized learning, AI systems can analyze data on students' performance, behavior, and learning styles in order to individualize educational programs. Adaptive learning systems powered by these technologies adjust to the needs of each learner (Hrytsenchuk, 2024). For example, in the United States, the prospects for personalized education are being addressed through the creation of a national public platform, EdSafe AI, aimed at facilitating societal adaptation to artificial intelligence technologies. One of its goals is to engage “diverse voices” in order to “educate policymakers and build a fair and safe AI-powered education ecosystem”. Priority is given to innovative technologies such as personalized learning systems, individualized assessments, AI-driven tutoring, and teaching support tools. Ultimately, broad cross-sector collaboration involving academia, industry, and government is expected to accelerate practical implementation, “transforming cutting-edge research into real-world solutions” (Opportunity at Scale, 2025).

At the heart of U.S. educational innovation lies the integration of AI techniques into master's and doctoral programs, with direct engagement of international talent. It is worth emphasizing again that developers of American educational strategies prioritize the systematic inclusion of students' knowledge, experience, and cultural heritage. In this context, artificial intelligence, as an effective tool, can provide “more customized learning resources to meet local needs” (Artificial Intelligence and the Future of Teaching and Learning, 2023).

However, in implementing AI-based educational practices, educators themselves have raised significant concerns. Some worry that they could be replaced by “digital agents”. The U.S. Department of Education, however, has firmly rejected the notion that AI can replace teachers. Many also see danger in algorithmic bias, particularly in imperfect exam monitoring systems that may unfairly flag certain student groups for disciplinary action. Moreover, the infrastructure-

level and often invisible use of AI tools may threaten transparency and trust. Even more troubling is that AI may generate information that appears credible but is, in fact, inaccurate or lacks any real foundation - a topic we will address in greater detail later.

These points to the broader phenomenon of algorithmic discrimination. This occurs when algorithms, trained on biased data or designed with flaws, reinforce existing social prejudices or produce unjust outcomes. In education, this may manifest in systemic inequality in learning opportunities and resources recommended to certain groups of students. Thus, the following forecasts are being made today: all participants in the educational process in the United States will require both time and autonomy to implement innovations that more effectively engage teachers and students as co-creators, with particular attention to balancing diverse interests. One anticipated positive outcome is the prioritization of digital equity through the expanded use of educational technologies and EdTech platforms (Artificial Intelligence and the Future of Teaching and Learning, 2023).

It is also worth paying attention to specific aspects of AI in the fields of business and production, especially given that this sector is key to the development of Industry 4.0. According to most experts, the “intervention” of artificial intelligence in these areas may have both positive and negative consequences. For instance, a contemporary PwC business platform highlights the concept of “AI agents” - digital workers. As noted by Anthony Abbattello, leader of PwC’s Workforce Transformation Practice, such agents can “easily double your knowledge workforce”, performing roles like sales and field support, while accelerating time-to-market, improving customer engagement, and transforming product design. Moreover, an AI agent can autonomously handle numerous tasks, such as processing routine customer requests, generating initial drafts of software codes, or converting human design concepts into functional prototypes. Nevertheless, the human factor remains critical - people instruct and supervise AI agents during the automation of simpler tasks; they collaborate with digital agents on more complex assignments like design for example; and they gradually “orchestrate” teams of agents, assigning

tasks and refining outcomes (2025 AI Business Predictions).

Despite these advantages, the use of digital agents in business can also bring drawbacks. Antonio Aloisi, co-author of the noteworthy study “Your Boss Is an Algorithm”, argues that managerial authority in the workplace can now be circumvented by technical tools that extend command-and-control mechanisms into intimate spaces - even beyond working hours. In this way, “algorithmic bosses” reveal their most controversial nature. However, big data can offer solutions to numerous complex questions. How can we assign workers to tasks they are best suited for? How can it be ensured during restructuring that the most loyal employees will stay with the company? How can diverse and well-balanced teams be compared to ensure winning results? Managers seek answers to these everyday dilemmas - and AI may have them (Aloisi, 2022).

Still, as in education, business applications of AI reveal a growing dependence on algorithms and the rising complexity of working with big data. This raises a fundamental question: can we avoid the gradual erosion of human participation in many processes? Such a trajectory provokes unavoidable concerns about justice, accountability, equality, and respect for human rights.

The rapid growth of artificial intelligence is accompanied by serious environmental costs. According to estimates by Strubell et al., training a single language model can result in CO₂ emissions equivalent to those produced by five cars over their entire lifetimes. This raises urgent concerns about the sustainability of AI development (Strubell et al., 2019).

Another important point: today - as confirmed by many practitioners - the development of modern AI and robotics technologies is giving rise, so to speak, to a new dimension of contemporary capital - cognitive capital. This consists of algorithms, knowledge, and ideas synthesized from streams of information and decisions exchanged between individuals and organizations. A key example is the AlphaGo program, which defeated the world champion in

the game of Go and possesses all the knowledge, experience, and strategies necessary to beat any human player. In other words, it has cognitive capital specific to Go. By analogy, business companies and various platforms are creating cognitive capital not only in gaming but also in other domains - particularly through the use of AI for Industry 4.0 - increasing its value, for instance, for hedge funds and innovative medical technologies (Rao, 2019). Modern technologies of Industry 4.0, as well as the current advancement of artificial intelligence, enable cyber-physical systems - focused on human-machine interaction - to transition from a purely physical interaction paradigm to a cognitive one. In this context, human monitoring and implementation of innovative solutions are more relevant than ever (Chacón et al., 2020). At the same time, the task of protecting and promoting democratic achievements, both globally and at the European level, is becoming increasingly important in the digital age - particularly in the fields of education and civic engagement (Baranets & Osadtsa, 2025, p. 57). With the rapid disappearance of physical barriers to communication and consumption, there is an urgent need to foster the development of responsible digital citizenship.

Since the objective of this analytical material is to critically reflect on the AI phenomenon, let us explore this issue further. Experts and observers increasingly argue that one of the fundamental questions in AI development is whether a machine can truly “think”, and if a certain level of “technological thinking” is ever achieved, whether such thinking would resemble human cognition. Related to this is the question: can we, so to speak, define the parameters of human thought in the same way we define an algorithm?

Turning to recent studies, let us consider the prospects of developing artificial superintelligence. One of the most prominent theorists and advocates of its creation, Nick Bostrom, in his seminal work *Superintelligence: Paths, Dangers, Strategies* (2014), predicts that by the second half of the 21st century, humanity may succeed in developing a “safe and beneficial artificial superintelligence”. On this challenging path, humanity may experiment with genetic and embryonic selection, as well as the creation of neural interfaces - which could

initially lead to an increase in “the number of people with more advanced minds”. However, Bostrom argues that the biological brain will ultimately play only an “auxiliary role”, as collective intelligence is quantitative rather than qualitative.

According to Bostrom, even an enhanced biological brain is incapable of rapidly performing countless sequential operations. Moreover, it tires after a few hours of work, and after “several decades of subjective time”, it begins to deteriorate. A superintelligence, operating millions of times faster than a human, would function in a virtual environment and focus primarily on the information economy (Bostrom, 2020, pp. 141-142). This, in the context of the ongoing development of Industry 4.0, will have critical significance.

Many theorists and experts believe we are entering the age of technological singularity, which will become an irreversible stage of innovation-driven technological evolution. One possible starting point for this phase is the so-called intelligence explosion, which could suddenly trigger the era of superintelligence. However, as noted by M. Ozhevan and D. Dubov, humanity may encounter a singularity that transforms into a state of “catastrophic complexity” - an extremely unstable and unpredictable world plagued by various crises from which escape becomes increasingly difficult and costly (Ozhevan & Dubov, 2017, p. 71).

As Bostrom writes, the road ahead is filled with major surprises and risks. One of the least favorable scenarios for humanity is the possible takeover of the world by AI. Such “governance” may lead to the emergence of a singleton, a global order in which a single authority controls all matters and determines the fate of intelligent life on Earth. Bostrom notes that the United States theoretically could have created a singleton after World War II using its nuclear monopoly, but this never became political reality.

The ideal scenario for developing artificial superintelligence would involve a system characterized by “scientific curiosity, kindness, and concern for others”. Yet this raises the question of whether an autonomous AI system would actually be friendly to humanity. A global

digital agent may possess non-anthropomorphic goals, being instrumentally motivated to acquire unlimited resources and neutralize any threats to itself or to the outcomes of its own activity. With internet access, such a superintelligence could continue to increase its power by “gaining control over hardware and data” (Bostrom, 2020, pp. 141-142).

Still, even if superintelligence were entirely benevolent, could it avoid making a fatal mistake? To prevent such a scenario - even in isolated segments of its “autonomous activity” - AI must be “fed with good knowledge”, and this process must be controlled. Otherwise, the approach could do more harm than good. It is worth remembering that AI is a tool meant to facilitate, improve, and optimize work (Tytarchuk, 2024).

While scientists and experts debate the properties and usefulness of AI for humanity, the race for global leadership in AI is taking on clear geostrategic dimensions. This primarily involves a military-political rivalry over technological dominance between China and the United States. In this context, the article “The AI Power Paradox” by Ian Bremmer and Mustafa Suleyman, published in *Foreign Affairs* in 2023, deserves attention. Experts argue that the central issue is AI governance and control. Policymakers often find themselves trapped in a dangerous false dilemma: should they use AI to expand national power or suppress it to minimize risk? The focus, they argue, must be on preventing a full-scale AI arms race between states. The difference between China and the West will be most evident in terms of organization. (Bremmer & Suleyman, 2023). In China, nearly all AI-related spheres are controlled by the state, while in the West - particularly in the United States - the trajectory of artificial intelligence in the near future will be largely shaped by the decisions of several private companies. These trends in global digital development are analyzed in the aforementioned book by A. Bradford. According to the author, the central challenge in the medium term will be the still uncompromising battle between techno-democracies and techno-autocracies. Since the United States and the European Union strive for greater alignment, the most intense horizontal battle is currently unfolding between the coalition of techno-democracies led by the U.S. and the EU - which supports a

rights-based version of the European regulatory model - and the group of techno-autocracies adhering to China's state-centric regulatory model. However, to date, democratic nations have neither proposed a viable alternative to China's Digital Silk Road nor achieved meaningful coordination in developing technological standards that reflect their shared values (Bradford, 2023). Moreover, the geostrategic rivalry between the two global powers - the U.S. and China - in the realm of AI hardly inspires optimism. We can only hope that a tragic confrontation involving artificial intelligence will not become humanity's future and that we will avoid a "nuclear holocaust" (Fasching, 1993, p. 69).

In other words, technologists - not politicians or bureaucrats - will wield control over a force capable of deeply transforming the power of nation-states and their foreign strategies. All of this makes the challenge of governing artificial intelligence unlike any other that governments have faced before (Bremmer & Suleyman, 2023).

At the same time, experts highlight that China's current AI strategy - as defined in the Next Generation Artificial Intelligence Development Plan and the Made in China 2025 program - is aimed not only at technological advancement but also at building a reliable base for innovation and self-sufficiency. Beijing is pursuing goals that intertwine economic growth with military objectives. However, export controls on semiconductors imposed by the U.S. underscore the vulnerability of China's strategy, revealing that while its ambition for global leadership in AI is strong, it is also deeply entangled in geopolitical challenges (Al Midfa, 2025). The competition between these two global actors over AI technologies is taking place in a space already fraught with the dangers of economic instability, large-scale disinformation, and the threat of autonomous warfare. According to Jieli Li, a researcher from Ohio University, one of the most urgent risks is the development of lethal autonomous weapon systems (LAWS) - so-called "killer robots". These AI-powered drones and robotic warfare systems operate with minimal human oversight. The lack of comprehensive international regulations concerning AI-based weapons

further accelerates the global arms race and increases the risk of warfare (Li, 2025). Such undesirable developments may fuel significant manifestations of regional imperialism.

Let us reiterate: the fundamental goal is to create an artificial superintelligence that is not aggressive, but beneficial to humanity. Most proponents of AI oversight argue that human values are too complex and poorly understood to be explicitly programmed into a superintelligence. Instead, it should be designed with “indirect normativity” - that is, the ability to learn and internalize human values through deep learning. However, this path is fraught with challenges and obstacles. One such challenge is the illusion that political control over technologies and their creators can be strengthened. The spread of such bureaucratic illusions is often fueled by mass media and social networks. Unfortunately, as American scholar Darrell Fasching writes, modern technological society is almost impervious to effective political control, with politicians frequently distracted by inefficient bureaucratic solutions (Fasching, 1993, p. 69).

There is, however, another vision of the prospects for technological development, including superintelligence. For example, the well-known American mathematician Richard Bellman, founder of the algorithmic method of dynamic programming (used for achieving optimal solutions), noted the complexity of the external world and speculated that it might take a computer several thousand years to fully comprehend the myriad of its interconnections and influences. Moreover, significantly more time would be needed to logically “make sense” of the acquired information (Bellman, 1980).

But even the notion of what it means for AI to “make sense” of information raises numerous unresolved questions. For instance, one major challenge lies in teaching machines to “think in images”. Ukrainian researcher Pavlo Semenenko observes that, in order to generate the concept of a chair, an AI system must process a detailed description: how many legs it has, their shape, the seat, the backrest, the material, whether it is standing or lying down, etc. AI must process a vast amount of data to form the image of a chair. Meanwhile, the human brain generates that image instantly. The issue is that while AI also processes such information instantly, it does so

in principle through a type of processing that lacks access to the human way of knowing (Semenenko, 2024).

It is also very difficult to predict when and where artificial intelligence will make a mistake, because the models are too large and opaque. It may give the correct answer to a complex question and the wrong answer to a simple one. Why? That remains one of the key questions that “neither AI nor its creators can currently answer”. As tech writer Devin Coldewey notes, models do not “understand” facts or concepts, but merely statistical relationships between words. Therefore, their responses are often closer to an “aesthetic impression” than a reliable answer. Their goal, essentially, is to “plausibly, not accurately, fill in the gap” - which is why, in principle, their responses cannot be fully trusted. Of course, this situation may change in the future - perhaps even the near future. Everything is evolving so rapidly and unpredictably that nothing can be said with certainty. But for now, this remains a useful mental model to remember: AI wants “you to believe it, and it will say anything to increase its chances” (Coldewey, 2023).

In our view, one of the major issues that will be addressed in the coming decades within the context of AI development is not only the formalization of natural language, but also the elimination of polysemy (semantic ambiguity) in interpreting meanings. As Heinrich Rickert aptly noted, excessive abstraction usually leads to a detachment from reality. Maximum formalization inevitably results in the loss of many meanings, and preserving and clearly defining those meanings during deep learning will become a crucial factor for AI evolution.

The path toward the widespread integration of AI raises numerous complex questions with no easy answers. Today’s emerging forms of self-understanding and social organization offer great potential - but they also demand delicate resolution of ethical issues, including equality, autonomy, and the unpredictable consequences of artificial intelligence development.

The pioneering work of British researcher Kevin Warwick, known for developing neuro-computer interfaces, has sparked critical discussions on the future of human identity and self-

perception. Indeed, the far-reaching impact of artificial intelligence on social structures and institutions - transforming how we work, communicate, and govern - adds new layers of complexity to the ongoing technological revolution. As we stand on the threshold of a potential post-labor society, the emerging questions are as philosophical as they are practical. Traditional paradigms of work, personal fulfillment, and social contribution are being upended, prompting us to reconsider what it means to lead a meaningful life in a world increasingly mediated by technology (Lykhats'kyi, 2024).

Engaged experts, when discussing broader approaches, increasingly address the pressing issues surrounding the development of Industry 4.0 - particularly questions of innovative transformation within the production sector. It is anticipated that by the 2040s and 2050s, in most countries integrated into the global socio-economic system, the paradigms of Industry 4.0 and 5.0 will become dominant and practically unavoidable. These industries will evolve within a triangular framework that unites technology, the Internet, and manufacturing processes.

A special report by the World Bank, aptly titled *Digital Dividends*, highlights that the optimal vision for global development lies in the creation of a unified digital global market, unrestricted by national infrastructure. This model would allow citizens of developing countries to access the same products and services as those available in more advanced economies (World Bank, 2016).

At the same time, the challenge of protecting and advancing democratic achievements on both global and European levels in the era of digitalization is becoming increasingly relevant in both educational and civic spheres (Baranets & Osadtsa, 2025, p. 57). In a context where physical barriers to communication and consumption are rapidly vanishing, there is an urgent need to foster the development of responsible citizenship. The Council of Europe's 2019 recommendations introduced the concept of the digital citizen. According to the definition presented in the document, a digital citizen is "a person who possesses the competences of democratic culture to engage competently and positively with digital technologies", actively

participates in social and civic life, embraces lifelong learning, and is committed to the ongoing protection of human rights and dignity (Recommendation CM/Rec(2019)10).

The foundation of civic responsibility in the digital age must lie in knowledge and critical understanding of the world, alongside the systematic development of autonomous learning skills - particularly those rooted in communicative abilities, adaptability, and flexibility, as well as the pursuit of cooperation and conflict resolution (Richardson & Milovidov, 2020).

Conclusions

Thus, in the authors' view, it is essential to develop adequate conceptual models and practical steps capable - within a medium-term perspective - of restraining AI as an overly formalized tool of influence on various essential dimensions of human existence.

First, in the coming decades, we must double or even triple our efforts to harness the transformational power of artificial intelligence to improve human well-being and promote social and economic development in the context of Industry 4.0.

Second, two major development scenarios are currently observable along the "local-global" axis. The first is relatively optimistic: in this case, AI serves as a tool for localized development and solving segmented problems, primarily within the business domain - platform capitalism being a prime example within the Industry 4.0 framework. The second, significantly more pessimistic scenario envisions a global, generative artificial intelligence that "seeks" to become a universal mediator and overseer, creating an autonomous and potentially hostile singleton system. In this context, it is necessary to build elite teams of high-level IT specialists capable of solving the most complex challenges in interacting with AI, with the goal of guiding its development in a human-centered direction.

A virtuoso expert in the field of information technologies, possessing both predictive insight and intuitive capacity, will be able to monitor, verify, and adjust AI-generated outputs -

and must be able to detect even hidden variables in its evolution. As noted by one of Silicon Valley's prominent visionaries, Carver Mead, "There is always something new threatening humanity". To be at least partially prepared for the arrival of groundbreaking innovations in AI, Mead suggested, one must learn to listen to technology and take from it what it wants to tell you (Interview with Carver A. Mead, 2000).

Third, without losing sight of the horizon of innovative technological solutions within Industry 4.0, we should continue to strengthen the ethical dimension embedded in the modern political approaches of the united Europe - especially as it relates to the practical implementation of conceptual models and practices in AI development. Such an approach will distinguish us, as Europeans, from the competing digital empires of today - namely, the American and Chinese models. Moreover, the development of an ethical foundation for addressing these deep transformations will define not only our relationship with technology but also the very essence of what it means to be human in the age of artificial intelligence and Industry 4.0.

Perhaps only by following this path will humanity be able to grasp the transformative dynamics and future mysteries of AI - as an anticipated and powerful instrument of positive global change.

References

- Aloisi, A. (2022, April 25). *Boss ex machina: Employer powers in workplaces governed by algorithms and artificial intelligence*. In A. Lo Faro (Ed.), *New technologies and labour law: Selected topics* (forthcoming). SSRN. <https://doi.org/10.2139/ssrn.4096870>
- Al Midfa, N. (2025, March 28). China's AI strategy: A case study in innovation and global ambition. <https://trendsresearch.org/insight/chinas-ai-strategy-a-case-study-in-innovation-and-global-ambition/>
- Banitaan, S., Al-Refai, G., Almatarneh, S., & Alquran, H. (2023). A review on artificial intelligence in the context of Industry 4.0. *International Journal of Advanced Computer Science and Applications*, 14(2). <https://doi.org/10.14569/IJACSA.2023.0140204>
- Baranets, V., & Osadtsa, I. (2024). Naukovi pidkhody do vyvchennia roli suchasnykh tekhnolohii u mizhnarodnykh vidnosynakh. *Mediaforum: Analityka, prohnozy*,

informatsiinyi menedzhment, 15, 155–170.

<https://doi.org/10.31861/mediaforum.2024.15.155-170>

Baranets, V., & Osadtsa, I. (2025). Mozhlyvi ryzyky ta perevahy vykorystannia ShI v dyplomatii ta mizhnarodnykh vidnosynakh. In *Mizhnarodni vidnosyny v hlobalnomu ta rehionalnomu vymiri: Materialy Vseukrainskoi naukovo-praktychnoi konferentsii* (pp. 56–60). V. N. Karazin Kharkiv National University.

https://dspace.wunu.edu.ua/bitstream/316497/54497/1/Міжнар.відносини_конф_2025%20%281%29.pdf

Bellman, R. (1980). *Analytic number theory: An introduction*. Benjamin/Cummings Publishing Company.

Bostrom, N. (2020). *Superintelligence: Paths, dangers, strategies*. Nash Format.

Bradford, A. (2023). *Digital empires: The global battle to regulate technology*. Oxford University Press. <https://doi.org/10.3917/poeu.083.0130>

Bremmer, I., & Suleyman, M. (2023). The AI power paradox: Can states learn to govern artificial intelligence—before it’s too late? *Foreign Affairs*.

<https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox>

Cardona, M. A., Rodríguez, R. J., & Ishmael, K. (2023, May). *Artificial intelligence and the future of teaching and learning*. U.S. Department of Education.

<https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf>

Chacón, A., Angulo, C., & Ponsa, P. (2020). Developing cognitive advisor agents for operators in Industry 4.0. In L. Romeral Martínez, R. A. Osornio Rios, & M. Delgado Prieto (Eds.), *New trends in the use of artificial intelligence for Industry 4.0*. IntechOpen.

https://library.oapen.org/bitstream/handle/20.500.12657/43835/external_content.pdf

Coldewey, D. (2023, April 3). The great pretender: AI doesn’t know the answer, and it hasn’t learned how to care. *TechCrunch*. <https://techcrunch.com/2023/04/03/the-great-pretender/>

Council of Europe. (2019, November 21). *Recommendation CM/Rec(2019)10 on developing and promoting digital citizenship education*. <https://rm.coe.int/090000168098de08>

Fasching, D. J. (1993). *The ethical challenge of Auschwitz and Hiroshima: Apocalypse or utopia?* State University of New York Press.

Fisanov, V. (2020). Vplyvy Marshalla Makliuena na hlobalni ta kanadski mediini praktyky: Pohliad iz XXI stolittia. *Istoryko-politychni problemy suchasnoho svitu*, 41, 73–82.

<https://doi.org/10.31861/mhpi2020.41.73-82>

Horkheimer, M. (2006). *Krytyka instrumentalnoho rozumu*. PPS.

Hrytsenchuk, O. (2024). Vykorystannia shtuchnoho intelektu v osviti: Tendentsii ta perspektyvy v Ukraini ta za kordonom. *Visnyk kafedry YUNESKO “Neperervna profesiina osvita XXI stolittia,” 10*. <https://lib.iitta.gov.ua/id/eprint/743864/1/123-Article%20Text-371-1-10-20241129%203.pdf>

- Kavoliūnaitė-Ragauskienė, E. (2024). Artificial intelligence in manipulation: The significance and strategies for prevention. *Baltic Journal of Law & Politics*, 17(2), 116–141. <https://doi.org/10.2478/bjlp-2024-00018>
- Li, J. (2025). Governing high-risk technologies in a fragmented world: Geopolitical tensions, regulatory gaps, and institutional barriers to global cooperation. *Fudan Journal of the Humanities and Social Sciences*. Advance online publication. <https://doi.org/10.1007/s40647-025-00445-4>
- Lykhatskyy, A. (2024). Konverhentsiia shtuchnoho intelektu ta liudskoi identychnosti v suchasnomu filosofskomu dyskursi. *Kulturolohichniy almanakh*, 1(9). <https://almanac.npu.kiev.ua/index.php/almanac/article/view/335/314>
- Ozhevan, M. A., & Dubov, D. V. (2017). *Homo ex machina: Filosofski, kulturolohichni ta politychni peredumovy formuvannia konverhentnoho suspilstva*. NISS. https://niss.gov.ua/sites/default/files/2018-01/HomoExMachina_druk-53c4b.pdf
- PwC. (2019). *Gaining national competitive advantage through AI: Policy making and national AI strategies*. <https://www.pwc.lu/en/technology/docs/gaining-national-competitive-advantage-through-ai.pdf>
- PwC. (2025). *2025 AI business predictions*. <https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html>
- Russell, S. J., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). Prentice Hall.
- Semenenko, P. V. (2024). Aspekty rozvytku samosti shtuchnoho intelektu. https://www.ipai.net.ua/docs/ai_conf_15.03.2024.pdf
- Shengnan, H., Eugene, K., Shahrokh, N., & Svee, E.-O. (2021). Aligning artificial intelligence with human values: Reflections from a phenomenological perspective. *AI & Society*. <https://d-nb.info/1243958855/34>
- Tytarchuk, S. A. (2024). Vplyv shtuchnoho intelektu na sferu hostynnosti. https://www.ipai.net.ua/docs/ai_conf_15.03.2024.pdf
- Vinnykova, N. A. (2021). Shtuchnyi intelekt yak novyi chynnyk heopolitychnoho vplyvu. *Visnyk Kharkivskoho natsionalnoho universytetu imeni V. N. Karazina, seriiia "Pytannia politolohii,"* 40, 21–28. <https://doi.org/10.26565/2220-8089-2021-40-03>
- EDSAFE AI Alliance. (2025, April). *Opportunity at scale: The case for public infrastructure for AI in education*. https://www.edsafeai.org/files/ugd/f384a6_14ca393d3f7440a781bb00a75a0ae120.pdf
- Interview with Carver Mead by Shirley K. Cohen. (2000). *Oral history project*. California Institute of Technology Archives. [https://worrydream.com/refs/Mead_1996_-_Oral_History_\(Cohen\).pdf](https://worrydream.com/refs/Mead_1996_-_Oral_History_(Cohen).pdf)