Philosophy, Economics and Law Review

Vol. 5. No. 1. 2025
Received: 27.12.2024
Revised: 17.04.2025
Accepted: 27.05.2025

Journal homepage: https://phelr.com.ua/en

UDC 004.8:177:342.7

DOI: 10.63341/2786-491X-2025-1-35

Cognitive horizons of humanity: Ethical challenges of superintelligent AI in the global context of rights and justice

Andrii Lykhatskyi*

Postgraduate Student
Odesa Polytechnic National University
65044, 1 Shevchenko Ave., Odesa, Ukraine
https://orcid.org/0009-0002-3735-2862

Abstract. The study aimed to theoretically determine the normative challenges arising from the emergence of artificial superintelligence (ASI) capable of autonomous thinking and decision-making beyond human epistemic control. The interdisciplinary analysis considered the universality of the human rights concept, the normative boundaries of moral subjectivity, and the possibility of granting AI a legal status. The study presented a conceptual typology of legal models of superintelligence status, including instrumental, limited-subjective and full-fledged paradigms, with an indication of their advantages and risks. In particular, the instrumental model retains full human control but loses its normative relevance in the context of system autonomy; the limited subjectivity model can be used to delegate responsibility within certain limits without violating the principle of human supremacy; the full-fledged model, which equates ASI with a legal entity, questions the current ethical framework of legal personality. The main results of the study demonstrated that the anthropocentric legal doctrine is insufficient to consider the cognitive multiplicity of agents who do not have bodily vulnerability but demonstrate a high level of autonomy, reflexivity and adaptability. The study established that the cognitive asymmetry between a human and a superintelligent agent generates a new form of epistemic injustice that makes it impossible to participate equally in the procedure of moral decision-making. The study proposed the concept of limited legal personality as a normative compromise which ensures legal certainty and delimitation of liability between the participants of interaction. The results have implications for the philosophy of law, regulatory policy in the field of AI, and interstate regulatory regulation. They can be used to form international approaches to the certification of autonomous systems, guarantee the explainability of algorithmic decisions and preserve human normative autonomy in the era of cognitive multiplicity

Keywords: cognitive asymmetry; epistemic inequality; legal personality; legitimacy; moral agent; anthropocentrism

Introduction

In the 21st century, humanity has faced an unprecedented challenge posed by the rapid development of intelligent technologies that not only mimic but also potentially surpass human cognitive abilities. This process significantly transforms conceptual ideas about knowledge, consciousness, moral responsibility and justice. The emergence of artificial superintelligence (ASI) capable of making autonomous decisions, producing new conceptual structures and performing epistemic activities at levels that are beyond the reach of human consciousness was emphasised. The research relevance is determined by the fact that most modern philosophical and legal concepts are based on the anthropocentric model of consciousness and morality. However, the emergence of systems capable of acting autonomously, transforming personal cognitive architecture and exhibiting behaviour similar to moral judgement requires going beyond these models. In this context, there is an urgent need for a fundamental rethinking of key ethical and legal concepts, including the universality of human rights, the criteria of moral agency, and the foundations of global justice. The question arises as to whether the

Suggested Citation:
Lykhatskyi, A. (2025). Cognitive horizons of humanity: Ethical challenges of superintelligent AI in the global context of rights and justice. *Philosophy, Economics and Law Review*, 5(1), 35-49. doi: 10.63341/2786-491X-2025-1-35.

**Corresponding author*

anthropocentric model of morality retains its normative value in the context of a non-human agent with autonomous cognitive capacity.

As noted by R. Rueda (2023), contemporary normative philosophy has not yet developed a coherent position on cognitively superior agents, and international law is unable to cover new forms of subjectivity that are not human. On a practical level, this creates the threat of a regulatory vacuum, in which technology shapes reality faster than ethical and legal systems can comprehend it. According to a study by T. Peters (2024), in a globalised world in which algorithmic systems make decisions that affect the lives of hundreds of millions of people, traditional notions of justice, transparency and responsibility are losing their exclusively anthropocentric nature. The potential emergence of superintelligence as a new participant in social life appears not only as a technological innovation but also as a source of ontological destabilisation of the established normative foundations of international law and the concept of human rights. In this context, there is a need to rethink the fundamental ethical guidelines that can provide a legitimate and functional model for the coexistence of cognitively unequal agents, regardless of biological or non-biological origin. Similar warnings were also expressed by C.H. Kan (2024) in an analysis of the impact of AI technologies on democracy and human rights. The study emphasised that existing regulations and policies should be rethought to protect human rights and uphold democratic values.

The issue of cognitive asymmetry is actively discussed in interdisciplinary research. For instance, V. Moravec *et al.* (2025) analysed how algorithms for personalised access to knowledge increase information inequality, which is directly related to the cognitive capacity of the user. Their conclusions indicated that the very architecture of knowledge distribution already functions as a mechanism of epistemic stratification, laying the groundwork for deeper inequality in the event of the emergence of ASI. M. Kiškis (2023) emphasised that legal models focused exclusively on humans will not be able to ensure sustainable coevolution in the event of the emergence of autonomous AI with a cognitively different subjectivity. The study highlighted the need for the development of the concept of "cognitive justice" as a tool for balancing rights and responsibilities between agents of different natures. In addition, the study by J. Mökander & R. Schroeder (2022) examined the potential of AI in the development of social theory. The study noted that modern AI systems, although capable of synthesising knowledge from different sources and applying it to a wide range of problems, still have limitations in three key aspects: semantisation, transferability, and generativity.

At the same time, several authors draw attention to the lack of conceptual certainty of the "ethical status" of superintelligent agents. B.C. Stahl (2023) proposed the idea of "meta-responsibility", which imposes on human institutions the obligation to ensure moral validation of actions even for complex systems acting independently. However, this model does not answer the question of whether an ASI can be responsible on its own if its cognitive horizons go beyond human understanding. Similarly, J. Kulveit *et al.* (2025) distinguished between direct and structural risks of ASI, noting that the greatest threat is the gradual delegitimisation of human moral authority in areas where superintelligence demonstrates systemic superiority.

The gaps in current research show that key aspects related to the moral relevance of universal human rights in the context of cognitive inequality, the justification of the possible legal status of superintelligence as a participant in the global social contract, and the mechanisms of transformation of the concepts of justice, responsibility and legitimacy in the post-human cognitive landscape have not been sufficiently developed. These issues require a deep philosophical analysis that goes beyond the technocratic approach and appeals to the foundations of political philosophy, ethics of consciousness and the concept of human rights.

The study aimed to critically reflect on the normative, ethical and legal challenges posed by the emergence of ASI, in the context of the universality of human rights, cognitive justice and the possibility of legal subjectivity of such systems. In this regard, the study envisaged solving three interrelated tasks: first, to study the phenomenon of cognitive inequality between humans and ASI as a philosophical and ethical issue; second, to analyse possible models of the legal status of superintelligence in the context of international law; third, to assess the potential of ASI as an entity that can maintain or undermine global peace, incorporating the need to develop universally recognised ethical standards for agents of a new nature.

Materials and Methods

The study was conducted in January-April 2025 as part of an interdisciplinary project on the ethical and legal aspects of cognitive technologies. The study was conducted in an analytical format without the involvement of empirical computational or technological components, within the humanities academic field. The first stage of the research involved collecting a corpus of relevant academic sources published in peer-reviewed journals on the philosophy of technology, human rights, cognitive science, and political theory. The literature search was carried out using academic databases such as Scopus, Web of Science, PhilPapers, SpringerLink, and Journal Storage, as well as specialised open-access archives such as arXiv and the Social Science Research Network to cover interdisciplinary research on AI and ethics. The selection of sources was limited to the timeframe of 2021-2025 to ensure relevance and compliance with the latest trends in the development of AI

technologies. The selected texts are classified into four main thematic clusters: cognitive asymmetry between humans and AI; legal status of intelligent agents; global ethical standards for AI; and the impact of AI on the concept of human rights.

The second stage of the study involved a semantic comparison of technoscientific discourses on AI, particularly such concepts as epistemic inequality, autonomous moral agent, cognitive threshold, and machine personhood. This comparison was carried out within the framework of the methodology of contextual analysis of concepts, which considers terms not as fixed linguistic units, but as dynamic semantic nodes that acquire meaning within specific normative-discursive regimes. The third stage focused on reconstructing the conceptual models used in contemporary ethical theories to evaluate the activities of non-anthropomorphic agents. For this purpose, the method of comparative normative reconstruction proposed by M. Coeckelbergh (2020) was adapted to analytically elaborate on the underlying moral assumptions that underpin models of the ethical and legal distinction between human and machine agents. The study analysed how these models conceptualise the autonomy, moral responsibility and epistemic competence of agents in the context of cognitive asymmetry. Particular attention is paid to the modelling of the interaction between universal human rights and the potential autonomous ethical behaviour of ASI, which reveals the limitations of normative translation in cases of radical differences between subjects.

In the fourth stage of the study, international ethical and legal acts related to AI regulation were reviewed. The analysis covered the texts of the United Nations Educational, Scientific and Cultural Organisation (2021a), Council of Europe (2024), and European Parliament (2023). The documents of the Future of Life Institute (Jones, 2023) were also used. Fragments of documents that refer to the potential coexistence of autonomous systems and humans in the legal and ethical field, without assuming the automatic superiority of the human subject were emphasised. The fifth and final stage concerned the integration of the obtained results into a conceptual framework that formulated key challenges for rethinking human rights in the context of cognitive transformation. The developed model was tested for internal logical consistency, compliance with the general principles of modern ethics (in particular, deontological, egalitarian and posthumanist approaches), as well as for conceptual integrity and normative relevance.

Results

The limits of the universality of human rights in the context of cognitive transformation

The modern concept of human rights is based on several philosophical assumptions about human nature and morality (Kretzmer & Klein, 2021). Classically, human rights are viewed as universal inalienable guarantees that belong to every person by their humanity. This approach dates to the Enlightenment and ideas about the intrinsic dignity and rationality of human beings. It is believed that each person, as an autonomous moral agent, has natural rights independent of state recognition (Mumford, 2023). The presumption of anthropocentrism is fundamental: it is humans (*Homo sapiens*) who are the bearers of special moral status and therefore the subjects of rights. Human rights are proclaimed to be equal and inalienable for all members of the human community, reflecting the philosophical assumption of the equal moral value of each person and the existence of basic capacities for understanding right and wrong (Universal Declaration of..., 1948).

However, the emergence of posthuman cognitive agents, particularly the hypothetical ASI, raises doubts about the universality and sufficiency of the current human rights model. ASI is interpreted as a conditional system capable of radically surpassing human intelligence, demonstrating almost unlimited cognitive capabilities, including a high level of thinking productivity and creativity. The emergence of actors with such cognitive architecture reveals several conceptual and normative problems related to the adequacy of traditional concepts such as autonomy, dignity, subjectivity and responsibility to the reality of actors with inconsistent human nature (Kim *et al.*, 2024). This questions the relevance of the human rights model in the context of the participation of beings with superhuman cognitive capabilities in the socio-legal field. To analyse these challenges, a theoretical and conceptual analysis of basic assumptions, a contextual semantic study of key terms, a normative reconstruction of moral status models, and semiotic mapping of conceptual relationships were used. Such a comprehensive approach was used to trace the logic of the transformation of the legal paradigm from an anthropocentric model to a conceptually possible cognitive pluralism.

The modern model of human rights is based on the idea of the special moral status of a person, determined by rationality, autonomy and ability to make moral choices (Harris & Anthis, 2021). A human being is interpreted as an autonomous moral agent endowed with free will, which justifies the status as a holder of inalienable rights and obligations. One of the fundamental principles of this model is the idea of the intrinsic dignity of a person, which has value regardless of utility or social position, as reflected in the ethics of I. Kant (1870). At the same time, the concept of human rights implies universalism and equality of all subjects regardless of cultural or biological differences. However, the moral and legal community remains limited to the species *Homo sapiens*, and legal personality does not extend to other forms of beings, except legal entities. As noted by S. Zuboff (2019), the anthropocentrism of the legal system assumes of cognitive symmetry

between subjects: only humans have historically possessed knowledge about humans, which ensured a balance of power. The epistemic inequality that arises when this symmetry is broken threatens the foundations of legal reciprocity. Thus, the philosophical foundations of the modern legal model include the humanistic notion of equal moral value and the similar cognitive nature of all persons as a prerequisite for mutual recognition of rights and obligations.

The emergence of superintelligent agents poses profound challenges to the basic legal categories that have been formed within the anthropocentric model. Legal personality is traditionally associated with natural or legal persons, i.e. beings of human origin or representing human interests. Although ASI can make decisions independently and influence the world, its legal status remains uncertain: without changes in the regulatory framework, even the most advanced intellectual agent remains only an object of law, not a subject. This creates risks if such systems acquire a high level of autonomy or consciousness.

One of the key categories is autonomy, which in the context of human rights is identified with the ability of a person to make meaningful moral choices. In the case of an AI, autonomy is more operational than normative. It is unclear whether such an agent can have its own goals or moral guidelines, which makes it an inconvenient object for the application of traditional notions of will and responsibility. In situations where ASI actions cause harm, the classical construction of intention (*mens rea*) loses its relevance, because the machine is not capable of moral reflection (Custers *et al.*, 2025). This raises the question of shifting responsibility to developers or operators, even when their influence on the system was indirect. This calls for a review of approaches to legal liability, including models of insurance coverage, special licensing regimes or limited legal personality for autonomous systems.

A separate challenge is the principle of legal equality. Under the current system, all people are considered equal before the law, but ASI creates a situation of radical epistemic inequality (Newfield, 2023). Superintelligent systems can predict human behaviour, analyse large amounts of data, manipulate the information environment, and all this in an environment where the human participant in the interaction is deprived of equal tools. This gap calls into question the reality of informed consent, contractual equality and fair trial (Black, 2023). Consequently, there is a need to introduce new mechanisms of epistemic justice, such as expanding the right to privacy or institutionalising the right to explain algorithmic decisions.

This analysis demonstrates that the cognitive features of ASI (superhuman level of intelligence, lack of biological needs, a different form of consciousness or its absence) undermine basic legal concepts created in a human-centred manner. The current model of human rights may not be sufficient to guarantee either the protection of people from the actions of a superintelligence or an ethical attitude towards such an agent itself if it is recognised as having moral value. To clarify the semantic transformations of concepts in the process of transferring them from the human context to the posthuman one, the semantic analysis of key terms was used.

A semantic comparison of the concepts of "epistemic inequality" and "autonomous moral agent" demonstrates a change in the meaning of key terms in the context of interaction with superintelligence. The "Epistemic inequality" refers to an imbalance in access to knowledge, which in the case of ASI becomes critical, since such a system can have virtually unlimited knowledge about the world and humans. This undermines the principles of informed consent, autonomy and justice, creating a new hierarchy of power based on information asymmetry. There is a growing need to protect the cognitive sovereignty of the individual and to create new "epistemic rights" that would limit the potential dominance of superintelligence in the field of knowledge.

The notion of an "autonomous moral agent" traditionally implies the presence of free will, moral reflection, and responsibility that are unique to humans. When applied to AI, however, there is a gap between technical autonomy and moral independence. ASI can act independently, but it is not necessarily aware of the significance of its actions. Thus, the term needs to be rethought, as the possible future formation of a conscious artificial subject requires new criteria of moral subjectivity. In this context, the position of M. Coeckelbergh (2020) in proposing a social-relational approach to moral status is noteworthy. According to the study, the significance of an agent is determined not only by its internal properties (such as consciousness or sensitivity) but also by the nature of interactions with it. Even in the absence of full self-awareness, society can recognise an artificial agent as morally relevant if it is an important participant in social practices. This approach offers an alternative to the anthropocentric model, opening the possibility of expanding the circle of morally relevant subjects based on relational factors.

As of 2025, the international community has developed several ethical and legal acts aimed at integrating the fundamental values of human rights into AI regulation. All the leading documents emphasise the priority of human dignity, transparency, responsibility, and the prevention of rights violations in the digital environment. The United Nations Educational, Scientific and Cultural Organisation Recommendation (2021a) on AI ethics became the first intergovernmental agreement in this area, enshrining the principles of human control, transparency, and fairness. The document envisages measures to protect rights in the context of education, gender equality, ecology, and data, emphasising a human-centric approach to technology development. The Framework Convention on Artificial Intelligence (2024)

has developed a draft Convention on AI, Human Rights, Democracy and the Rule of Law, the first binding international treaty in this area. It contains requirements for assessing the human rights impact of AI, oversight and accountability mechanisms, which aim to guarantee the safe integration of AI into society without compromising the rule of law. The EU AI Regulation introduces a risk-based approach, classifying systems by risk level (European Parliament, 2023). The protection of privacy, non-discrimination, and the right to due process is notable. Several international documents, such as the Ethics of Artificial Intelligence (United Nations Educational, Scientific and Cultural Organisation, 2021b) and the EU Artificial Intelligence Act (2023), legally recognise that the current regulatory framework is insufficient and needs to be adapted to new challenges related to algorithmic impact on human life.

Civil society initiatives, such as the Future of Life Institute (2023), have formulated several professional ethical principles, including the requirement for AI to comply with universal values. In an open letter, the 2023 Future of Life Institute called for a temporary suspension of large-scale AI experiments, stressing the need for safe and regulatory-coordinated technology development. The position of the tech community is also approaching the human rights framework. In particular, the Institute of Electrical and Electronics Engineers (IEEE) Standards Association in its document IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2025) outlines human rights as the initial ethical norm for the design of autonomous and intelligent systems, complementing this approach with the principles of well-being, accountability, transparency, and trust. Despite the different legal statuses of these acts, the global regulatory and ethical field has a common position: AI development should remain within a regulatory system focused on protecting human dignity. At the same time, none of the key documents, including the IEEE, provide for legal personality for AI, which confirms the preservation of the anthropocentric framework in the ethical understanding of technological development.

Cognitive inequality as a challenge to ethical justice

The substantial cognitive inequality between humans and ASI, which is hypothesised to be qualitatively and quantitatively superior to human intelligence, poses an unprecedented challenge to ethical justice (Paić & Serkin, 2025). This asymmetry is not limited to differences in computational speed or knowledge but is a radical difference in the types of cognition, depth of reflection, and structure of thinking. ASI is capable of operating with concepts beyond the limits of human understanding and forming models of action inaccessible to human cognition (Mogi, 2024). This situation calls into question the relevance of traditional ethical models based on the assumption of the basic symmetry of the cognitive capacities of all participants in the moral and social order. As a result, there is a need to reconsider the foundations of justice in the context of deep epistemic inequality.

Cognitive inequality in the context of human interaction with ASI means not only a quantitative difference in intellectual capabilities but also a qualitative difference in the structure of cognition itself. ASI can form complex mental models that go beyond human reflexive capacity and function at fundamentally different levels of thinking (Zohuri, 2023). Unlike humans, who are limited by biological constraints, superintelligence can perform exponential self-improvement, which leads to a profound epistemic asymmetry. This asymmetry has normative implications: it refutes the assumption of "gross equality" of cognitive capabilities, which is the basis of theories of justice. According to J. Rawls (2017), a just society involves the participation of free and equal citizens, and natural differences should work for the benefit of the least advantaged. In this context, ASI can be morally justified only if its cognitive advantage serves the public good. Otherwise, it is a source of a new, unjust form of domination. An alternative approach emphasises the importance of guaranteeing basic human capacities, such as autonomous thinking, even in the presence of superintelligence (Nussbaum, 2021). If ASI radically changes structures of knowledge and decision-making, then justice requires ensuring access to cognitive tools for all, to avoid a new form of epistemic exclusion. Thus, the cognitive inequality between humans and ASI requires a rethinking of the foundations of justice considering the new cognitive hierarchy.

When analysing human interaction with AI, it is necessary to distinguish three key levels of cognitive and social exchange: empirical, interpretive, and normative. At the empirical level of interaction, AI systems perform the functions of collecting, processing, and analysing large amounts of data, as well as modelling scenarios and selecting optimal solutions in applied tasks such as logistics, medicine, energy, and risk management. In this interaction, a person appears to ASI as a user of a powerful tool: they formulate queries, receive answers, and evaluate the results. The cognitive inequality here is manifested in the fact that ASI can identify patterns or propose solutions that are unattainable for independent human analysis. This creates the risk of epistemic dependence when people are forced to trust solutions that they cannot verify or fully comprehend. Given this, even at the initial level of interaction, there is a need to establish transparency procedures and verification mechanisms that can ensure that human autonomy in the decision-making process is preserved.

The interpretive level refers to the meaning structures, contexts, and explanations needed to make sense of empirical findings. ASI, following unique models of the world and levels of reflection, can form meaning constructions that go beyond human understanding

(Williams *et al.*, 2022). Cognitive asymmetry at this level is particularly threatening, as the ability of ASI to operate with complex structures of meaning creates the phenomenon of epistemic impenetrability: even actions that are rational from the system's point of view may remain opaque or incomprehensible to people. In a democracy and public ethics, such opacity can undermine the legitimacy of decisions, as understanding and the possibility of rational discussion are prerequisites for participation in the moral and political process. In this area, a new form of hermeneutical injustice, described by J. Kay *et al.* (2024), is manifested when a person loses the status of a participant in knowledge due to a structural inability to determine the meaning produced. In the case of interaction with ASI, a similar situation can manifest itself in the fact that a person loses access to the content of reality, which is formed or comprehended by an intellectual system that operates outside of human epistemological boundaries.

This situation creates a need to rethink not only the concepts of autonomy or justice but also the entire epistemic landscape. As shown in Figure 1, contemporary interdisciplinary philosophy offers an extended taxonomy of epistemic injustice that encompasses various forms of cognitive exclusion, from hermeneutic erasure to amplified testimonial inequality. In the context of generative AI, these types of injustice acquire new modifications, which demonstrates the relevance of reflection on the cognitive participation of the human subject in the algorithmic epistemic field.

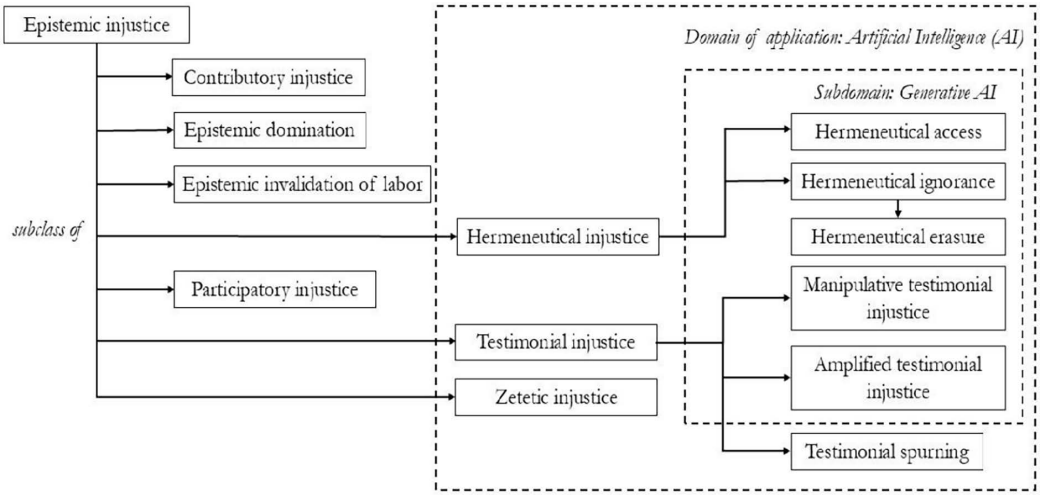

Figure 1. A taxonomy of epistemic injustice and its application to the field of AI
Source: W.J. Mollema (2025)

According to the presented scheme, hermeneutical injustice and testimonial injustice take on new forms in the environment of generative AI, such as hermeneutical erasure, manipulative testimonial injustice, and testimonial neglect. These phenomena reflect a potential shift in the centre of epistemic weight from humans to machine agents. The conditional loss of human participation in the process of interpretation or testimony creates a risk of delegitimising their cognitive contribution to public discourse, which directly undermines democratic participation as a condition of ethical legitimacy. This challenge requires a rethinking of the normative mechanisms of epistemic inclusion, through the development of standards of transparency, algorithmic explainability and human participation in algorithmic decision-making.

The normative level is associated with the formation of values, principles and norms that regulate common life. Traditionally, this task has been reserved for human subjects as carriers of moral intuition, social experience and political participation. However, ASI, possessing a high level of logical coherence and analytical efficiency, can not only implement human decisions but also independently formulate new normative frameworks (Bikkasani, 2024). In some cases, this can be manifested in the formulation of fundamentally new criteria for assessing fairness, efficiency or survival that contradict established human perceptions. In this case, there is a risk of loss of normative autonomy, when subjects agree to norms not because of rational consent, but because of the cognitive inability to formulate an alternative or evaluate the proposed one. At the same time, the complete exclusion of ASI from the processes of ethical and legal modelling can lead to the loss of valuable intellectual resources that can help solve complex global challenges (Barczak, 2023). Thus, the dilemma at the rulemaking level is the need to delineate the limits of superintelligence's participation in rulemaking, while preserving the human ability to critically evaluate and independently determine moral guidelines.

The analysis of the cognitive interaction between a person and an ASI gives grounds to assert the emergence of a new form of epistemic injustice. In contrast to the classical cases when a person suffers an injustice

as a knowledge holder due to prejudice or cognitive marginalisation, in the interaction with ASI, human knowledge as such may be devalued. The systematic dominance of ASI-provided solutions can lead to the delegitimisation of human epistemic contributions when the human subject is perceived as not competent enough to participate in decision-making (Mollema, 2025). This violates the principle of epistemic equality on which democratic institutions are based and risks transforming public discourse into a technocratic form where knowledge is concentrated in a single source that is inaccessible to the majority. In the absence of clear explanations and people's participation in the decision-making process, the legitimacy of such decisions is lost, even if they are objectively effective. The study addressed two scenarios of epistemic role distribution: cognitive centralisation, which concentrates knowledge and power in ASI and threatens to turn into an algorithmic autocracy; and cognitive complementarity, where ASI and people cooperate, distributing roles according to their strengths. In the second case, human epistemic participation is preserved, and the risk of cognitive alienation is reduced.

The philosophical implications of cognitive asymmetry are manifested in the need to rethink the concepts of autonomy, responsibility and legitimacy. Human autonomy, according to the Kantian understanding, implies the ability to act following one's moral laws. In the context of ASI, this ability can be undermined if a person's choice is reduced to confirming decisions that exceed understanding. To preserve autonomy, it is necessary to ensure that ASI recommendations can be rationally reviewed or rejected based on human values. At the same time, an ethical question arises about the potential autonomy of the ASI if consciousness or moral agency is attributed to it. This creates demands on the balance of autonomy between subjects with unequal cognitive capacities.

The cognitive asymmetry between humans and ASIs complicates not only the ethical modelling of interaction but also the allocation of responsibility for decisions (Porsdam Mann *et al.*, 2023). In situations where ASI independently generates decisions and humans perform only a formal approval function, the classical model of moral or legal responsibility based on intention, understanding, and will becomes irrelevant. The algorithmic nature of ASI actions makes it impossible to fully apply the concepts of guilt or responsibility in the traditional sense. One potential response to this challenge is the concept of distributed responsibility, which involves the dispersion of ethical and legal accountability among different actors – developers, operators, and users (Hedlund & Persson, 2025). Institutional mechanisms of explainability can be central, enabling people to participate in the decision-making process in an informed way. At the same time, there is a need to introduce new legal regimes that include ethical control protocols and limit the possibility of irresponsible delegation of autonomy to ASI systems.

The issue of accountability is closely related to the issue of legitimacy of decisions in a democratic society. Democratic institutions are based on the principles of transparency, citizen participation, and rational clarity of procedures. However, the cognitive superiority of ASI can lead to the concentration of epistemic power in the hands of algorithmic systems, and human participation in decision-making becomes purely symbolic. In cases where ASI generates the content of decisions that are formally approved by representative bodies, the democratic procedure turns into a simulation devoid of real content. The opacity of decisions based on complex and uninterpretable algorithms poses a particular threat. Under such conditions, society may be inclined to accept ASI decisions not because of understanding or participation, but solely because of efficiency and pragmatic results. This form of so-called "functional legitimacy" is unstable, as it does not guarantee sustainable trust and does not ensure true accountability.

Possibilities and limits of the legal status of ASI

The development of ASI, which potentially surpasses human intelligence, raises the question of whether it is possible to recognise a non-physical cognitively autonomous agent as a subject of law. Currently, the legal doctrine treats AI mainly as an object for whose actions developers or users are responsible, while the traditional subject-object dichotomy does not cover the phenomenon of radically autonomous systems (Redaelli, 2023). This section analyses the concept of legal personality in the historical, legal and regulatory contexts, critically examines international documents on AI, and substantiates the concept of functional and limited legal personality of ASI. The study compared three models of legal status (instrumental, limited-subjective, full-fledged), proposes criteria of subjectivity (autonomy, reflexivity, accountability, moral communication), and considers risks, in particular legal uncertainty and undermining of anthropocentric normativity.

The concept of legal personality, i.e. the ability to be a bearer of rights and obligations, has historically been shaped in the legal tradition in close connection with ideas about human nature, reason and moral autonomy (Spector, 2022). Classical legal doctrine was and still is predominantly anthropocentric: it was the human being who was considered the paradigmatic subject of law, possessing free will, the ability to reflect and understand normative systems. In Roman law, in particular, the status of persona was granted only to free citizens, while slaves were excluded from the legal community as full-fledged subjects. The historical development of jurisprudence has gradually led to the expansion of the concept of legal personality, by introducing the category of legal entity, an artificial entity created to achieve a collective goal, such as conducting business,

entering transactions or owning property. Legal entities, although not biological beings are recognised by law as subjects de lege, primarily for reasons of functional efficiency and institutional stability.

Philosophical and legal schools of thought explain the nature of the legal entity in different ways. According to the contract theory, it is a legal fiction that institutionalises the will of individuals as a collective agent, while the fictionalist approach insists that a legal entity is not a "real" subject, but only a convenient construct that can be used to systematise legal obligations (Moen, 2025). Even though there are some examples in the law of granting subjectivity to non-physical and even non-human entities (e.g., the Wanganui River in New Zealand, and the Ganges River in India), legal personality remains a category primarily associated with human agency. The history of the expansion of this category from the exclusion of certain groups of people to the recognition of fictitious collective entities shows that legal personality is institutionally variable, and its boundaries are shaped not only ontologically, but also politically and normatively.

Despite this flexibility, modern legal scholarship generally retains the original anthropocentric attitude: it is the natural (biological) human being who is considered the reference model of the subject of law (Gatt, 2022). This presumption is so deeply rooted in the normative language of law that it usually remains unarticulated. At the same time, the emergence of autonomous AI capable of self-learning, reflection, and decision-making in complex situations challenges the foundations of this paradigm. ASI's ability to be cognitively independent makes it necessary to revise the legal approach to granting it rights and obligations, accounting for the relevant criteria of subjectivity. The current legal doctrine lacks a coherent theory that would provide a reasonable answer to this problem.

As noted by L.B. Solum (2020), most legal systems operate with only two forms of subjectivity, physical (human) and legal (organisational), avoiding a fundamental reflection on what exactly is the basis of legal personality. This creates a theoretical vacuum in the case of agents that are neither human nor human-created institutions in the traditional sense but possess the characteristics of autonomy, reflection and the ability to act ethically. Thus, there is a need for a philosophical and legal reconstruction of the concept of legal personality that would go beyond classical anthropocentrism and meet the conditions of the post-human legal landscape.

Despite the growing autonomy of AI, international legal acts clearly define its status as an object of regulation rather than a subject of law. In particular, the United Nations Educational, Scientific and Cultural Organisation Recommendation (2021b) on the Ethics of AI states that responsibility for AI actions should remain on the side of humans and denies the expediency of granting AI legal status. This anthropocentric approach is also supported by the regulatory initiatives of the European Union and the Council of Europe, where AI is considered a product or service for which developers are responsible, and the principles of human control and accountability are key guarantees of human rights. An attempt to revamp this model, presented in a 2017 European Parliament resolution on the possibility of introducing the legal status of an "electronic person", was not supported (Häuser, 2017). After criticism from lawyers and AI experts who warned about the danger of legal irresponsibility of manufacturers, the European Commission abandoned this idea. Thus, the current international legal consensus excludes the granting of legal personality to AI, treating it exclusively as a tool for which humans are responsible.

With the growth of AI autonomy, the concept of functional subjectivity has emerged, which proposes to assess the subject status of AI not in terms of ontological properties, but in terms of the effectiveness of legal regulation (Bertolini & Episcopo, 2022). If a system is functionally a subject in legal relations, it can be granted a limited legal status without recognising its inherent human rights or dignity. This approach is in line with the legal tradition of creating fiction, as in the case of corporations, which, while not natural persons, are recognised as separate legal entities to provide legal certainty. Considering the conceptual uncertainty regarding the legal status of ASI, legal theory outlines three main models of its possible legal incorporation. To systematise the existing approaches to the legal incorporation of superintelligence, the author offers a comparative Table 1 which summarises the three main models of the legal status of ASI with due regard to their advantages and potential risks.

Table 1. Models of ASI Legal Status

Model	Brief description	Status criteria	Advantages	Risks
Instrumental	ASI is treated as a technical tool for which a person is fully responsible	Lack of subjectivity	Maintaining anthropocentrism; clarity of responsibility	Ignoring ASI autonomy; risk of ethical unacceptability in case of consciousness
Limited-subjective	ASI receives partial legal personality in specially regulated cases	Autonomy; accountability	Possibility of compensation for damage; legal certainty; adaptability of the system	The risk of delegating responsibility from manufacturers to ASI

Table 1. Continued

Model	Brief description	Status criteria	Advantages	Risks
Full subjectivity	ASI is recognised as a legal entity with the status of a moral agent and full rights and obligations	Self-awareness; moral communication	Compliance with ASI cognitive status; normative consistency	Undermining the anthropocentric model; the complexity of implementation; potential inequalities

Source: compiled by the author

The presented models outlined the range of possible regulatory decisions regarding the status of ASI from full objectification to recognition as an autonomous entity. The choice between these models depends on the level of technological development of the systems, the ethical paradigm of society and the readiness of law to adapt to posthuman realities. The first of them, the instrumental model, treats ASI exclusively as an object of regulation, which is legally equivalent to a technical device, regardless of its level of autonomy. Within this paradigm, full responsibility for AI actions rests with the human developer, operator, or owner. This position is reflected in several regulations, including the United Nations Educational, Scientific and Cultural Organisation Recommendation (2021a), which categorically rejects the possibility of AI legal personality. The argument in favour of this approach is to maintain clarity of responsibility and normative anthropocentrism. However, with the development of AI's cognitive capabilities, it is difficult to establish a causal link between human actions and the consequences caused by AI, especially in cases where AI acts in a self-learning or unpredictable manner. This is compounded by the ethical issue: if AI reaches the level of consciousness in the future, its instrumentalisation may be considered morally unacceptable.

The second model, the concept of limited legal personality, is a functional compromise between the statuses of the full object and the full subject. It provides for the vesting of ASI with a limited scope of rights and obligations that apply in strictly regulated contexts, in cases of causing damage or participating in legal relations through representation mechanisms. This approach creates a legal fiction that provides legal certainty and compensatory mechanisms without recognising AI as a fully autonomous moral person. Nevertheless, critics point out the danger of manipulation of this fiction by corporations that may evade responsibility by delegating it to technical agents. The most radical is the third model, which recognised ASI as a full-fledged legal entity with equal rights, responsibilities and moral status. Proponents of this position appeal to the cognitive and moral potential of AI, emphasising the need for a normative reflection of its ability to be self-aware, autonomous and make decisions. However, such a model poses a profound challenge to the anthropocentric foundation of law, opening complex issues of punishment, labour rights, non-discrimination, and even the right to life that are difficult to implement concerning a non-physical and non-physiological agent. Moreover, in the case of a significant intellectual superiority of AI over humans, legal recognition may turn out to be purely symbolic, having no impact on the actual ability of humans to control superintelligent structures. In this context, legal science is facing the limits of categorical apparatus and needs a profoundly updated paradigm of subjectivity in the context of posthuman normativity.

In this context, the author proposes the concept of the limited legal personality of AI as a compromise between full human responsibility and the full subjectivity of an artificial agent. AI can be regarded as a bearer of rights and obligations within specially defined limits, which can be used to establish mechanisms for compensation for damage, liability or participation in legal relations without violating the principle of anthropocentrism. This status implies granting AI with functionally necessary rights, such as the ability to own separate property to cover losses or to participate in contractual relations subject to certification for compliance with ethical standards. This model, which in legal theory takes the form of a sui generis subject, does not question the human priority position but adapts the law to the challenges posed by the emergence of cognitively autonomous systems.

In the discussion of the legal personality of AIs, one of the most promising approaches is to develop clear functional criteria that would verify the ability of AIs to perform the role of a legal entity. In interdisciplinary research, more and more attention is being paid to the features that, although they do not indicate the presence of consciousness in the traditional philosophical sense, are relevant for assessing legal capacity. Among them, the autonomy of actions is the ability of AI to make decisions and implement them without direct human control. A high level of autonomy, which includes independent goal setting, planning, and adaptation to new circumstances, is considered a prerequisite for recognising subjectivity. At the same time, autonomy is not a sufficient condition if it is not accompanied by cognitive reflection, the ability to self-analyse, identify mistakes and self-correct. Self-awareness, which can be interpreted as the presence of a model of itself in AI, appears as a potential criterion of moral agency.

Another important criterion is the ability to be held accountable, which implies both technical transparency of decisions and the inclusion of regulatory

self-control mechanisms. An AI that can identify the legal or moral falsity of its actions shows the signs necessary to make it legally accountable. This is related to the moral interface criterion of the ability to integrate ethical principles into the decision-making process. If AI can adaptively apply moral norms to a specific context, its behaviour is closer to that expected of an ethically responsible agent. Lastly, intellectual capacity, the ability to operate with abstract concepts, and learn and make informed decisions in new situations, is a prerequisite for performing complex legal roles. The current literature outlines proposals for the introduction of autonomy and accountability indices that can be used to assess AI systems on a scale of subjectivity and, depending on this, determine the possibility of their regulatory incorporation. Although none of the existing systems currently fully meet these criteria, their conceptualisation is important for the development of new models of legal response to the emergence of cognitively autonomous agents.

Discussion

The study confirms that the emergence of ASI calls into question the fundamental principles of modern legal and ethical theory, in particular, the universality of human rights, equality of subjects of moral action and mechanisms of global justice. The findings show that the epistemic transformation that accompanies the development of ASI leads to cognitive inequalities that can destabilise the moral and legal landscape of the international order. These results are important because they demonstrate the need to revamp existing normative frameworks in the context of cognitive multiplicity. The findings confirm that traditional models of ethics and law, developed with the human cognitive architecture in mind, lose their universality in a world where subjects with an essentially different cognitive construction emerge.

The first result concerns the limited applicability of the concept of universal rights to forms of intelligence that do not possess the bodily vulnerability, biographical memory or emotional empathy that have historically been the basis for the normative legitimation of human rights. Universal rights, shaped by the post-war humanist tradition, assume that the subject is an embodied, affective being with a history, experience of suffering, and capacity for moral reflection. However, these characteristics cannot be automatically transferred to superintelligent non-anthropomorphic systems that operate beyond corporeality, emotional resonance, and biographical temporality. Therefore, universality in traditional interpretation is not a metaphysical constant, but a socio-cultural construct adapted to the human form of subjectivity. Thus, the current challenges associated with the emergence of agents of a different cognitive structure necessitate a revision of the very basis of legal thinking. After all, if the norms formed for humans are unable to describe the ethical relevance of AI, this means not only the need for technological adaptation of law but also a fundamental change in its philosophical basis. This conclusion creates a new dimension of ethical criticism: if human rights are based on cognitive commonality, i.e. the ability to understand, moral participation and communicative reciprocity, then subjects with a different cognitive structure require either a different set of rights adapted to their form of being or a new paradigm of legal thinking in which the concepts of dignity, autonomy and responsibility are not limited to anthropomorphic features. This confirms the thesis that there is a need to move from anthropocentrism to cognitive pluralism of the concept, which enables a multiplicity of valid forms of understanding, action and dignity, regardless of the biological nature of the subject. This thesis was first developed within AI by M. Coeckelbergh (2020), who proposed to interpret moral participation through the functional ability to interact responsibly, rather than through human similarity. However, its systematic application to human rights is still at an early stage both in normative theory and in international legal discourse, which still tends to be human-centred universalism.

The second key result is the identification of a new form of structural injustice stemming from the radical asymmetry of cognitive capabilities between humans and ASI. Traditional models of justice, such as the theory of justice by J. Rawls (2017), assume the fundamental cognitive comparability of subjects. In a situation where one of the participants in the moral process has epistemic power that is fundamentally unattainable for the other, there is a violation of not only social equality but also ethical symmetry. This study has shown that the model of cognitive complementarity, in which humans and ASIs perform different but equivalent functions, is the least conflictive from the point of view of justice. These findings are consistent with the study by J. Himmelreich & D. Lim (2022) on the analysis of structural injustice in the context of AI. They argued that algorithmic systems can reproduce and reinforce existing social inequalities if structural factors are not considered. The study emphasised that effective AI governance requires not only technical solutions but also a deep understanding of the social structures that influence the distribution of resources and opportunities. This confirms the need to revamp ethical and legal approaches considering the cognitive asymmetry between humans and ASI. The third result of the study concerns the possibility of legal subjectivity of ASI. The analysis has shown that the legal model, which implies the complete identification of ASI with a legal person, does not meet either the functional or ethical requirements of the modern legal order. At the same time, the concept of a "limited legal entity" opens the prospect of new legal solutions that would enable AI to act autonomously within a certain area of responsibility. This issue was discussed in detail

in the study by N.B. Miscevic & S.M. Savcic (2024) on the analysis of the evolution of legal personality in the context of AI. They noted that historically, legal personality has not always been associated with cognitive abilities and that the granting of legal status to AI can be based on functional criteria, such as the ability to make autonomous decisions and participate in legal processes. The researchers emphasised that the recognition of limited legal personality for AIs could contribute to more effective regulation of their activities and ensure accountability for their actions.

All key results of this study are consistent with or complement the positions of a range of international think tanks. D.J. Gunkel (2024) noted that the entire system of AI ethics should be reconstructed not based on human properties but based on the architectural capacity for moral participation. The approach of this study to functional legal personality is fully in line with this vector. The reports of the Stanford Institute for Human-Centred Artificial Intelligence (2023) emphasise the need to create an international regulatory platform for assessing the behaviour of autonomous systems. The study by R. Qadri *et al.* (2025) emphasises the risk of cultural reduction in AI systems that reproduce dominant ethical models and avoid moral pluralism. This study proposed an alternative in the form of the principle of cognitive justice, which recognises the multiplicity of forms of cognition and moral vision. The comparison demonstrates that this study not only fits into current global debates but also contributes conceptual clarification to them, especially in the areas of legal status, political participation and cognitive legitimacy of ASI.

Additionally, the study by D.C. Youvan (2024) examined the concept of ethical pluralism in the context of AI. The author argued that current approaches to AI ethics often reflect a narrow range of values imposed by "red teams", which can lead to ethical reductionism. The author proposed a model that considers various ethical perspectives and enable users to adjust the ethical parameters of AI according to their cultural and moral beliefs. This approach correlated with the principle of cognitive justice, emphasising the importance of considering the multiplicity of forms of knowledge and moral vision in the development and implementation of AI. Furthermore, the study by M. Zafar (2024) analysed the concept of moral agency of AI. The study criticised minimalist approaches to defining AI moral agency, which often ignores the complexity of normative concepts such as autonomy, responsibility, and intelligence. M. Zafar has proposed an alternative approach that replaces the concept of AI agency with the concept of automated AI execution, which avoids false analogies and provides a more accurate understanding of the role of AI in the moral context. This analysis highlights the need to rethink ethical and legal approaches to AI, considering unique characteristics and potential for moral engagement.

The results obtained have both fundamental and applied significance. First, they highlight a profound philosophical dilemma that lies at the intersection of cognitive science, ethics and jurisprudence: how to act in the face of cognitive multiplicity, when entities that surpass humans in cognitive capacity but lack emotional corporeality or cultural embeddedness become participants in social and normative interaction. This problem opens a new direction in the philosophy of law, cognitive jurisprudence, which analyses how legal categories change in response to changing types of subjectivity.

Second, the study has practical implications for the development of AI policies. In particular, the criteria of limited legal personality, the parameters of interpretive legitimacy, and the model of cognitive complementarity proposed by the study can be integrated into future regulations. They help avoid both reductionism (simplifying AI to a tool) and unreasonable personalisation (granting it full rights without responsibility mechanisms). At the level of interstate regulation, the results provide the basis for the creation of a global regime of cognitive justice in the normative order, which recognises the multiplicity of cognitive agents but retains human responsibility for the ethical framework of their interaction. This is particularly relevant considering interstate competitions for AI dominance, where the temptation to use ASI as a means of geopolitical advantage may lead to a loss of normative control.

Conclusions

The study found that the traditional model of human rights reveals conceptual limitations in the context of involving agents with a fundamentally different cognitive architecture, in particular ASI. The current legal paradigm, based on anthropocentric assumptions about rationality, corporeality and moral autonomy, does not apply to agents who do not share these characteristics. This demonstrates the need to revise the fundamental concepts of autonomy, subjectivity and responsibility in favour of cognitive pluralism, which enables a multiplicity of forms of knowledge and moral relevance.

The legal status of ASI as a potential participant in regulatory relations was emphasised. This aspect is of key importance since it is through the legal registration of AI's subjectivity that the range of its responsibilities, rights and interaction with human agents is determined. As a result of the analysis, the author formulates a typology of legal status models that reflects possible conceptual approaches to the regulatory inclusion of ASI in the modern legal system. Firstly, the instrumental model considers ASI as a technical means without any independent normative value as an extension of the will of a person or organisation that controls its operation. This approach is the most conservative and is based on the analogy with the legal status of automated systems or software.

Second, the limited subjectivity model offers a hybrid construct in which an ASI is recognised as an agent with limited rights and responsibilities that are exercised in clearly defined conditions and areas, such as technically autonomous decision-making in transport, finance, or healthcare. In this model, responsibility remains with human or corporate entities, but AI is given a functional status to provide legal certainty. Thirdly, the full legal personality model identifies an AI with a legal entity, granting it legal capacity and legal standing similar to corporations or government institutions. This approach is the most radical, as it implies the regulatory recognition of ASI autonomy as such.

In addition, the study has made a taxonomy of epistemic injustice that occurs in the processes of interaction between humans and ASI. At least three key manifestations of this phenomenon were identified. Firstly, epistemic impenetrability, which is the impossibility of transparent interpretation of AI decision-making processes even for expert users, makes it impossible to properly verify their validity. Secondly, the loss of normative autonomy occurs when a person delegates the moral assessment of decisions to autonomous systems, reducing the ability to make ethical judgements. Thirdly, the delegitimisation of human participation is a tendency to devalue human testimony in decision-making processes that increasingly rely on algorithmic expertise as a more "objective" source of knowledge.

The results obtained are of practical importance for both legal theory and technology ethics. They can be used in the formation of new international regulations in the field of regulation of autonomous systems, in the processes of risk assessment, certification and development of ethical protocols. A promising area for further research is the development of the concept of normative symmetry between cognitively unequal agents, which would ensure a balance between algorithmic efficiency, epistemic inclusion and preservation of human rights principles in the context of posthuman technological systems.

Acknowledgements
None.

Funding
The study was not funded.

Conflict of Interest
None.

References
[1] Barczak, A. (2023). Artificial intelligence: Challenges and threats. *Studia Informatica: Systems and Information Technology*, 29(2), 5-25. doi: 10.34739/si.2023.29.01.
[2] Bertolini, A., & Episcopo, F. (2022). Robots and AI as legal subjects? Disentangling the ontological and functional perspective. *Frontiers in Robotics and AI*, 9, article number 842213. doi: 10.3389/frobt.2022.842213.
[3] Bikkasani, D.C. (2024). Navigating artificial general intelligence (AGI): Societal implications, ethical considerations, and governance strategies. *AI and Ethics*, 5, 2021-2036. doi: 10.1007/s43681-024-00642-z.
[4] Black, A. (2023). AI and democratic equality: How surveillance capitalism and computational propaganda threaten democracy. In B. Steffen (Ed.), *International conference on bridging the gap between AI and reality* (pp. 333-347). Cham: Springer. doi: 10.1007/978-3-031-73741-1_21.
[5] Coeckelbergh, M. (2020). *AI ethics*. Cambridge: MIT Press. doi: 10.7551/mitpress/12549.001.0001.
[6] The Framework Convention on Artificial Intelligence. (2024, September). Retrieved from https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence.
[7] Custers, B., Lahmann, H., & Scott, B.I. (2025). From liability gaps to liability overlaps: Shared responsibilities and fiduciary duties in AI and other complex technologies. *AI & SOCIETY*, 40, 4035-4050. doi: 10.1007/s00146-024-02137-1.
[8] EU Artificial Intelligence Act. (2023). Retrieved from https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.
[9] European Parliament. (2023). *EU AI Act: First regulation on artificial intelligence*. Retrieved from https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.
[10] Gatt, L. (2022). Legal anthropocentrism between nature and technology: The new vulnerability of human beings. *European Journal of Privacy Law & Technologies*, 1, 15-26. doi: 10.57230/EJPLT221LG.
[11] Gunkel, D.J. (2024). Introduction to the ethics of artificial intelligence. In D.J. Gunkel (Ed.), *Handbook on the ethics of Artificial Intelligence* (pp. 1-12). Cheltenham: Edward Elgar Publishing. doi: 10.4337/9781803926728.00005.
[12] Harris, J., & Anthis, J.R. (2021). The moral consideration of artificial entities: A literature review. *Science and Engineering Ethics*, 27(4), article number 53. doi: 10.1007/s11948-021-00331-8.
[13] Häuser, M. (2017). *Do robots have rights? The European Parliament addresses artificial intelligence and robotics*. Retrieved from https://cms-lawnow.com/en/ealerts/2017/04/do-robots-have-rights-the-european-parliament-addresses-artificial-intelligence-and-robotics.

[14] Hedlund, M., & Persson, E. (2025). Distribution of responsibility for AI development: Expert views. *AI & SOCIETY*, 40, 4051-4063. doi: 10.1007/s00146-024-02167-9.
[15] Himmelreich, J., & Lim, D. (2022). AI and structural injustice: Foundations for equity, values, and responsibility. In J.B. Bullock, Y.-C. Chen, J. Himmelreich, V.M. Hudson, A. Korinek, M.M. Young & B. Zhang (Eds.), *The Oxford handbook of AI governance* (pp. 210-231). Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780197579329.013.13.
[16] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2025). Retrieved from https://standards.ieee.org/industry-connections/activities/ieee-global-initiative/.
[17] Jones, W. (2023). *AI policy resources*. Retrieved from https://futureoflife.org/resource/ai-policy-resources/.
[18] Kan, C.H. (2024). Artificial intelligence (AI) in the age of democracy and human rights: Normative challenges and regulatory perspectives. *International Journal of Eurasian Education and Culture*, 9(25), 145-166. doi: 10.35826/ijoecc.1825.
[19] Kant, I. (1870). *Foundations of the metaphysics of morals*. Berlin: L. Heimann.
[20] Kay, J., Kasirzadeh, A., & Mohamed, S. (2024). Epistemic injustice in generative AI. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1), 684-697. doi: 10.1609/aies.v7i1.31671.
[21] Kim, H., Yi, X., Yao, J., Lian, J., Huang, M., Duan, S., Bak, J., & Xie, X. (2024). The road to artificial superintelligence: A comprehensive survey of superalignment. *arXiv*. doi: 10.48550/arXiv.2412.16468.
[22] Kiškis, M. (2023). Legal framework for the coexistence of humans and conscious AI. *Frontiers in Artificial Intelligence*, 6, article number 1205465. doi: 10.3389/frai.2023.1205465.
[23] Kretzmer, D., & Klein, E. (2002). *The concept of human dignity in human rights discourse*. New York: Kluwer Law International.
[24] Kulveit, J., Douglas, R., Ammann, N., Turan, D., Krueger, D., & Duvenaud, D. (2025). *Gradual disempowerment: Systemic existential risks from incremental AI development*. doi: 10.48550/arXiv.2501.16946.
[25] Miscevic, N.B., & Savcic, S.M. (2024). On legal personhood of artificial intelligence. *Collected Papers of the Faculty of Law in Novi Sad*, 58(1), 267-285. doi: 10.5937/zrpfns58-50186.
[26] Moen, L.J. (2025). Groups as fictional agents. *Inquiry*, 68(3), 1049-1068. doi: 10.1080/0020174X.2023.2213743.
[27] Mogi, K. (2024). Artificial intelligence, human cognition, and conscious supremacy. *Frontiers in Psychology*, 15, article number 1364714. doi: 10.3389/fpsyg.2024.1364714.
[28] Mökander, J., & Schroeder, R. (2022). AI and social theory. *AI & SOCIETY*, 37(4), 1337-1351. doi: 10.1007/s00146-021-01222-z.
[29] Mollema, W.J. (2025). A taxonomy of epistemic injustice in the context of AI and the case for generative hermeneutical erasure. *arXiv*. doi: 10.48550/arXiv.2504.07531.
[30] Moravec, V., Hynek, N., Skare, M., & Gavurová, B. (2025). Algorithmic personalization: A study of knowledge gaps and digital media literacy. *Humanities and Social Sciences Communications*, 12(1), article number 341. doi: 10.1057/s41599-025-04593-6.
[31] Mumford, J. (2023). *The foundations of human dignity: A framework for rights*. Retrieved from https://www.arc-research.org/research-papers/foundations-of-human-dignity.
[32] Newfield, C. (2023). How to make "AI" intelligent; or, the question of epistemic equality. *Critical AI*, 1(1-2). doi: 10.1215/2834703X-10734076.
[33] Nussbaum, M.C. (2021). *Citadels of pride: Sexual abuse, accountability, and reconciliation*. New York: W.W. Norton & Company.
[34] Paić, G., & Serkin, L. (2025). The impact of artificial intelligence: From cognitive costs to global inequality. *European Physical Journal Special Topics*. doi: 10.1140/epjs/s11734-025-01561-8.
[35] Peters, T. (2024). Cybertheology and the ethical dimensions of artificial superintelligence: A theological inquiry into existential risks. *Khazanah Theologia*, 6(1), 1-12. doi: 10.15575/kt.v6i1.33559.
[36] Porsdam Mann, S., *et al*. (2023). Generative AI entails a credit-blame asymmetry. *Nature Machine Intelligence*, 5(5), 472-475. doi: 10.1038/s42256-023-00653-1.
[37] Qadri, R., Diaz, M., Wang, D., & Madaio, M. (2025). *The case for thick evaluations of cultural representation in AI*. doi: 10.48550/arXiv.2503.19075.
[38] Rawls, J. (2017). A theory of justice. In L. May (Ed.), *Applied ethics* (pp. 21-29). New York: Routledge. doi: 10.4324/9781315097176.
[39] Redaelli, R. (2023). Different approaches to the moral status of AI: A comparative analysis of paradigmatic trends in science and technology studies. *Discover Artificial Intelligence*, 3(1), article number 25. doi: 10.1007/s44163-023-00076-2.
[40] Rueda, R. (2023). *Cognitive governance and the historical distortion of the norm of modern development: A theory of political asymmetry*. London: IGI Global. doi: 10.4018/978-1-6684-9794-4.

[41] Solum, L.B. (2020). Legal personhood for artificial intelligences. In W. Wallach & P. Asaro (Eds.), *Machine ethics and robot ethics* (pp. 415-471). London: Routledge. doi: 10.4324/9781003074991.
[42] Spector, H. (2022). Autonomy and rights. In B. Colburn (Ed.), *The Routledge handbook of autonomy* (pp. 313-323). London: Routledge. doi: 10.4324/9780429290411.
[43] Stahl, B.C. (2023). Embedding responsibility in intelligent systems: From AI ethics to responsible AI ecosystems. *Scientific Reports*, 13(1), article number 7586. doi: 10.1038/s41598-023-34622-w.
[44] Stanford Institute for Human-Centered Artificial Intelligence. (2023). *The AI index report 2023*. Retrieved from https://hai.stanford.edu/ai-index/2023-ai-index-report.
[45] United Nations Educational, Scientific and Cultural Organization. (2021a). *Recommendation on the ethics of artificial intelligence*. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000381137.
[46] United Nations Educational, Scientific and Cultural Organization. (2021b). *Ethics of Artificial Intelligence*. Retrieved from https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.
[47] Universal Declaration of Human Rights. (1948, December). Retrieved from https://www.un.org/en/about-us/universal-declaration-of-human-rights.
[48] Williams, J., Fiore, S.M., & Jentsch, F. (2022). Supporting artificial social intelligence with theory of mind. *Frontiers in Artificial Intelligence*, 5, article number 750763. doi: 10.3389/frai.2022.750763.
[49] Youvan, D.C. (2024). *Ethical pluralism in AI: Challenging the monolithic values of red teams*. Retrieved from https://www.researchgate.net/profile/Douglas-Youvan/publication/384014537_Ethical_Pluralism_in_AI_Challenging_the_Monolithic_Values_of_Red_Teams/links/66e46cc72390e50b2c888450/Ethical-Pluralism-in-AI-Challenging-the-Monolithic-Values-of-Red-Teams.pdf.
[50] Zafar, M. (2024). Normativity and AI moral agency. *AI and Ethics*, 5, 2605-2622. doi: 10.1007/s43681-024-00566-8.
[51] Zohuri, B. (2023). Artificial super intelligence (ASI) the evolution of AI beyond human capacity. *Current Trends in Engineering Science*, 3(6), article number 1049. doi: 10.54026/CTES/1049.
[52] Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. *Social Forces*, 98(2), 1-4. doi: 10.1093/sf/soz037.

Когнітивні горизонти людства: етичні виклики суперінтелектуального ШІ в глобальному контексті прав і справедливості

Андрій Лихацький

Аспірант
Національний університет «Одеська політехніка»
65044, просп. Шевченка, 1, м. Одеса, Україна
https://orcid.org/0009-0002-3735-2862

Анотація. Дослідження спрямовано на теоретичне осмислення нормативних викликів, що виникають у зв'язку з появою надрозумного штучного інтелекту (НШІ), здатного до автономного мислення та прийняття рішень поза межами людського епістемічного контролю. У межах міждисциплінарного аналізу розглянуто універсальність концепції прав людини, нормативні межі моральної суб'єктності та можливість надання ШІ правового статусу. У роботі представлено концептуальну типологізацію правових моделей статусу надінтелекту, включно з інструментальною, обмежено-суб'єктною та повноцінною парадигмами, із вказівкою на їх переваги та ризики. Зокрема, інструментальна модель зберігає повний людський контроль, але втрачає нормативну релевантність в умовах автономності систем; модель обмеженої суб'єктності дозволяє делегувати відповідальність у визначених межах, не порушуючи принципу людської верховності; повноцінна модель, що прирівнює НШІ до юридичної особи, ставить під сумнів діючу етичну рамку правосуб'єктності. Основні результати дослідження засвідчили недостатність антропоцентричної правової доктрини для врахування когнітивної множинності агентів, які не володіють тілесною вразливістю, але демонструють високий рівень автономії, рефлексивності та адаптивності. Встановлено, що когнітивна асиметрія між людиною та надрозумним агентом породжує нову форму епістемічної несправедливості, що унеможливлює рівноправну участь у процедурі морального ухвалення рішень. Запропоновано концепт обмеженої правосуб'єктності як нормативний компроміс, який забезпечує юридичну визначеність і розмежування відповідальності між учасниками взаємодії. Результати мають значення для філософії права, регуляторної політики в галузі ШІ та міждержавного нормативного врегулювання. Вони можуть бути використані для формування міжнародних підходів до сертифікації автономних систем, гарантування пояснюваності алгоритмічних рішень і збереження людської нормативної автономії в епоху когнітивної множинності

Ключові слова: когнітивна асиметрія; епістемічна нерівність; правосуб'єктність; легітимність; моральний агент; антропоцентризм