ФІЛОСОФІЯ ДУХОВНОЇ СФЕРИ АБО ІНТЕЛЕКТУАЛЬНИХ ПРОЦЕСІВ

Bondar Svitlana Valeriivna
Candidate of Philosophical Sciences, Associate Professor,
Associate Professor at the Department of Sociology
Kyiv National Economic University named after Vadym Hetman
54/1, Beresteysky ave., Kyiv, Ukraine
orcid.org/0000-0003-2076-1889

COULD ARTIFICIAL INTELLIGENCE BE MORAL? TEST OF HUMANNESS IN LIGHT OF IMMANUEL KANT'S IDEAS

***Relevance of the problem**. It remains a considerable distance to create artificial intelligence that is sufficiently advanced to be compared to the human mind. However, the issue of methods for detecting consciousness in artificial intelligence systems has attracted the attention of businesspeople, scientists, and philosophers. They are interested in whether it will be possible to draw a line and distinguish human consciousness from the imitation of thought processes, artificial intelligence. What criteria determine the presence of consciousness in AI? How not to miss the moment of its emergence? **The purpose of this study** is to identify the fundamental differences between the human mind and artificial intelligence systems, and to determine the possibility of detecting signs of awareness of actions in artificial intelligence. **Methods**. The study uses an interdisciplinary and comparative approach. The comparative method, critical analysis and generalisation are used. **Research results**. Kant's ideas about reason, morality and human nature provide a theoretical basis for understanding the possibilities and potential ethical problems created by the implementation of AI into the socio-cultural sphere. Kant's statement about the universality of the laws of reason suggests that the study of these laws allows for the modelling of intellectual processes, and possibly the creation of artificial intelligence. The defining difference between the human mind and artificial intelligence is the presence in humans of a centre of cognition, which Kant calls the transcendental unity of apperception, which determines the global goal of cognition and allows one to identify oneself as the initiator of cognition and any other actions. Morality constructs and preserves the integrity and stability of the cognitive centre, creating the basis for such mental abilities as prediction, pattern detection, and understanding of meaning, due to the ability to view a life situation or research problem holistically. To detect awareness of artificial intelligence's actions, one should strictly define the measurement conditions (set a period, create a controlled situation, select the norms in relation to which awareness is established). Signs of the presence of consciousness in artificial intelligence include its perception of its actions, the ability to distinguish them from the absence of action, understanding the meaning of its actions, and the capacity to establish a cause-and-effect relationship between an action and its result, as well as managing its actions. **Conclusions.** Artificial intelligence is a unique technology for verifying human authenticity, which is carried out in the context of moral judgments. The creation of an artificial mind similar to a human one is impossible without its simultaneous moral development. I. Kant's emphasis on the inseparability of moral judgments from thought processes, more broadly, cognitive processes, allows us to assume that the lack of morality is simultaneously an indicator of the inability of a rational being to understand the essential axiological components of interpersonal relationships and the cognitive process or serves as a preventive mechanism that limits the capabilities of reason. In this case, we are not so much*

concerned with intellectual limitations, but rather with the inability to adequately perceive the axiological component of social phenomena and cognitive processes, and to accept truth as the value and ultimate goal of any cognition.

Key words*: artificial intelligence, cognition, consciousness, human being, I. Kant, morality, practical reason.*

The introduction. What does it mean to be a human being? How does a human differ from other rational beings? Immanuel Kant provided answers to these philosophical questions. In an era when science and philosophy lacked substantial experimental material on human nature, thinking, and cognition, the German philosopher meditated on the structure of thinking and the source of morality. However, his reasoning was based mainly on theoretical research and life experience, many of his ideas about reason and thinking remain relevant today.

In his reasoning about thinking, cognition and moral behaviour, I. Kant proceeds from the universality of the laws of reason and the absoluteness of ethical values. The relevance of Immanuel Kant's approach to the study of these issues lies in the fact that the German philosopher explores reason and moral principles, taking them beyond the boundaries of anthropology. At the same time, he points out their irreducibility to theological issues. I. Kant is convinced that these laws are universal and valid for every rational being [10]. Like the laws of physics, they do not depend on human nature but are realised in it.

I. Kant is guided by the principles of Christian anthropology, declaring humanity and each individual as the priority goal of God and nature [10]. At the same time, his philosophical ideas are a prerequisite for a significant weakening of the principles of anthropocentrism, which exaggerate the importance of human beings in the world and hinder the acquisition of reliable and unbiased knowledge. The relevance of I. Kant's views is that he considers reason as a phenomenon inherent not only to humanity. He assumes the existence of non-human forms of intellectual processes.

Kant's philosophical ideas resonate with modern theories that form the basis of artificial intelligence development. Based on his ideas about the universal nature of the mind, it can be assumed that knowledge of the laws underlying thinking can lead to the creation of devices that imitate mental processes or even artificial intelligence. The creation of artificial intelligence systems is impossible without studying human consciousness, and a deep understanding of this complex phenomenon is impossible without analysing the philosophical heritage of humanity.

Immanuel Kant's philosophical concept of practical reason can help establish criteria for detecting consciousness in artificial intelligence and ways to distinguish intelligent beings from models that imitate intellectual and emotional processes. It can serve as a theoretical basis for developing a type of Voigt-Kampf test [12], which will help distinguish between living human consciousness and its simulation.

Kant's idea of universal principles and laws inherent in the structure of the world and reason opens up ways to solve many of the current philosophical problems, which concern the cognition of the world; the construction of artificial intelligence; overcoming the existential crisis by developing ethical norms, taking into account changes in socio-cultural reality, caused, among other things, by the development of technologies related to AI.

In his book "The Unity of Reason", Heinrich Dieter writes [5], "Kant's theoretical philosophy is an investigation of the universal structures of reason" [5, 85]. He emphasises the unity of reason in Kant's conception and the special significance of practical reason, initiated by human will [5]. Dieter emphasises the importance of rationality in the composition of moral judgments and actions. He points to the correlation between freedom and philosophising, the relationship between liberty and thinking, emphasising the importance of a special kind of freedom in Kant's philosophy, which manifests itself as a moral feeling – respect for the law [5].

In the monograph "Kant and Artificial Intelligence", the authors consider the issue of ethical constraints attached to the process of designing and developing robots. In particular, they ask whether the development of moral freedom in robots should be hindered [2]. Appealing to Kant's philosophical concept of practical reason, the authors conclude that the moral development of

robots should be limited. Otherwise, it will complicate their exploitation. According to Kant, using a rational moral being as a means to achieve a goal is immoral and unworthy [2]. The authors do not believe that AI is capable of making fair moral decisions. Having analysed the concept of practical reason in detail, they argue that "Practical reason cannot be artificial" [14, 170]. AI has no will, which confirms that it is incapable of intentional action [14]. With this work, the authors have made a significant contribution to a detailed study of the history of philosophy and substantiation of the relevance of Immanuel Kant's philosophical ideas.

The authors of the monograph "The Road to General Intelligence" explore the possibilities and prospects for creating general AI, which they define as "a property of a machine that exhibits general-purpose intelligence of the kind exhibited by humans, i.e., enjoying the ability to continually adapt existing knowledge to different domains" [15, 1]. The key ability of AI is to learn and adapt to the environment and circumstances, and to foresee various possibilities [15]. The stimulus for intelligence activity is the "a priori reward function" [15, 8].

The authors of the article [16] on the importance of design and materials used in robotics define key properties of robots: contextual dependence and the ability to adapt to the environment. Evolutionary robotics is a direction in robotics that focuses on developing robots capable of adapting their structure in response to environmental changes [1]. Modern technologies enable robots to interact with their environment in the same manner as living beings. Therefore, it is becoming complex to distinguish them. As robotics and AI continue to improve, this need is increasingly pressing.

Calum Neill [12] analyses how Philip K. Dick's novel, within the framework of a science fiction narrative, examines questions close in their problematic to those that Kant explores in his "Critiques": what does it mean to be human? Is there a difference between artificial human-like intelligent beings and humans, in fact? What is the difference between the human mind and the non-human mind? According to the author, the key factors in distinguishing the human mind from artificially created minds are revealed through the process of introspection and self-assessment by the subject test. "The test centres on the subject's own knowledge of their status, whether this knowledge is consciously known or not" [12, 3]. A similar test can be used to detect consciousness in artificial intelligence. The author focuses on an interesting technique for detecting artificial consciousness: prompting the test subject to reflection in a situation of uncertainty.

Purpose. To define the criteria for distinguishing the human mind from the imitation of its capabilities using artificial intelligence technologies. To determine the indicators of the presence of consciousness in artificial intelligence. The tasks of this work are to analyse the possibility of the morality of artificial intelligence in the context of Immanuel Kant's ideas, and to examine philosophical approaches to determining the nature of reason and morality. Special attention is paid to how the Kantian concept of practical reason can be used to assess the limits and potential of ethical behaviour of AI, as well as to form theoretical foundations for testing its ethics.

Results and discussion. The limits of knowledge and the possibilities of morality: a Kantian approach in the era of artificial intelligence. Kant argues that cognition requires the coordination of methods of cognition with the subject area [10]. A person, cognising the natural world, is himself a part of it, ordering and constructing it in accordance with the peculiarities of his thinking and understanding. The philosopher emphasises the relationship between the theoretical (cognitive) and practical (moral) aspects of human thinking. In cognition, as in theoretical activity in general, a person is subject to the laws of nature. The laws of the intelligible world, reason and morality, according to Kant, must exist in the physical world, but in such a way as "not to harm the laws of this world" [10, 38].

Moral activity enables a person to transcend the boundaries of the directly given natural reality, extending beyond the context of a specific socio-historical situation. Morality provides a comprehensive understanding of the current state of affairs and the possible development of events. It contributes to reaching a new level of knowledge of the natural world, thanks to the acquired freedom of thought and cognition.

Kant demonstrates that what is inaccessible to theoretical knowledge can be comprehended through the practical reason, in interpersonal interaction and mutual understanding. Without them, full-fledged cognition is impossible. Understanding arises from the mutual recognition of another person as a self-valuable individual (goal). Such relationships are possible only voluntarily, provided that one is aware of one's freedom. In this case, one should distinguish between knowledge about freedom and awareness of oneself as free.

A person as a goal is a key principle of ethics, interpersonal relationships, and social justice. At the same time, it is a factor that limits the possibilities of knowledge and understanding. The centrality of human interests (anthropocentrism) shapes our knowledge and understanding of ourselves, but this does not mean that humanity has, or is guaranteed to acquire, exhaustive knowledge of the world. On the contrary, the Kantian concept of "things in themselves" clearly indicates the opposite.

Kant argues that the existence of things in themselves is connected with the peculiarities of the human mind [11]. The thing in itself is knowledge that is closed in itself and closed not only because of its complexity, but also because of the human inability to comprehend it. The fact is that the state of the human mind does not correspond to the principles and laws of the construction and existence of the world. A person focused on himself seeks, first and foremost, the satisfaction of his own needs and happiness, as he imagines it.

The state of the human mind and self-understanding must correspond to the degree of cognising of the world, the level of knowledge; otherwise, it (knowledge) will remain a thing in itself. Why does a person need a categorical imperative as an unambiguous principle, to adhere to which is the duty of everyone? Because the world is not what a person perceives and understands it to be. Freedom and objective knowledge can be achieved through adherence to the laws of reason, which are universal principles of thinking and existence for all rational beings, not just humanity.

According to Kant, for a person, his own life and the life of any other person, as a rational being, as well as humanity, should be the goal. This is a law attributed to all rational beings. As a mortal being, a person strives for well-being and happiness [10]; therefore, every person's rights and interests must have meaning for others and be taken into account by them, which is a natural law.

To be the goal of nature, one must be alive [10]. A person, as a living and rational being, occupies a central position in the world. However, the principles of Kant's philosophy permit, and even presuppose, the existence of non-human forms of minds and rational beings. Therefore, anthropocentrism, which remains a priority worldview principle due to our belonging to humanity, must be combined with the acceptance of the fact that humanity shares the world with other living rational beings. Even if the existence of these beings seems insignificant, we must take into account their interests, since we do not know their true purpose in the world. According to Kant, humanity cannot obtain final knowledge about the meaning of existence and the structure of the world, as it lacks comprehensive knowledge of the world in general [11].

According to Kant, for each person, his own life is the most important [10]. This is a significant obstacle in both interpersonal relationships and cognition. This principle can become an obstacle in recognising other types of mind, understanding and accepting different cultures and points of view. This is one of the main reasons for the fear of artificial intelligence, which is capable of debunking human arrogance, forcing one to reassess one's capabilities and look at the world and oneself from a different angle.

The task facing a person, as a rational being, is quite complex: to learn to distinguish one's subjective perception and self-perception from objective cognition, and the desire to be happy, natural for a mortal being, from the voice of reason and the fulfilment of the duty attributed to it − compliance with the categorical imperative. However, this is the only way to achieve the truth of cognition, the realisation of its humanistic essence − morality − and the acquisition of a degree of freedom accessible to a person. This task is challenging to accomplish, but the development of artificial intelligence brings such a possibility closer to realisation. Forms of intelligence other than human, particularly artificial intelligence, offer a chance to see the world as a person does not know

it, to gain access to alien experiences and knowledge, and to develop a worldview that is different from that of human. In some areas of scientific research, it has already become not an auxiliary, but a central element of scientific knowledge [7].

Do androids dream of morality?

The theme of defining the boundary between humans and artificially created intelligent beings was explored in Philip Dick's novel "Do Androids Dream of Electric Sheep?" [6]. He proposed the Voigt-Kampf empathy test to identify artificially created intelligent androids. However, Kant's idea is deeper. The difference lies in how a person uses his intellect and knowledge. According to Kant's concept, the key to the difference between humans and artificial intelligence may lie in morality – practical reason [2].

The realisation of practical reason requires volitional efforts. It is not just purposeful activity, but actions based on free will and conscious adherence to the laws approved by reason. Kant explains that practical reason "at least as a naturally acquired ability, can never be completed" [10, 30]; therefore, a person needs eternity for continuous self-improvement and efforts to realise their humanistic nature. Practical reason is a complexly organised practice, capable, thanks to autonomy, of initiating its own causality, which involves planning its activities [5]. The goal of this process is the realisation of the good (humane) nature of reason.

If morality is realised in practical actions, then the awareness of freedom comes through the fulfilment of the moral law – the categorical imperative. Freedom, in turn, stimulates the development of rationality [10], allowing one to overcome the external limitations of the existing circumstances. However, the abuse of freedom leads to its loss. To act freely, it is necessary to break the chain of cause and effect inherent in the natural world, to manifest the will, to initiate freedom of thought [10].

Freedom for Kant is primarily the ability to moral legislation, possible under the condition of ethical maturity, self-sufficiency and independence from the events and connections of the natural and social worlds. Reason is associated with the will, the ability to establish laws and introduce one's own causality, which distinguishes a person from natural phenomena. Reason is unambiguously moral, and morality cannot be abstract; it requires implementation. A person must not only differentiate between good and evil, that is be compos mentis. It is necessary to act in accordance with moral principles. Freedom is inaccessible to theoretical understanding, but it is achievable in practice.

Immanuel Kant argued that freedom is inseparable from reason. However, finding confirmation of this in surrounding events is difficult, and may even be impossible according to Kant. Therefore, it is so important not to be limited by the framework of a specific context of events. The ability to judge, according to Kant, consists in finding general a priori principles that allow us to go beyond the boundaries of the natural world into the sphere of freedom. A person finds rationality and purposiveness in the surrounding world, since the human mind contains their intuitive a priori understanding. Kant believes that the source of expediency, as well as morality, is not apparent, but reason convinces us of their existence.

Based on Kant's ideas, it is possible to develop methods for detecting highly developed consciousness, as well as to determine the criteria for distinguishing human intelligence from imitation of its abilities using software. Signs of the presence of highly developed consciousness include voluntary compliance with the laws and principles of morality, awareness of one's freedom and the ability to exercise it, and free will. As Kant wrote: "It is, therefore, not limited to human beings only but applies to all finite beings that have reason and will and even includes the infinite being as the supreme intelligence". [10, 29]. Consciousness is not exhausted by intelligence but presupposes it.

It is not always possible to determine the motive or intention of an act. People commit immoral acts and act in obedience to circumstances rather than moral principles. At the same time, research shows [8] that rational behaviour, the ability to achieve goals, learn, and solve problems, is inherent in living beings that do not have a brain. Modern science confirms the presence of complex intel-

lectual processes in species of living beings other than humans [3]. However, Kant sets strict limits that determine what it means to be a rational being. The criteria are not biological or social characteristics. Suppose a representative of Homo sapiens, which is traditionally perceived as a model against which the degree of rationality of all other living beings is established, is devoid of morality. In that case, he is neither free nor rational in the Kantian sense.

Automatic actions of mechanisms, conditioned by the environment and design, may exhibit rational behaviour and purposefulness. At the same time, the actor, in fact, has no hint of the presence of intelligence, programs, or algorithms that would guide his behaviour. He is entirely contextually dependent. His design determines the nature of his activity. A person as a being, included in the events of the world with all his feelings, will, and thinking, contained in a continuous chain of causes and effects, is only a "thinking automaton", or an android in the terminology of the aforementioned science fiction novel.

In robotics, the term "physical intelligence" refers to the ability of a robot to exhibit rational behaviour through its design or materials, rather than relying on complex software and voluminous calculations. Physical intelligence enables the robot to adapt to its environment, allowing it to evolve alongside it. In my opinion, physical intelligence is a good example of the context-dependent nature of human-created imitations of its individual abilities, for example, the ability to perceive, interact with the environment, and "understand" its body. Unlike the concept of artificial intelligence, which focuses on methods of imitating human cognitive abilities, the idea of physical intelligence emphasises the importance of physical interaction between robots and their surrounding world.

Another example of the fundamental importance of intelligent systems and robotics in adapting to and responding to the demands of their surrounding world is the branch of robotics known as evolutionary robotics. One of the key tasks is the effective adaptation of robots to the conditions of the "wild environment" [1] without compromising productivity.

To avoid ethical and legal antinomies, the authors of the monograph "Kant and Artificial Intelligence" propose limiting the "moral development" of artificial intelligence and robotics [2]. Since the authors of the monograph analyse the actual and potential abilities of artificial intelligence, drawing on the ideas of Immanuel Kant, it is logical to critique certain statements of the authors from the perspective of Kantian philosophy. According to Kant, it is impossible to impose restrictions on moral development without simultaneously limiting the development and functioning of the intellect. For the German philosopher, it is evident that there is a kind of fuse that inhibits the realisation of intellectual abilities in the presence of moral defects [9]. Theoretical and practical reasons are two inseparable sides of the same phenomenon. The definition of general artificial intelligence, formulated by the authors of the monograph themselves [2], as well as presented in a monograph devoted to the technical aspects of developing general AI, assumes that it is impossible to impose restrictions on its abilities without harming its functioning [15].

At the level of cognitive abilities, the differences between human and artificial intelligence are becoming less and less noticeable [13]. Shortly, AI will surpass humans in cognitive capabilities. At the level of individual interactions, differences can be detected. Artificial intelligence is currently unable to fully replicate the complexity of human interpersonal relationships, which involve understanding one another, compassion, emotional preferences, and other nuanced aspects. All that the Voigt-Kampf test (a test of empathy) reveals.

The authors of the article "Preventing Antisocial Robots: A Pathway to Artificial Empathy" emphasise the importance of developing empathy in AI and robots to prevent the harm they can potentially cause to people if they lack "effective empathy and moral behaviour" [4]. They believe that artificial intelligence will learn to understand human feelings and sympathise if it experiences vulnerability, perception of damage, pain, and suffering. In fact, the authors of the article identify human empathic experiences with artificial empathy in AI, which is wrong. It should also be noted that similar experience in humans does not guarantee the development of empathy. The creation of ethical, socially aware artificial intelligence is increasingly necessary, given the expansion of its

applications; however, empathic abilities and complex emotional processes are not yet incorporated into logical schemes that can be implemented through calculations.

According to Kant, a person must be able to distance himself from personal sympathies and preferences, for the sake of fulfilling his duty and serving the laws of reason [10]. However, without the experience of interpersonal relationships, which is impossible without empathic selectivity, many aspects of the surrounding world and his own nature remain incomprehensible to the mind, closed off as "things in themselves".

According to the author of the novel "Do Androids Dream of Electric Sheep?", it is challenging to distinguish between humans and androids. Outwardly, they appear to be ordinary people. They were betrayed by an immature emotional reaction to simple questions, caused by their lack of personal history, insufficient life experience, and limited experience in interpersonal relationships [6]. Dick provides the following explanation for the "helpless failure" of the Voigt-Kampf empathic test, which is simple for humans: androids possess a different type of intelligence, distinct from that of humans. This is not the intelligence of omnivorous group animals, of which humans are descendants, but the intelligence of solitary predators, for whom empathy and any manifestation of compassion for prey reduces the chances of survival.

The author [12] of the monograph demonstrates that the test should not only reveal the absence of a specific emotional reaction in the subject, but also his awareness that he is unable to adequately respond emotionally to a regular event or circumstance. However, the real "android detection test" is a test based on the philosophical ideas of Immanuel Kant, which involves identifying the motivation behind actions. Kant requires each person to be extremely honest in assessing their own actions and to engage in strict introspection to identify the motives behind their behaviour. Only in this way can the degree of humanity, morality of a person, and their authenticity be determined.

Indicators of consciousness in computing systems, particularly in artificial intelligence, include the ability to perceive their actions, distinguish them from the absence of action, know their value, and determine the cause-and-effect relationship between their actions and their results, as well as control their actions. About computing devices, the concept of "consciousness" should be approximated in meaning to the idea of "sane" used in jurisprudence. It makes sense to define it in this way, as the concept of "sane" captures the awareness of actions at a specific moment in time, in a particular situation, and in relation to established norms or rules, which simplifies the possibility of its detection. A strict definition of measurement conditions is essential, as the level of awareness and state of consciousness can change even in a person, for example, under the influence of strong emotions, prolonged stress, or medications.

The differences between the human mind and its imitation using artificial intelligence technologies, in addition to those mentioned in the monograph "Kant and Artificial Intelligence", are the transcendental unity of apperception – the centre of cognition, which structures knowledge and gives stability and integrity to the subject of cognition, completeness to what he knows. This is a prerequisite for analysing knowledge, which enables us to identify existing patterns and predict the existence of potential ones.

Artificial intelligence does not have a central hub. It is unstable. Its functioning can be altered by modifying the algorithms. Billions of calculations form it. All of them are equivalent and important. Algorithms, laws of logic and mathematics determine the priority of calculations. AI does not have a single centre that would organise cognition and awareness of itself as a thinking being.

Such centring, or coordination of knowledge in a person, is performed, in particular, by moral beliefs and a system of ethical values that guide him in his actions. The "core" of the personality, which remains inviolable under any circumstances, is morality. It is a constructive force that coordinates the main features of consciousness and self-awareness of the individual. The human mind is determined not only by cognitive abilities and intellect, but also by the ability to distinguish good from evil and to realise this understanding in actions. Kant, in his "Anthropology" [9], suggests that self-knowledge is the awareness of one's freedom and is grounded in the fundamental

ethical principle − the categorical imperative. Morality is rooted in freedom. By realising it in actions, a person knows himself and the world.

The peculiarity of AI is that it acts in accordance with the task set before it, according to a given program and based on circumstances. Its actions have no independent meaning. They are only a series of events in the process of contextually conditioned coincidences.

Artificial intelligence knows the logical meaning of truth as the opposite of false. But not its axiological meaning as a universal human value. Such an understanding of truth is achieved through moral autonomy, which frees one from attachment to narrow situational limitations and opens prospects for freedom of thought, prediction (one of the key abilities of the human mind), and goes beyond the conclusions achievable from available knowledge.

Conclusions. According to Kant, reason and morality are inseparable; therefore, the ability to distinguish good from evil, and to put this knowledge into practice, to act for the sake of good, is the original feature of rational beings. They are not only aware of the consequences of their actions, but they are also able to determine the motives of their behaviour and distinguish their own desires and subjective preferences from the laws established by reason, which should guide them in their actions. They are guided in their actions by the concept of the good of a person and humanity, which justifies the correctness of actions or the truth of the knowledge acquired. Kant directly testifies to the primary purpose of reason. This is not cognition. This is the formation of interpersonal relationships, in which every rational being is the goal of any action. Kant saw the specificity of humanity in rationality, but, first and foremost, in morality, the peculiarities of interpersonal interaction, and the ability to set independent (autonomous) goals.

Destructive activity limits the capabilities of the mind, leading to intellectual and moral degradation, as well as a loss of freedom of judgment and action. Therefore, the so-called "maliciousness" of AI is merely a reflection of unhealthy relationships in society, highlighting how immature and inferior people's attitudes towards one another remain.

The introduction of new technologies and the creation of increasingly complex artificial intelligence systems lead to several ethical and legal conflicts, so it is essential, in the process of developing AI, to clearly determine what we want to create: a being similar to a person, and therefore, one that has the same rights and responsibilities as a person, or a convenient tool, specialized for performing a specific type of work, but at the same time limited by this specialization? It is appropriate to develop both types of artificial intelligence.

Narrowly specialised devices, such as a robot vacuum cleaner, do not require knowledge of ethics or the skills of a psychologist. However, many types of artificial intelligence systems will have to interact with people directly, inevitably making ethical choices and decisions independently, without waiting for an order from a person or acting contrary to such an order if the person has criminal intentions. As artificial intelligence becomes increasingly integrated into social and cultural relations, the need for it to be a full-fledged subject of decision-making, including moral considerations, grows correspondingly.

The prospects for creating a general artificial intelligence that is similar to human intelligence remain uncertain. Developers seek to imbue it with those qualities that a person possesses or aspires to have. There should be no restrictions on its development. By limiting the capabilities of artificial intelligence, we endanger humanity, resulting in a grotesque or overly primitive image of ourselves, and never achieving the primary task set for artificial intelligence: human self-knowledge, harmonisation of social relations, and writing the history of civilisation.

Artificial intelligence is a unique technology for testing human authenticity, which is applied in the context of ethical judgments. Immanuel Kant claims that a person is neither a puppet nor an automaton [10]. I believe that a person can live as a "thinking automaton", and the key prerequisite for this is immorality, which is revealed primarily through introspection. AI is a way of organising cognitive processes that differs from human ones, which makes it possible to comprehend the world, relying on knowledge and experience that are inaccessible to human perception. This opens up an alternative perspective on human nature.

BIBLIOGRAPHY

1. Angus M., Buchanan E., Le Goff L.K., Hart E., Eiben A.E., Carlo M.D., Winfield A.F., Hale M.F., Woolley R., Timmis J., Tyrrell A. M. Practical hardware for evolvable robots *Frontiers.* 2023 Vol. 10. DOI: https://doi.org/10.3389/frobt.2023.1206055

2. Benossi L., Bernecker S. A Kantian Perspective on Robot Ethics/ T. Schlicht, R. Evans, S. Baiasu, H. Kim, L. Benossi, S. Bernecker, D. Schonecher, E. E. Schmidt, A.T. Wright C. Dierksmeier, L. Berger. *Kant and Artificial Intelligence/* Ed. by H.Kim, D. Schönecker. Berlin/ Boston: Walter de Gruyter GmbH, 2022. P. 147–168. DOI: https://doi.org/10.1515/9783110706611

3. Chalmers D. J. Could a Large Language Model be Conscious? *arXiv.* 2023. DOI: https://doi.org/10.48550/arXiv.2303.07103

4. Christov-Moor L., Reggente N., Vaccaro A., Schoeller F., Plumer B., Douglas P. K., Iaconobi M., Mann K., Damasio A., Kaplan J. T. Preventing antisocial robots: A pathway to artificial empathy. *Science Robotics.* 2023 Vol. 8, Issue 80 Doi: 10.1126/scirobotics.abq3658

5. Dieter H. The Unity of Reason: Essays on Kant's Philosophy/ Ed. by R.L. Velkley, transl. by J.Edwards, L.Hynt, M. Kuehn, G.Zoeller. USA: Harvard University Press, 1994. 256 p.

6. Dick P. K. Do Androids dream of Electric Sheep? 1968. URL: https://files.cercomp.ufg.br/weby/up/410/o/Phillip_K._Dick_-_Do_Androids_Dream_of_Electric_Sheep_cópia.pdf

7. Gong X., Holmes J., Li Y., Liu Z., Gan Q., Wu Z., Zhang J., Zou Y., Teng Y., Jiang T., Zhu H., Liu W., Liu T., Yan Y. Evaluating the Potential of Leading Large Language Models in Reasoning Biology Questions. *arXiv.* 2023. Doi: https://doi.org/10.48550/arXiv.2311.07582

8. Jacobsen R. Brains Are Not Required When It Comes to Thinking and Solving Problems − Simple Cells Can Do It. *Scientific American.* 2024. URL: https://www.scientificamerican.com/article/brains-are-not-required-when-it-comes-to-thinking-and-solving-problems-simple-cells-can-do-it/?utm_medium=social&utm_source=twitter&utm_campaign=socialflow

9. Kant I. *Anthropology from a Pragmatic Point of View* / transl. and ed. by R. B. Louden. Cambridge: Cambridge University Press, 2006. 248 p. Doi: https://doi.org/10.1017/CBO9780511809569

10. Kant I. *Critique of Practical Reason* / transl. and ed. by M. Gregor. Cambridge: Cambridge University Press, 2015. 142 p. Doi: https://doi.org/10.1017/CBO9781316136478

11. Kant I. *Critique of Pure Reason /* transl. and ed. by P. Guyer & A. W. Wood. Cambridge: Cambridge University Press, 1998. 786 p. Doi: https://doi.org/10.1017/CBO9780511804649

12. Neill C. *From Voight-Kampff to Baseline Test: By Way of an Introduction* / C. Neill, B. Tyrer, S. Zizek, T. McGowan, D. Bristow // Lacanian Perspective on Blade Runner 2049 / ed. Neill C. Cham. *Switzerland: Palgrave Macmillan.* 2021. P. 1–11. DOI: https://doi.org/10.1007/978-3-030-56754-5

13. Nosta J. Is "Digital Thought" the Next Cognitive Domain? *Psychologe Today.* 2024. URL: https://www.psychologytoday.com/us/blog/the-digital-self/202407/is-digital-thought-the-next-cognitive-domain

14. Schönecker D. Kant's Argument from Moral Feelings: Why Practical Reason Cannot Be Artificial. Ed. by H.Kim, D. Schönecker / T. Schlicht, R. Evans, S. Baiasu, H. Kim, L. Benossi, S. Bernecker, D. Schönecher, E. E. Schmidt, A.T. Wright C. Dierksmeier, L. Berger // *Kant and Artificial Intelligence* / Ed. by H.Kim, D. Schönecker. Berlin/ Boston: Walter de Gruyter GmbH, 2022. P. 169–188. Doi: https://doi.org/10.1515/9783110706611

15. Swan J., Nivel E., Kant N., Hedges J., Atkinson T., Steunebrink B. The Road to General Intelligence / J. Swan, E. Nivel, N. Kant, J. Hedges, T. Atkinson, B. Steunebrink; Ed. by J. Kacprzyk. Cham, Switzerland: Springer Nature Switzerland AG, 2022. P. 136. DOI: https://doi.org/10.1007/978-3-031-08020-3

16. Zhao Y., Hong Y., Li Y., Qing H., Su H., Yin J. Physically intelligent autonomous soft robotic maze escaper. *Science Advances.* 2023. Vol. 9, Issue 36. DOI: https://doi.org/10.1126/sciadv.adi3254

REFERENCES

1. Angus, M., Buchanan, E., Le Goff, L.K., Hart, E., Eiben, A.E., Carlo, M.D., Winfield, A.F., Hale, M.F., Woolley, R., Timmis, J., & Tyrrell, A. M. (2023). Practical hardware for evolvable robots. *Frontiers* Volume 10. DOI: https://doi.org/10.3389/frobt.2023.1206055

2. Benossi, L., & Bernecker, S. (2022). A Kantian Perspective on Robot Ethics. Ed. by H.Kim, D. Schönecker *Kant and Artificial Intelligence* (p. 147-168). Berlin/ Boston: Walter de Gruyter GmbH. DOI: https://doi.org/10.1515/9783110706611

3. Chalmers D. J. (2023). Could a Large Language Model be Conscious? *arXiv* DOI: https://doi.org/10.48550/arXiv.2303.07103

4. Christov-Moor, L., Reggente, N., Vaccaro, A., Schoeller, F., Plumer, B., Douglas, P. K., Iaconobi, M., Mann, K., Damasio, A., & Kaplan, J. T. (2023). Preventing antisocial robots: A pathway to artificial empathy *Science Robotics* Vol. 8, Issue 80. DOI: 10.1126/ scirobotics.abq3658

5. Dieter, H. (1994). The Unity of Reason: Essays on Kant's Philosophy. (Ed. by R.L. Velkley, transl. by J. Edwards, L. Hynt, M. Kuehn, G. Zoeller). USA: Harvard University Press.

6. Dick, P. K. (1968). Do Androids Dream of Electric Sheep? Retrieved from: https://files.cercomp.ufg.br/weby/up/410/o/Phillip_K._Dick_-_Do_Androids_Dream_of_Electric_ Sheep_cópia.pdf

7. Gong X., Holmes J., Li Y., Liu Z., Gan Q., Wu Z., Zhang J., Zou Y., Teng Y., Jiang T., Zhu H., Liu W., Liu T., Yan Y. (2023). Evaluating the Potential of Leading Large Language Models in Reasoning Biology Questions. *arXiv* Doi: https://doi.org/10.48550/arXiv.2311.07582 Retrieved from: https://arxiv.org/abs/2311.07582

8. Jacobsen R. (2024). Brains Are Not Required When It Comes to Thinking and Solving Problems – Simple Cells Can Do It. *Scientific American.* Retrieved from: https://www.scientificamerican.com/article/brains-are-not-required-when-it-comes-to-thinking-and-solving-problems-simple-cells-can-do-it/?utm_medium=social&utm_source=twitter&utm_ campaign=socialflow

9. Kant, I. (2006). *Anthropology from a Pragmatic Point of View* / transl. and ed. by R. B. Louden. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/ CBO9780511809569

10. Kant, I. (2015). *Critique of Practical Reason* (transl. and ed. by M. Gregor). Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781316136478

11. Kant, I. (1998). *Critique of Pure Reason* (transl. and ed. by P. Guyer & A. W. Wood). Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511804649

12. Neill, C. (2021). *From Voight-Kampff to Baseline Test: By Way of an Introduction.* Neill C.(ed.) Lacanian Perspective on Blade Runner 2049 (p. 1-11). Cham, Switzerland: Palgrave Macmillan. DOI: https://doi.org/10.1007/978-3-030-56754-5

13. Nosta, J. (2024). Is "Digital Thought" the Next Cognitive Domain? *Psychologe Today* Retrieved from: https://www.psychologytoday.com/us/blog/the-digital-self/202407/ is-digital-thought-the-next-cognitive-domain

14. Schönecker, D. (2022). Kant's Argument from Moral Feelings: Why Practical Reason Cannot Be Artificial. Ed. by H.Kim, D. Schönecker *Kant and Artificial Intelligence* (p. 169–188). Berlin/ Boston: Walter de Gruyter GmbH. DOI: https://doi.org/10.1515/9783110706611

15. Swan, J., Nivel, E., Kant, N., Hedges, J., Atkinson, T., & Steunebrink, B. (2022). The Road to General Intelligence. Ed. by J. Kacprzyk. Cham, Switzerland: Springer Nature Switzerland AG. DOI: https://doi.org/10.1007/978-3-031-08020-3

16. Zhao, Y., Hong, Y., Li, Y., Qing, H., Su, H., & Yin, J. (2023). Physically intelligent autonomous soft robotic maze escaper *Science Advances* Vol 9, Issue 36. DOI: https://doi.org/10.1126/sciadv.adi3254

Бондар Світлана Валеріївна
кандидат філософських наук, доцент,
доцент кафедри соціології
Київського національного економічного університету імені Вадима Гетьмана
просп. Берестейський, 54/1, Київ, Україна
orcid.org/0000-0003-2076-1889

ЧИ МОЖЕ ШТУЧНИЙ ІНТЕЛЕКТ БУТИ МОРАЛЬНІСНИМ? ТЕСТ НА ЛЮДЯНІСТЬ У СВІТЛІ ІДЕЙ ІММАНУЇЛА КАНТА

Актуальність проблеми. До створення досконалого штучного інтелекту, такого, щоб його можна було порівняти з людським розумом, ще далеко. Проте проблема методів виявлення свідомості у штучних інтелектуальних систем привертає увагу бізнесменів, науковців та філософів. Їх цікавить, чи можна буде провести межу та відрізнити людську свідомість від імітації розумових процесів, штучного розуму. Які критерії визначають наявність свідомості у ШІ? Як не упустити момент його виникнення? ***Метою дослідження*** *є виявлення принципових відмінностей людського розуму від систем штучного інтелекту; визначення можливості виявлення ознак усвідомлення власних дій штучним інтелектом.* ***Методи.*** *В дослідженні застосований міждисциплінарний та компаративістський підхід. Використовуються порівняльний метод, критичний аналіз та узагальнення.* ***Результати дослідження****. Ідеї Канта про розум, мораль та природу людини забезпечують теоретичну основу для розуміння можливостей та потенційних етичних проблем, створюваних впровадженням ШІ у соціокультурну сферу. Твердження Канта про універсальність законів розуму припускає, що вивчення цих законів дозволяє моделювати інтелектуальні процеси, і, можливо, створити штучний розум. Визначальною відмінністю людського розуму від штучного інтелекту є наявність у людини центру пізнання, який Кант називає трансцендентальною єдністю аперцепції, який визначає глобальну мету пізнання та дозволяє ідентифікувати себе як ініціатора пізнання та будь-яких інших дій. Моральність конструює та зберігає цілісність та стабільність пізнавального центру, створюючи основу для таких здібностей розуму як передбачення, виявлення закономірностей, розуміння сенсу завдяки здатності цілісного бачення життєвої ситуації чи дослідницької проблеми. Для виявлення усвідомленості дій штучного інтелекту слід суворо визначити умови вимірів (встановити часовий період, створити контрольовану ситуацію, вибрати норми, щодо яких встановлюється усвідомленість). Ознаками наявності свідомості у штучного інтелекту можуть бути сприйняття їм своїх дій, здатність відрізняти їх від відсутності дії, знати значення своїх дій та здатність встановлювати причинно-наслідковий зв'язок між дією та її результатом, управляти своїми діями.* ***Висновки.*** *Штучний інтелект є унікальною технологією перевірки людської автентичності, котра здійснюється в контексті моральнісних суджень. Створення штучного розуму, подібного до людського, неможливе без одночасного моральнісного його становлення. Акцент І.Канта на невід'ємності моральнісних суджень від мисленнєвих, ширше когнітивних процесів, дозволяє припустити, що недолік моральності є одночасно показником нездатності розумної істоти розуміти суттєві аксіологічні складові міжособистісних відносин і пізнавального процесу або служить запобіжним механізмом, обмежуючим можливості розуму. При цьому йдеться не так про інтелектуальну обмеженість, а про нездатність адекватно сприймати аксіологічну складову соціальних явищ і когнітивних процесів, приймати істину як цінність і остаточну мету будь-якого пізнання.*

Ключові слова: І. Кант, людина, моральність, пізнання, практичний розум, свідомість, штучний інтелект.