DOI https://doi.org/10.30525/2592-8813-2025-4-17

ARTIFICIAL INTELLIGENCE AND THE PROBLEM OF RESPONSIBILITY: THE NEW BOUNDARIES OF ETHICS

Laman Garayeva Ahmad,
Lecturer at the Department of Social Sciences,
Azerbaijan State Agricultural University (Ganja, Azerbaijan)
ORCID ID: 0009-0007-3180-2718

Tamilla Rzayeva Zahir,
Lecturer at the Department of Social Sciences,
Azerbaijan State Agricultural University (Ganja, Azerbaijan)
ORCID ID: 0000-0002-4638-6485

Abstract. The article philosophically analyses the issues of responsibility, morality, and ethical decision-making that arise in the context of the development of artificial intelligence (AI). The author focuses on the distribution of responsibility between humans and technological agents, as well as on the nature of intention and moral values. The main objective is to examine whether artificial systems are capable of making moral decisions and to show how this relates to the ethical position of the human being. The study compares the behavioural models of AI with the particularities of human consciousness and evaluates the integration of technology into ethical principles along with its social implications. The author notes that as AI advances, the concept of morality acquires not only an individual but also a systemic character; that is, responsibility becomes the joint product of human–machine interaction rather than belonging solely to a single agent. At the same time, adopting a posthumanist perspective, the article questions the central position of the human and explores the possibility of technological entities occupying a place within the moral order. Thus, artificial intelligence is presented not merely as a technical phenomenon but as a philosophical event that stimulates the emergence of a new moral paradigm.

Key words: artificial intelligence, responsibility, morality, ethical decision-making, philosophy of technology, posthumanism.

Introduction. One of the most significant turning points of twenty-first-century human civilization is the emergence and rapid development of artificial intelligence (AI). Machines' ability to learn, make decisions, and even engage in creative thinking has moved beyond the realm of science fiction and now produces tangible social, legal, and ethical consequences in real life. However, this rapid progress also gives rise to new questions: Who is responsible for the behaviour of artificial intelligence? Can the decisions of machines be evaluated through moral values? Do the boundaries of morality change with technology? Addressing these questions requires not only technological insight but also profound philosophical and ethical reflection. AI is no longer merely an instrument created by humans; it functions as an extension of human intellectual and cultural capacities. Through technology, the human expands cognitive possibilities while simultaneously facing the risk of distributing moral responsibility. Whereas human actions were traditionally judged based on intention, algorithmic decisions lack the very notion of intention. The nature of artificial intelligence is ethically and ontologically different from that of human consciousness. Machines evaluate the boundary between "right" and "wrong" on the basis of mathematical and statistical probabilities, whereas humans determine these boundaries through moral, historical, and cultural experience.

Thus, the problem of responsibility is not merely a legal issue but also a test of moral identity. Today, AI decision-making mechanisms directly affect human lives in areas such as healthcare, the judiciary, education, and the military. The management of this process requires the integration of

ethical principles not only into human behaviour but also into technological systems. The fundamental question here is: Who bears responsibility for the outcomes produced by artificial intelligence–the programmer, the user, or the system itself? Consequently, the development of AI necessitates a re-evaluation of human essence, free will, and the concept of morality. This also reflects a form of metaphysical pressure exerted by technology upon human existence: the human has objectified their own intellect and transformed it into "another entity." As a result, the classical boundaries of morality are destabilized, giving rise to a new philosophical stage–the *techno-ethical era*.

Discussion.

The Philosophical Nature of the Responsibility Problem

In the history of philosophy, the concept of responsibility has been closely associated with human free will and the capacity for conscious choice. Aristotle's virtue ethics, Kant's deontology, and Hans Jonas's "ethics of responsibility" all demonstrate that human behaviour is evaluated through moral criteria because humans possess the ability to comprehend the consequences of their actions (Jonas, 1984). However, these principles become disrupted in the case of artificial intelligence: the algorithm lacks consciousness and moral intention. Nevertheless, it is able to make decisions that affect human life (Floridi, 2013). This situation renders the subject of responsibility ambiguous: who is accountable for a moral failure–the programmer, the user, or the algorithm itself? According to Kant, the moral worth of an action lies in the intention, not in the outcome. Yet AI decisions are outcome-oriented–they compute optimal scenarios but cannot evaluate moral motivations (Bryson, 2018). This creates a new philosophical condition referred to as the "responsibility gap." Jonas (1984) describes this gap as "the ethical problem of technological civilization" and warns that as humanity gains technological power, its sphere of responsibility must expand accordingly. Philosophers in Azerbaijan and Türkiye likewise emphasize the need to reassess responsibility in the context of advancing technology. For instance, Aliyev (2019) argues that the moral evaluation of AI decisions cannot rely solely on legal frameworks but must also incorporate human ethical and social responsibility. Similarly, Yıldırım (2021) notes that the rise of autonomous systems brings forth a "distributed responsibility" model, which requires an ethical analysis of human–machine interaction. Thus, responsibility must be evaluated not only individually but also in systemic, collective, and techno-cultural dimensions.

Nevertheless, the decision-making mechanisms of artificial intelligence still lack deep moral layers such as intention and empathy. Therefore, the classical philosophical meaning of responsibility must be reconsidered in light of the increasing role of technological agents–an issue that remains complex both legally and metaphysically (Floridi, 2013; Bryson, 2018; Aliyev, 2019; Yıldırım, 2021).

Artificial Intelligence and Moral Agency

In traditional ethical theories, moral agency is explained through the intentional structure of human consciousness. The key criterion here is that a human performs an action on the basis of a particular intention, purpose, and a considered attitude toward the anticipated consequences of that purpose. Kantian ethics links this to rational will and self-legislation, whereas in Aristotelian virtue ethics, agency is explained through the formation of moral character. What unites these perspectives is the assumption that consciousness, intentionality, and the capacity for moral judgement are necessary conditions for moral agency.

For this reason, the question of whether artificial intelligence (AI) systems can possess moral agency long appeared to be closed. Existing AI systems neither form their own intentions nor construct ethical criteria through conscious, experience-based decisions. However, in recent decades, the integration of AI into complex social environments has significantly increased its "practical impact." Coeckelbergh (2020) argues that grounding moral agency solely in internal psychological structures is no longer sufficient, as modern technological systems have begun to exhibit behaviours that produce real moral consequences in the world.

The Rise of Autonomous Systems and Functional Agency

In many contexts, the decisions of autonomous systems are formed outside direct human control. Autonomous vehicles make life-critical decisions in seconds; military drones can identify targets and carry out operations; medical diagnostic systems analyse patient data and generate clinical recommendations. In these situations, AI behaviour is not merely an algorithmic output but a decision with real implications for human life. Consequently, many philosophers and technology ethicists describe such systems as "functional agents." Even without intention, the presence of effects and outcomes makes it necessary to keep them within the scope of moral evaluation.

Braidotti (2019) notes that technological agency is not merely a technical process; it is a unique mode of action that reshapes social relations, creates new power dynamics, and influences human behaviour. This perspective belongs to the posthumanist framework, which rejects limiting agency to humans alone. Social reality, in this view, is formed through a wide network of humans, machines, infrastructures, and information flows. Since AI produces social consequences within this network, it demonstrates a degree of "impact-based agency."

Dialogical Agency and the Sharing of Responsibility

Coeckelbergh (2020) explains AI agency through a "dialogical agency" model. In this model, the human–machine relationship is not merely one of tool-use but a dynamic interaction in which outcomes are co-produced. The human assigns tasks to the machine, yet the machine's behaviour and the resulting outcomes arise from both human input and the system's own functional agency. Here, the question of shared responsibility becomes central: who is responsible – the programmer, the user, the system, or all participants collectively?

According to the dialogical agency model, conscious intentions are not a prerequisite for AI to be considered a moral agent. What matters is its capacity to generate outcomes within a social and normative framework. Thus, the moral value of operational results applies not only to the human but also to the system's activity. However, this model has been criticized. Critics, including Hasanov (2021), argue that intention is a fundamental category in ethics, and evaluating agency solely based on outcomes can lead to flawed conclusions. In their view, moral agency cannot be grounded merely in functional results, since moral evaluation requires consideration of motivations behind choices.

The Agency Spectrum and Semi-Autonomous Systems

Contemporary ethical discussions reveal that agency is no longer understood through a binary model–either present or absent. Many scholars now conceptualize agency as a spectrum. At one end of this spectrum lie fully conscious humans; at the other, purely mechanical and deterministic systems. Artificial intelligence occupies an intermediate position – as "semi-autonomous agents." These systems are neither fully autonomous nor fully under human control. They co-produce adaptive behaviours with humans, respond to new information, modify previous patterns, and thus do not follow a wholly predetermined operational trajectory.

Autonomous vehicle decision-making in crash scenarios is a clear example. In critical moments, a driverless car must decide whose safety to prioritize. Although such decisions are pre-programmed, real-world conditions are so variable that the system resorts to adaptive responses. This demonstrates that agency cannot be attributed solely to the programmer or the user.

Local Philosophical Perspectives

Azerbaijani philosopher Aliyev (2019) emphasizes that moral agency should not be defined solely through consciousness. In his view, if a system makes decisions that affect human life, it is inevitable that such activity carries a certain degree of moral significance. Thus, agency must be reconceptualized in a functional and outcome-based form. This approach also requires the creation of new normative models within legal systems. Three main criteria become central here: the distribution of responsibility, the system's degree of influence, and its level of autonomy. If a balance is not established among these criteria, both legal and ethical gaps emerge.

The Shared Agency Model: The Ethical Framework of the Future

Yıldırım (2021) presents what he considers the most realistic future approach as the "shared agency model." In this framework, the human and the AI system function as two actors within the same operational structure. The human defines the normative framework, goals, and desired outcomes, while the AI generates optimal decisions aligned with those goals. This interdependence gives rise to a multilayered structure of responsibility. Responsibility is no longer attributed to a single agent but is distributed across the entire system's operation. This approach necessitates new regulatory mechanisms at legal, ethical, and technological levels. Such interaction also requires rethinking the role of the human as an agent. The human is no longer a controller or final decision-maker but becomes an actor who coordinates, frames, and defines objectives. While this may seem like a weakening of human agency, it is in fact the transformation of the human role within new technological networks.

It should be noted that whether artificial intelligence can be considered a moral agent remains a contested philosophical issue. However, it is clear that the real-world consequences, decision-making power, and social impact of AI systems do not allow them to be excluded from moral discourse. The notion of agency is shifting from a human-centered framework toward a distributed, multi-directional, and outcome-focused phenomenon embedded within technological networks. This indicates the necessity of new methodological and ontological inquiries within future ethical theories. The central ethical question is increasingly becoming: *How should we assign moral status to systems that do not require consciousness but nonetheless produce real effects?* The answer to this question will shape not only philosophical discourse but also technology policy, legal regulation, and the societal value system.

The Boundaries of a New Ethics: A Posthumanist Approach

Posthumanist philosophy critiques the anthropocentric worldview constructed within traditional humanism and re-evaluates the interdependence of humans with technological, ecological, and cybernetic systems from a new perspective. According to this view, ethics is not solely the product of human experience, intention, or reality; rather, moral responsibility must be reinterpreted within broader ontological networks – ecosystems, machines, information flows, and the relational structures linking human interactions (Coeckelbergh, 2020). From this standpoint, the ethics of the future may shift from "ethics for humans" to a "network-based ethics of responsibility." In this emerging ethical model, values shape not only individual behaviour but also the decision-making mechanisms of technological systems.

The Ontological Rise of Technological Entities

One of the central claims of posthumanism is that technological entities–artificial intelligence, robots, and cyber-physical systems–are no longer merely instrumental tools but "ontological partners" (Braidotti, 2019). This should not be interpreted as the loss of human superiority. On the contrary, this perspective expands the scope of human responsibility and calls for the sharing of moral obligations with new actors. Latour's (2005) "actor-network theory" demonstrates that social reality arises from the joint activity of humans, technical objects, infrastructures, data flows, and natural elements. Within this framework, ethics can no longer be restricted to interpersonal human relations. Posthumanist ethics forms a dynamic and mutually influential system composed of human–machine–nature interactions. In this system, the human begins to understand themselves not as a central being but as one element of a broader ecosystem – a "node of relations."

Infosphere and the Expanded Ethical Space

Floridi (2013) approaches posthumanist ethics through the lens of the philosophy of information and describes this process as the "expansion of the infosphere." The infosphere is the space in which all informational entities interact–humans, machines, data structures, platforms, and algorithms all participate in this domain. Within this framework, the role of ethics shifts from being a mechanism of control to becoming a "normative communication system" that organizes coordinated interactions. Responsibility is no longer an individual matter but assumes a systemic character: programmers, users, platform owners, policymakers, and even, to some extent, artificial intelligence agents become actors within the network of ethical relations.

In this expanded ethical space, human responsibility extends beyond regulating one's own behaviour; it also encompasses accountability for the social impacts of technological algorithms. For example, while human oversight is needed to prevent AI systems used in healthcare from making diagnostic errors, the system itself also generates moral value within the boundaries of its functional agency.

The Human as an "Ethical Node"

One of the most far-reaching interpretations of posthumanism describes the human not as a fixed biological being but as an "ethical node" composed of multilayered connections. From this perspective, technology is neither a mere extension of the human nor an entity that replaces it; rather, it is a partner that co-creates a new ontological reality alongside the human. As artificial intelligence becomes increasingly involved in decision-making processes in medicine, security, governance, and education, human responsibility is defined not only at the level of oversight but also at the level of co-creation. Today, the human is no longer a solitary decision-maker; they are a "co-actor" who develops decisions together with machines. This transformation does not weaken human agency; instead, it restructures and renders it more complex.

The Posthuman Subject and a New Moral Configuration

In her theory of the "posthuman subject," Braidotti (2013) argues that technological progress generates a new moral configuration on both bodily and cognitive levels. The posthuman subject is neither a fully autonomous human nor an entirely technical system; it is a hybrid entity emerging from the mutual synthesis of the human body, biology, algorithms, data flows, and emotional response models. According to Braidotti, the ethics of the future will not be determined solely by the preservation of humanist values. Moral systems will be enriched by new forms of empathy generated through technological partnership, cyber-social relations, and algorithmic reflection. This issue extends far beyond AI's ability to simulate empathy: in this context, AI is regarded as a "participant" in the ethical system, and its behavioural logic is linked to normative frameworks.

New Requirements of Posthumanist Ethics

Posthumanist ethics is not only a theoretical framework but also a set of emerging approaches with practical implications:

1. Development of algorithmic empathy models – the integration of emotional and social data into algorithms.
2. An expanded system of responsibility – responsibility for one's own behaviour as well as for the behaviour of technological systems one uses.
3. The principle of co-agency – the joint formation of decisions by humans and machines.
4. Integration of ecological and technological ethics – AI must consider not only social but also ecological impacts.
5. Cultural posthumanism – a new social and moral culture shaped through human–technology interaction.

Azerbaijani researcher Aliyeva (2022) emphasizes that although the moral status of the human changes in the post-humanist era, responsibility does not become obsolete. In her view, the most significant ethical model of the posthuman period is the "ethics of human–machine collaboration." This model demonstrates that in the future, artificial intelligence will carry not only technical functions but also social ones. This, in turn, requires both the development of new empathy models and the integration of emotional data into algorithms. Aliyeva notes that the transfer of part of the human moral burden to technological systems does not signify a loss of responsibility; rather, it reflects the redistribution of responsibility. Such redistribution necessitates the emergence of new structures in ethical, legal, and social regulation.

Cyber-Ethical Culture and the Normative Framework of the Future

The expanded posthumanist perspective indicates that the ethics of the future will be less a fixed set of rules and more a dynamic culture formed through continuous social negotiations among humans,

machines, and ecosystems. Ferrando (2019) refers to this culture as "cyber-ethical culture." In this environment, technology functions not merely as an object of control but as a moral partner. Future societies cannot construct ethical reasoning solely on anthropocentric grounds, as technology already plays the role of an autonomous actor that transforms the essence of social relations. Therefore, the boundaries of future ethics will be shaped by principles of mutual empathy, distributed responsibility, and co-evolution.

Posthumanist ethics demonstrates that the ethics of the future will not take shape around rigid humanist structures but will emerge within the context of expanded ontological relations. This approach portrays the human not as powerless before technology, but as a responsible participant in technological processes. The new ethical framework co-created with artificial intelligence does not weaken human agency; on the contrary, it renders it more multidimensional, networked, and systemic.

Conclusion. The development of artificial intelligence expands the boundaries of ethics and redefines the essence of human responsibility. Whereas responsibility was traditionally associated solely with the conscious individual, today it has transformed into a complex structure distributed among technological systems, programmers, and users. From a philosophical perspective, the fundamental question remains unchanged: *What does it mean to be moral?* However, the answer is now sought not only in human behavior but also in the behavior of the artificial systems humans create. Contemporary technological reality demonstrates that artificial intelligence is not merely a technical instrument but is increasingly becoming an active participant in social and moral relations. Nevertheless, it is still impossible to attribute full moral responsibility to machines, as they lack qualities inherently tied to human moral agency–such as intention, empathy, and ethical reflection. Therefore, future ethical systems must aim to balance human oversight with technological autonomy. The principal challenge of the new era lies in the deep integration of ethical principles into artificial intelligence decision-making processes. Otherwise, technology risks drifting away from human values and turning into an unregulated source of power. Moral evolution is now measured not only through individual responsibility but through the formation of collective and systemic responsibility. This represents one of the greatest intellectual and moral challenges facing both philosophy and science.

Thus, the renewal of moral reasoning in the age of artificial intelligence is indispensable. The ethical role of the human being is no longer limited to making decisions but extends to serving as an "architect of values" who determines the moral direction of technology. This constitutes a fundamental philosophical and cultural task for shaping the humanist society of the future.

References:

1. Arslan, M. (2022). *Yapay zekâ etiği ve sosyal adalet sorunları*. İstanbul: Beta Yayıncılık.
2. Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
3. Braidotti, R. (2013). *The posthuman*. Polity Press.229 p.
4. Braidotti, R. (2019). *Posthuman knowledge*. Polity Press.240 p.
5. Bryson, J. (2018). Patiency is not a virtue: AI and the design of ethical systems. *Ethics and Information Technology, 20*(1), 15–26.
6. Coeckelbergh, M. (2020). *AI ethics*. MIT Press. 248 p.
7. Əliyev, R. (2019). *Süni intellekt və məsuliyyət: Azərbaycan konteksti*. Bakı: Nurlan Nəşriyyatı.
8. Əliyeva, S. (2022). *Posthumanizm və etik məsuliyyətin transformasiyası*. Bakı: Elm Nəşriyyatı.
9. Ferrando, F. (2019). *Philosophical posthumanism*. Bloomsbury Publishing. 296 p.
10. Floridi, L. (2013). *The ethics of information*. Oxford University Press. 357 p.
11. Həsənov, E. (2021). *Texnoloji etik düşüncə və süni intellekt problemləri*. Bakı: Elm Nəşriyyatı.
12. Jonas, H. (1984). *The imperative of responsibility: In search of an ethics for the technological age*. University of Chicago Press. 263 p.
13. Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. Oxford University Press. 301 p.
14. Yıldırım, S. (2021). *Otonom sistemler ve etik sorumluluk*. İstanbul: Beta Yayınları.