The Ethics of Responsibility in the Context of the Use of Intelligent Machines and the Problem of the Technosystem

Mariusz Wojewoda

Abstract This article is devoted to the problem of responsibility as it arises in the context of the technosystem—where the latter is enhanced by the work of intelligent devices. In this case, the system in question refers to the relationship between man, functioning in the various roles of creator, trainer, owner and user of intelligent agents, and machines equipped with artificial intelligence. Suppose we assume that the aim of technological development is the well-being of the human being living now and the human being of the future. In that case, attention must be focused on the value of responsibility. The system of technology is embedded in a cultural and social context. The author diagnoses the causes of the disappearance of responsibility (the adjacency of actions) in the context of using smart devices, and considers what should be done to counteract this. The background for the analyses undertaken is furnished by ecosystem theory, together with the related concept of instrumentalization as construed by Andrew Feinberg, and Hans Jonas' ethical theory of responsibility—considered as they relate to the analysis of selected cases.

Keywords adiaphorization; artificial intelligence; Andrew, Feenberg; Hans, Jonas; moral responsibility; system of technology; technical phronesis

✍ Mariusz Wojewoda, University of Silesia in Katowice, Poland ✉ mariusz.wojewoda@us.edu.pl
ORCID 0000-0003-0732-7500

Forum Philosophicum 30 (2025) no. 2, 199–213 Subm. 15 January 2024 Acc. 25 March 2025
ISSN 1426-1898 e-ISSN 2353-7043 DOI: 10.35765/forphil.2025.3002.10

> The technosystem is based on a specific conception of life on which we cannot avoid taking a stand, whether consciously and explicitly, or passively in submission to the uncontested facts. . . . Progress is not technical or moral but technical and moral. In a society based on technical rationality the process of transcendence must itself have a rational structure—it must make sense in technical terms just as technical change must make sense in moral terms.
> (Feenberg 2017, 220)

Introduction: Intelligent Machines

In today's world we are increasingly—and willingly—making use of devices equipped with artificial intelligence. This applies to various areas of human activity. The design of intelligent machines means that we obtain a brilliant technical 'partner' for human activities in business, science, medicine and institutional management. It is also an essential 'companion' in public transport, tourism and entertainment activities. Intelligent machines use artificial intelligence (AI) in their operations. However, it is not easy to define this clearly. AI is a collection of different technologies that allow data mining, exploiting the potential of deep machine learning, natural language processing, artificial neural networks, logic programming, automation of decision-making processes, and a whole range of virtual and machine elements (Konotos 2021).

In the context of AI, 'intelligence' means dealing with an algorithm created on the basis of a dataset containing exemplary behavioural models. The program independently searches for relationships between the data and proposes solution patterns and specific courses of action (Przegalińska and Oksanowicz 2023). The speed of information transfer between humans and intelligent devices, together with intelligent machines (the Internet of Things), plays an important role here. We currently use 'weak/narrow AI' to solve specific problems. Its specialized skills allow it to perform certain tasks better than humans. Work is underway to create 'strong/general AI' with multi-threaded knowledge and cognitive abilities that will enable intelligent machines to solve functions at a more advanced level than humans. The next stage in the development of AI is expected to be self-aware superintelligence—the 'singularity'—which will be capable of making conscious choices. So far, we remain at the level of 'weak AI-ANI,' but work is underway to create 'strong AI-AGI' (Boden 2018).

The use of AI is linked to the concept of an 'expert system.' Its operation consists of emulating the decision-making process executed by a human expert in a given field. Emulation does not mean simulation of the decision-making process, but rather indicates a ‘parallel’ process, so to speak, alongside that of human decision-making. Expert systems use databases to ‘extend’ our scope of action. They can significantly help humans optimize the decision-making process, and improve companies’ level of business efficiency (Jackson 1998).

When referring to intelligent machines, we use the term ‘agent.’ This denotes an ‘entity’ operating in a specific environment, capable of communicating, checking, and reacting to environmental changes. Intelligent machines can be thought of as (1) agents that extend our ability to act, (2) agents equipped with a certain degree of autonomy of action, or (3) systems whose task is to optimize human actions.

For example, an intelligent agent can help manage information by searching for it in online resources and selecting and filtering data. An intelligent avatar can represent a user in cyberspace, facilitate online commerce, or facilitate business management. Another critical aspect of the use of AI is in medicine. Here we have in mind algorithms for collecting information on patients’ health. For this purpose, we can use smartphones with applications that allow instant information on the changing state of this. ‘Speed’ of access to information will enable us to undertake effective treatment options. An intelligent program can help create an effective vaccine. An exciting application of AI can be seen in the use of algorithms to find potential organ donors and recipients requiring a specific tissue match (Fry 2018). This is also when the speed of access to information matters—if we are to take effective medical action promptly.

The intelligent agent is unaware of the subjective distinctiveness of its existence (Floridi and Sanders 2004). From the perspective of cognitive science, it is recognized that humans are beings that are conscious of their existence, but we ourselves do not know why we are aware of anything (Dehaene 2014). In this dreamlike scenario, it is not easy to entertain the thought that we could create an artificial consciousness. The latter could just happen by accident, uncontrollably, thanks to the actions of some programmer. This is one of the deepest human fears relating to our technological future. When we talk about humans, we link consciousness with free will. We do not equip intelligent machines with an artificial will to act. By analogy with free will, we speak of ‘machine autonomy,’ especially when we want to draw attention to the operation of intelligent devices with limited or no human supervision (Wojewoda 2023).

Where intelligent machines are concerned, it is necessary to distinguish intelligence from consciousness. Robots, without being conscious, can be intelligent. They can be taught to follow legal and ethical rules that forbid endangering humans (Lin et al. 2017). However, it is challenging to expect intelligent machines to solve moral dilemmas when choosing between values of similar importance, such as life and truth. For example, a clever medical diagnosis robot can be a valuable assistant for medical services. Oriented towards the value of providing truthful information based on the patient's health data, it will inform the patient that their condition is dire, and that there is no point in further treatment. This can then trigger a radical mental crisis and suicidal thoughts on the part of the patient. However, such consequences clearly conflict with the affirmation of the value of life (Alowais et al. 2023). Doctors are expected to show sensitivity and be capable of communicating information to patients about their condition in ways that are appropriately nuanced. Currently, only humans can solve dilemmas and communicate with patients based on a specific understanding of values and the complex nature of what is involved in realizing them. Creating intelligent yet conscious machines (AGIs) will probably render them capable of solving human dilemmas. The acquisition of consciousness by AI will involve further challenges. The next step will be to assign rights to conscious machines similar to human rights—such as the right to have their existence protected by not being disconnected from a power source, or to have their mental health preserved in the context of possible software changes. It is difficult for us now to imagine the complexity of the problems that would have to be solved in this situation.

The autonomy of intelligent machines is a matter of degree. First, machines are not autonomous when it comes to their energy source—they must be powered. Second, they serve to 'complement' and 'extend' human capabilities. Third, it is the human being that determines the specific goals and how the machine will achieve them. Fourth, the human being sets the general objectives, while the machine itself selects the ways and circumstances of their implementation: so the human still controls task performance, but only at the level of overall objectives. Fifth, the machine defines its own tasks and chooses how to realize them. The human only deals with the result of the task performed (Glaser and Rossbach 2011). The autonomization of machines, to be sure, offers excellent opportunities for humans, and does not in and of itself generate any novel threat. A revolt by robots will be possible, though, once they become fully autonomous.

Machines equipped with AI are not conscious, so it is difficult to assign a moral characterization (good/bad, right/wrong) to their actions. However, decisions made by humans with significant participation from intelligent machines are ethical. It is now recognized that intelligent machines are not axiologically neutral. Things, including intelligent things, are bearers of specific values. Indeed, the realization of values depends to a large extent on such bearers. The choice of the bearer with which a value is to be realized is something that can influence the moral decision-maker's reading of the situation. For example, using AI-equipped drones fundamentally changes the nature of warfare. It distances the user of the tool from the consequences of the action. The introduction of intelligent tools for various activities affects the assessment of the user's skill—we can perform many activities faster and more precisely. At the same time, being quicker and more efficient does not necessarily mean it is an instance of morally good use.

When using technical devices, we focus on usability and functionality in achieving our intended purpose. In assessing specific situations, other values should also be considered, such as life, one's health, or the health of others, as well as the value of the common good or the good of humanity. Values of this kind provide a meta-level for reading values such as functionality or effectiveness. Value can be understood in three ways: (1) substantively, as an entity in its own right; (2) attributively, as a feature of a thing; (3) in terms of utility, as the equivalent of something handy to us (Fleischer 2010). Values concern the relational sense of the connections between creators, technical objects and their users. From the point of view of axiological relationalism, values are an essential part of our cultural endowment; they are the backdrop for our individual and social valuations (Wojewoda 2022). Consciously accepted moral responsibility is linked to, among other things, the value of life and health, and concern for the well-being of humanity understood as concern for the future. Responsibility opens up a meta-level with respect to our appreciation of the value of life, health and the common good. For example, we link the functionality of a piece of equipment to concern for the quality of life.

The problem to be posed here is the following: in what sense does the technological mediation of action through intelligent tools distance humans from the consequences of decisions? In this article, the issue of responsibility as it pertains to intelligent machines will be analysed from the perspective of Andrew Feenberg's technosystem theory and Hans Jonas' ethics of responsibility. The idea is to diagnose the related depersonalization of decision-making and blurring of moral responsibility affecting savvy machine users.

The topic of the ethical use of intelligent machines already has a considerable literature. It is primarily concerned with developing ethical and legal principles for artificial intelligence (Lin et al. 2017; Coeckelbergh 2020). The traditional model for constructing rules is based on Isaac Asimov's concept of ethics for robots. In his view, ethical rules form a set of rules similar to legal procedures or instructions created by a superior (manager) for a subordinate (employee). This relationship resembles the relationship between master and slave (Asimov 1950). However, this solution is inadequate. Therefore, solutions are also being formulated to teach intelligent machines human values and equip them with empathetical abilities (Minsky 2007). This idea of ethical principles for AI is complex and multi-stage. However, it requires the creation of an informed artificial intelligence—a powerful AI or superintelligence (Bostrom 2014; Floridi 2023). The present author seeks to draw attention to yet another theme, which concerns the use of intelligent machines inside organizational structures. The use of intelligent machines can help enhance the efficiency and effectiveness of an institution's operations. However, an intelligent system can also itself become a precise tool for improving an organization's level of efficiency. Andrew Feenberg's conception deserves attention in this regard, and will be discussed below.

THE TECHNOSYSTEM IN ITS SOCIAL AND CULTURAL CONTEXT

We can construe the term 'technique' (*techne*) in four ways: (1) as tools, these being ones that did not arise naturally in nature; (2) as the ability to perform certain activities using the tools just mentioned (Heidegger 1977); (3) as products of action, in the sense of technical artefacts (machines, equipment), as well as procedures and rules specifying a certain correct model of acting (Lizut 2014); and (4) as a system of technology. This last way of understanding technology points to the close relationship between man, civilization, and the products of technology (Dusek 2006). This dependence involves the environment, culture, and individual humans. It is also a two-way relationship—we are constantly changing and improving the tools we use and seeking new ways to employ them. The use of technical tools ultimately affects us: we change at the level of our functioning in the world and our social understanding of reality. By using such tools, we become participants in a technical way of life, and even define our own very identity through practices relating to their use (Feenberg 2006). We are currently experiencing such change with a particular intensity. The knowledge society needs constant and fluid access to information, and information that is appropriately selected. The products of technology and their use influence social, cultural and political change.

The issue of systems of technology was addressed in the 1960s by the French philosopher of technology Jacques Ellul. He mainly pointed out the dangers associated with the influence of technology on the human understanding of reality. He believed the technical mindset was fundamentally directed towards calculating and achieving specific goals. The reason for introducing technical thinking into organizations was to eliminate what is accidental in human action and to anticipate future risks. What we have in mind here is not a full-scale anticipation of the future, but the preparation of procedures for dealing with complex crises. According to Ellul, the fundamental purpose of setting up a system is to speed up and optimize the decision-making process and develop mechanical habits in employees. In his view, what is human in the decision-making process is related to arbitrariness and the influence of emotional factors. Rationalizing the decision-making process means that we eliminate such subjective and arbitrary factors from human agency. Rationality is understood here as the pursuit of a technical model for exercising organizational power. For example, systemizing activities involves subjecting employees' behaviour to procedures, and a specialized model for assessing task performance.

In Ellul's view, systemic organizational management is supposed to lead to a mechanization of human behaviour that is not pleasing to employees but is helpful for the organization. However, this reveals a critical danger in the form of the adiaphorization of action and the disappearance of the sense of responsibility (Ellul 1980, 35–40). Adiaphorization (from the Ancient Greek *adiaphoron*, meaning 'morally indifferent act') consists in a subject's making a decision mediated by technical and systemic rules, treating their action as based on pragmatic utility and at the same time as morally indifferent. Sigmund Bauman analysed this issue, seeing it as an oppressive tool of the bureaucratic model of exercising power in an organization. The disappearance of any ethical consciousness associated with the work performed means that the decision is no longer analysed from a moral perspective. Instead, efficiency, functionality, and the achieving of operational goals are introduced (Bauman 1991, 440–42). These are values reflecting a qualitatively lower level of axiological engagement, and in this way employees become cogs in the organization's machine. Moreover, Hannah Arendt wrote about adiaphorization in similar terms, understanding it to be the cause of moral indifference to the mass extermination of the Jewish population. Adiaphorization causes the individual to embody a mechanical attitude, such that they follow orders/commands and are not responsible for the consequences. This involves an attitude of self-justification and moral indifference (Arendt 2003, 62–63).

An analysis of the system of technology from a social and cultural point of view has also been developed by the contemporary philosopher Andrew Feenberg. He distinguishes four main ways of understanding technology:

(1) as technical determinism, in which technological creations fundamentally influence human action and lifestyles; (2) as socio-cultural determinism, in which technology is adapted to social, political, and cultural changes; (3) in terms of interactionism, where it is assumed that technology and users constitute a specific whole within which it is difficult to determine which element is dominant; and (4) critical constructivism, in which the system entails the close interdependence of humans and technology, but is also so malleable that it can be modified (Feenberg 2009). In his later research, the American philosopher himself adopts this last position.

Feenberg recognizes that human activities, including technology-related ones, are embedded in a world of values. Some of this involves moral values, based on which we consider human behaviour right or wrong. The author of the book *Technosystem* poses an important question: does not technology, or its use, lead to the disappearance of our moral sensitivity?

Among other things, Feenberg analyses the value of the utility of technical tools. The subjective aspect of usability is based on the belief that using the device in question is correct. In contrast, the objective element of usability concerns knowledge of the device's technical capabilities. In this case, practical rationality is based on justifying for what purpose and to what extent the device should be used appropriately in a specific social, business, medical or political context. The purpose of use depends on the characteristics of the object itself, the individual beliefs of the subject, and also on culturally and institutionally approved ways of using it (Feenberg 2017).

This goal is linked to our mental map and our systemic imaginaries. Such imaginaries relate to the capabilities inscribed in the machine, the subjective skills involved in using these devices, and the social imaginary in which the creator and user of the device operate. For example, the smartphone we use is constantly equipped with new affordances. We can use them meaningfully when we have mastered specific skills, and when we function in a developed technological society where individuals and institutions have produced appropriate strategies for using such devices. Here, forms of essential and occasional use should also be considered. Understood as a system, technology is more than the sum of its constituent parts: there is, in addition, whatever quality it has, *qua* part, that results from our human relationship with *techne*-related creations. It is generally recognized that we do change under the influence of technology, and certainly we are mindful of how we use devices now and how we used them in the past.

The culture surrounding *techne*-related creations fulfils an important role: namely, it points to a system of meanings and practices of technical use developed within a community. It is, therefore, not only about technical feasibility, but also about the social justification and psychological credibility of a particular use of *techne*-related artefacts. For example, we usually assume that we are safe using an app to carry out banking operations via a notebook or smartphone. In this case, we rely on our own 'common sense,' as well as a socially entrenched belief that such an action is right. Critical constructivism assumes that the use of technical tools depends on a cultural and axiological framework. The function of the tool is subject to interpretation. It depends on the technological imagination, socially established habits of behaviour, and approved action practices. Nowadays, the selection of artefacts of technology is primarily made within the framework of institutions, associated procedures of action, and specific technical disciplines.

Feenberg refers to 'instrumentalization theory' to clarify the question of the functionality of contemporary technological artefacts. Identifying affordances requires a decontextualization of *techne*-related objects. Here, we have two levels of reflection—the technical and the cultural. In the process of interpretation, we determine the meaning of artefacts by referring to legible values. For example, the usefulness of a thing means that we, as users, select the rules of its effectiveness. We do this because of our expectations. User preferences are embedded in an axiological background that determines in which situations a thing counts as functional, safe and sensible and under which conditions as a form of misuse—e.g., when it leads to dependence on the artefact. Values provide a cultural background giving credence to human practices. For example, when we use digital tools to work online, we recognize that this kind of work is just as productive as working in the office or at university.

In the process of interpretation, we consider the AI-equipped artefacts, their creators (in the case of AI, we are talking here about programmers and skills trainers), and the users of these tools. The meanings within these relationships are not fixed once and for all: every so often, we reinterpret our reading of the relationships between them (Feenberg 2017). In addition to the technical dimension, the design of tools also involves a recontextualization of their meaning. This process then equips the devices with new capabilities and meanings to justify their use. The next stage is to change our mental map, modify our conceptual models, and create imaginary systems for the effective and meaningful use of our digital tools (Norman 2023). Behind all these technical proposals lies a set of background axiological commitments, which must often be articulated. This implies, among other things, a focus on improving the speed of access to information, communicative convenience, and a concern to be in control of one's health. It is not just a matter of suggesting new possibilities related to the use of digital devices, but also of overcoming individual and social fears, changing personal habits, modifying the social and mental map, and creating a socially approved view that these devices are helpful and at the same time safe—that is, that our sensitive data is adequately protected.

Ways of using tools can result from imitation and the social compulsion to participate in certain practices. The relationship between the social and the technological compulsion to use tools is interesting. On the one hand, the range of professional, institutional activities that an individual can (and is now supposed to) carry out using digital tools is constantly expanding. On the other hand, the spread of these tools may widen the scope of exclusion, pushing large groups of people to the margins of social and professional life (Suchacka et al. 2021). Therefore, it is essential to design responsibly, equip devices with clear markers, and define understandable ways of using technical devices. What we have in mind here is a move away from a model based on compulsion, and toward creating positive habits of use. We should maintain a critical attitude towards technopression—the compulsion to use digital devices. Technological culture should lead to the development of human beings and their social and creative skills, and not to their objectification (Krzykawski 2023).

Techne and Moral Responsibility

Following Hans Jonas, we shall distinguish between moral responsibility and legal liability. The latter concerns compliance with rules laid down by a state, or by international law, while the former, in its legal sense, addresses the negative consequences of a subject's conduct as these relate to some negative state of affairs resulting from their freely chosen actions. Legal liability concerns acts whose significance is determined post-factum—as well as, more generally, ones that have happened with specific consequences (Peno 2015). In this article, we shall focus on moral responsibility. Jonas (1984) highlighted several factors to consider when discussing moral responsibility as it relates to the use of traditional technical tools dating from the second industrial revolution. We, on the other hand, are now at the stage of the fourth and projected fifth industrial revolution (the use AI). In modernizing Jonas' conception, we will relate it to the issue of the use of intelligent machines. It consists of three essential elements.

1) *Reflecting on the future.* Here we assume that we are dealing with probabilities of future events. Analysing scenarios pertaining to such possible courses of events prepares us for their undesirable consequences. As long as the danger is unknown, we do not know how to protect ourselves—in this case, we identify the sources of risk and diagnose the causes

of user misbehaviour (Jonas, 1984). Describing and explaining the causes of wrongdoing is supposed to make us work out scenarios for counteracting the negative consequences of activities related to the use of intelligent machines.

2) *Asymmetry of action.* In defining the rules of moral responsibility, we cannot rely on the law of reciprocity. Reciprocity means that we expect behaviour of a similar kind to our own, whereas the intelligent machine is assumed not to be a conscious subject but rather an agent whose assistance we use. The intelligent agent is an integral part of the expert team, because the information it provides, or its skills, influence the decision-making process and the action strategy adopted. The asymmetry is that we do not expect the intelligent machine to behave responsibly. It is a valuable tool for acquiring and collecting data, but it is the human being who designs, trains and supervises the intelligent machine that is responsible for its actions, even when using AI.

3) *Causal power.* Moral responsibility is linked to the category of causal power. In using the term 'power,' we are describing the impact of an individual or organization on the external environment and the lives of others. Being responsible means that the subject can be considered a conscious executor of their action: in other words, there is a causal relationship between the action and its consequences. By introducing intelligent machines, we increase the power of our impact on the environment. Still, paradoxically, we separate ourselves from the consequences of this action significantly when we systematically increase the autonomy of intelligent machines. This separation makes us systematically lose the ability to analyse the action in moral terms (Jonas 1984).

A good example of the adiaphorization phenomenon—the suspension of responsibility in the context of the use of technical tools equipped with artificial intelligence—is the Lavender program created by the Israeli army to eliminate Palestinian Hamas or Islamic Jihad fighters (Brigadier General YS 2021). The Israeli military used AI to mark Gaza zones as suspected of carrying out attacks against Israelis. This included actual militants, but also those individuals who were Hamas or Islamic Jihad sympathizers. Human oversight of combat drone operations was reduced to a minimum. The machine was given significant autonomy regarding who to eliminate and what places to bomb. As a result of the indiscriminate bombing, many innocent people were killed who were in the firing zone of the bombs being dropped. The human supervisor of the Lavender system was limited to determining the algorithm of people to eliminate, while agreeing to kill random victims of the attack. The Israeli army decided to use the intelligent system even though it made the wrong decisions in about 10 percent of cases. Sometimes, it flagged as dangerous individuals people who had only a loose connection to militant groups. The operation of the Lavender intelligence program is an example of the systematic narrowing of the field of human responsibility through smart devices. Transferring this responsibility to algorithms means there is no connection between the person making the decision and the consequences of his/her action (Yuval 2024).

What gives moral responsibility its specificity is that it does not seek to analyse the consequences themselves, but rather to consider the overall well-being of the persons associated with the action taken. Practical rationality involves considering what needs to be done to avoid moral evil, or how to restore the good that has been violated. We are concerned with identifying irresponsible behaviour in advance, as it were, before it is actualized. From this perspective, our aim is to consider possible consequences, not those already transpiring. Irresponsibility is an omission or failure to undertake certain activities: one from which negative consequences may arise. We do not expect moral reflection in this sense from intelligent machines. It is a peculiarity of human thinking and human ethical sensitivity. In a root sense, moral responsibility stems from the awareness that human well-being may be endangered by my actions or my inaction. The discovery of the causes of this threat makes it necessary to engage in activities to counteract it. We are not yet in a position to algorithmize this kind of awareness and transfer it to machine activity.

CONCLUSION

Technological mediation makes it increasingly difficult for us to identify the link between a conscious choice, the resulting action, and the consequences of that action. Autonomous devices distance the user from the consequences of their actions primarily when the machine determines the specific goal and the means of achieving it. It is essential to recognize that no command can enforce moral responsibility where the latter is based on voluntarily accepting a commitment to responsibility. While it is true that we are accountable for someone or something to someone or an institution, the key is the conscious acceptance of a commitment to a particular action. Moral responsibility is about acting freely and consciously. We do not expect this kind of action from animals or intelligent machines. When we think of responsibility, we consider the subject's responsibility for the intentions and consequences of their actions. With this in mind, Feenberg invokes the concept of *phronesis* from Aristotle—practical reason associated with the subject's trained ability to make accurate moral choices (Feenberg 2017).

Currently, it seems that the exercising of such a skill only takes place in a cultural and social vacuum. We need a ‘frenetic background’ embedded in the ecosystem, in which the creation of functional and valuable machines is linked to responsibility for their use. A background of this sort often goes unarticulated, but this is surely a condition for any rational debate over the extent of individual and social responsibility for the rules governing the development of smart devices. In this sense, we need an institution with an appropriate organizational culture, in which the laws of responsible behaviour are followed out of habit—technical phronesis.

When analysing the issue of responsibility in the context of the creation and use of intelligent devices, we are talking about the responsibility of programmers, trainers, robot builders, owners and users. The design and use of intelligent machines should be linked to responsibility for our quality of life and concern for future generations. Referring back to Jonas’ thought, when increasing the power of our actions through the design and use of intelligent artefacts we cannot overlook the issue of responsibility for the possible consequences of our actions. On the other hand, Feenberg emphasizes that the technosystem’s coherence not only refers to the utility of things but also has a normative significance: i.e., it speaks about what the system should be like.

In addition to practical values, a human-friendly technical system must consider values at the meta-level, such as concern for the quality of life, and responsibility for the well-being, of both those currently living and future generations. In a moral sense, practical reason must be employed not only to seek new solutions but also to define the limits of a model of technological development that otherwise would be undesirable for humanity. The social acceptance of intelligent machines should not lead its designers and users to lose sight of critical thinking and the ability to make independent moral judgments. If AI is an ethical tool, it is because its programmers, trainers and users are honest. If its development is done without analysing moral issues, it could lead to serious social problems.

The progress associated with the development of AI is exponential, so it is difficult to predict the future course of discoveries, or new applications of artificial intelligence tools. An important issue is the development of ethical rules for AI. Currently, we think of these rules in terms of task regulations. However, this may not be enough: we associate moral responsibility with consciousness, and research into creating conscious AI is undergoing intensive development. AGI—a powerful form of AI/superintelligence—may be achieved soon, and then the task of artificial intelligence developers and trainers will be to create subjective rules of moral action for it.

References

Alowais, Shuroug A., Sahar S. Alghamdi, Nada Alsuhebany, Tariq Alqahtani, Abdulrahman I. Alshaya, Sumaya N. Almohareb, Atheer Aldairem, Mohammed Alrashed, Khalid Bin Saleh, Hisham A. Badreldin, Majed S. Al Yami, Shmeylan Al Harbi, and Abdulkareem M. Albekairy. 2023. "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice." *BMC Medical Education* 23 (689): 1–15. https://doi.org/10.1186/s12909-023-04698-z.

Arendt, Hannah. 2003. *Responsibility and Judgment.* Edited and with an introduction by Jerome Kohn. New York: Schocken Books.

Asimov, Isaac. 1950. *I, Robot.* New York: Doubleday.

Bauman, Zygmunt. 1991. *Modernity and the Holocaust.* Cambridge: Polity Press.

Boden, Margaret A. 2018. *Artificial Intelligence: A Very Short Introduction.* Oxford: Oxford University Press.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies.* Oxford: Oxford University Press.

Brigadier General YS. 2021. *The Human-Machine Team: How to Create Synergy between Human and Artificial Intelligence That Will Revolutionize Our World.* Self-published.

Coeckelbergh, Mark. 2020. *AI Ethics.* Cambridge, MA: MIT Press.

Dehaene, Stanislas. 2014. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts.* London: Penguin Books.

Dusek, Val. 2006. *Philosophy of Technology: An Introduction.* Oxford: Blackwell Publishing.

Ellul, Jacques. 1980. *The Technological System.* Translated by Joachim Neugroschel. New York: Continuum Publishing.

Feenberg, Andrew. 2006. "What Is Philosophy of Technology?" In *Defining Technological Literacy*, edited by John R. Dakers, 5–16. New York: Palgrave Macmillan.

Feenberg, Andrew. 2009. "Critical Theory of Technology." In *A Companion to the Philosophy of Technology*, edited by Jan Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendricks, 146–53. Oxford: Wiley-Blackwell.

Feenberg, Andrew. 2017. *Technosystem: The Social Life of Reason.* Cambridge, MA: Harvard University Press.

Fleischer, Michael. 2010. "Wartości w wymiarze komunikacyjnym." In *Teorie komunikacji i mediów*, vol. 2, edited by Marek Graszewicz and Jerzy Jastrzębski, 9–54. Wrocław: Oficyna Wydawnicza ATUT.

Floridi, Luciano, and J.W. Sanders. 2004. "On the Morality of Artificial Agents." *Minds and Machines* 14 (3): 349–79. https://doi.org/10.1023/B:MIND.0000035461.63578.9d.

Floridi, Luciano. 2023. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities.* New York: Oxford University Press.

Fry, Hannah. 2018. *Hello World: How to Be Human in the Age of the Machine.* London: Doubleday.

Glaser, Horst Albert, and Sabine Rossbach. 2011. *The Artificial Human.* Frankfurt: Peter Lang.

Heidegger, Martin. 1977. *The Question Concerning Technology, and Other Essays.* Translated by William Lovitt. New York: Garland Publishing.

Jackson, Peter. 1998. *Introduction to Expert Systems.* 3rd ed. Boston: Addison-Wesley.

Jonas, Hans. 1984. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age.* Chicago: University of Chicago Press.

Konotos, John. 2021. "Artificial Intelligence, Machine Consciousness and Explanation." *Academia Letters*, Article 1709. https://doi.org/10.20935/AL1709.

Krzykawski, Michał K. 2023. "Krytyka w technopresji." *Śląskie Studia Polonistyczne* 20 (2): 1–28. https://doi.org/10.31261/SSP.2022.20.09.

Lin, Patrick, Keith Abney, and Ryan Jenkins, eds. 2017. *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence.* New York: Oxford University Press.

Lizut, Rafał. 2014. *Technika a wartości: Spór o aksjologiczną neutralność artefaktów.* Lublin: Wydawnictwo Academicon.

Minsky, Marvin. 2007. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind.* New York: Simon & Schuster.

Norman, Donald A. 2023. *Design for a Better World.* Cambridge, MA: MIT Press.

Peno, Michał. 2015. "Prawna odpowiedzialność z tytułu pełnienia roli społecznej." *Acta Universitatis Lodziensis. Folia Iuridica* 74: 42–57. https://doi.org/10.18778/0208-6069.74.04.

Przegalińska, Aleksandra, and Paweł Oksanowicz. 2023. *Sztuczna inteligencja: Nieludzka, arcyludzka. Fenomen świata nowych technologii.* Kraków: Znak.

Suchacka, Małgorzata, Rafał Muster, and Mariusz Wojewoda. 2021. "Human and Machine Creativity: Social and Ethical Aspects of the Development of Artificial Intelligence." *Creativity Studies* 14 (2): 430–43. https://doi.org/10.3846/cs.2021.14316.

Wojewoda, Mariusz. 2022. "Autointerpretacja osoby i problem 'mocnego wartościowania' w filozofii Charlesa Taylora." *Analiza i Egzystencja* 57: 71–90. https://doi.org/10.18276/aie.2022.57-04.

Wojewoda, Mariusz. 2023. "Artificial Intelligence as a Social Utopia." *Ethos* 36 (3): 109–26. https://doi.org/10.12887/36-2023-3-143-08.

Yuval, Abraham. 2024. "System of Lavender." +972 Magazine, April 3. https://www.972mag.com/lavender-ai-israeli-army-gaza. Accessed October 17, 2025.