Research article
DOI: https://doi.org/10.21202/jdtl.2023.22

Ethical-Legal Models of the Society Interactions with the Artificial Intelligence Technology

Dmitriy V. Bakhteev

Ural State Law University named after V. F. Yakovlev
Ekaterinburg, Russian Federation

Keywords

Artificial Intelligence,
ChatGPT,
digital technologies,
ethics,
law,
machine learning,
model,
regulation,
robot,
society

Abstract

Objective: to explore the modern condition of the artificial intelligence technology in forming prognostic ethical-legal models of the society interactions with the end-to-end technology under study.

Methods: the key research method is modeling. Besides, comparative, abstract-logic and historical methods of scientific cognition were applied.

Results: four ethical-legal models of the society interactions with the artificial intelligence technology were formulated: the tool (based on using an artificial intelligence system by a human), the xenophobia (based on competition between a human and an artificial intelligence system), the empathy (based on empathy and co-adaptation of a human and an artificial intelligence system), and the tolerance (based on mutual exploitation and cooperation between a human and artificial intelligence systems) models. Historical and technical prerequisites for such models formation are presented. Scenarios of the legislator reaction on using this technology are described, such as the need for selective regulation, rejection of regulation, or a full-scale intervention into the technological economy sector. The models are compared by the criteria of implementation conditions, advantages, disadvantages, character of "human – artificial intelligence system" relations, probable legal effects and the need for regulation or rejection of regulation in the sector.

Scientific novelty: the work provides assessment of the existing opinions and approaches, published in the scientific literature and mass media, analyzes the technical solutions and problems occurring in the recent past and present. Theoretical conclusions are confirmed by references to applied situations of public or legal significance. The work uses interdisciplinary approach, combining legal, ethical and technical constituents, which, in the author's opinion, are criteria for any modern socio-humanitarian researches of the artificial intelligence technologies.

Practical significance: the artificial intelligence phenomenon is associated with the fourth industrial revolution; hence, this digital technology must be researched in a multi-aspectual and interdisciplinary way. The approaches elaborated in the article can be used for further technical developments of intellectual systems, improvements of branch legislation (for example, civil and labor), and for forming and modifying ethical codes in the sphere of development, introduction and use of artificial intelligence systems in various situations.

For citation

Bakhteev, D. V. (2023). Ethical-legal Models of the Society Interactions with the Artificial Intelligence Technology. *Journal of Digital Technologies and Law, 1*(2), 520–539. https://doi.org/10.21202/jdtl.2023.22

Contents

Introduction

Numerous advantages of artificial intelligence systems (including fast learning ability, ability to solve a wide range of tasks, higher that in humans efficiency) together with their increasing penetration into various sphere of life forces a question whether an artificial intelligence system is (or will be) able to perceive itself as an autonomous personality, independent of developers and users, whether artificial intelligence will realize its advantages over people, how it will estimate its position and what it will do if it wants to change it. These questions are at the intersection of the objective areas of ethics, law, and technology; hence, they should be resolved with interdisciplinary research methods (Kazim, 2021). Depending on the answers to these questions and the degree of technology development, the models/scenarios of social reaction described below are probable and, consequently, the rights to artificial intelligence technologies.

1. The tool model

A tool is a product of human activity and a means for manufacturing other objects, including tools. Karl Marx and Martin Heidegger in their works emphasized the differences between a tool and a machine. The former pointed out that cardinally different notions are often mixed up: while a tool demands immediate participation of a human in the labor process, a machine "supersedes the workman, who handles a single tool, by a mechanism operating with a number of similar tools, and set in motion by a single motive power" (Marx, 2001). He also noted that a machine differs from a tool in sufficient autonomy (mostly a resource one, as a machine is directed and controlled by a human anyway). Martin Heidegger, in turn, relying upon Hegel's works, lists such criteria of an object being a machine as autonomy, self-reliance and independence (Heidegger, 1993). For the present research, we may interpret it as follows: at the stage of development and testing, artificial intelligence within the tool concept serves more as a tool, not as a machine, as its functioning is connected with human activity at three levels: during development (a tool and a machine are similar in this aspect), during implementation of the activity previously inherent exclusively to a human, and during the human control over the results of the artificial intelligence activity. This approach is also sometimes called pragmatic (Morley et al., 2021). The "machine character" of the artificial intelligence is, in this case, a derivative characteristic from autonomy, but the studied model denies informational autonomy of the artificial intelligence systems. Accordingly, within this model artificial intelligence is viewed as a tool, including for implementing the needs of the humanity (Watkins & Human, 2023).

The theory of autonomy (informational and resource ones) described in the author's works (Bakhteev, 2021) wholly correlates with the fact that the artificial intelligence systems may be perceived as machines (which, in turn, is proved by the existence of the terms "machine learning" and "machine vision"), and as entities comparable with cognoscitive biological objects. This said, one should bear in mind that, while a tool (in traditional scientific interpretation) is intended for easing human labor, the artificial intelligence may substitute human labor with its activity. At that, it would be incorrect to compare artificial intelligence with machines of the industrial revolution era. Those machines put a lot of people out of work, but they created new jobs at the same time. In case of automation in general and using artificial intelligence systems in particular, we already observe elimination of certain positions, mainly associated with mediation services: consultants, dispatchers, marketers, etc. Standard industrial robots even now, without intellectual modules, have largely optimized assembly line production, which allowed reducing costs and manifoldly increase the product quantity and quality and put out of work a lot of unqualified and low-qualified workers; actually, we observe a new industrial revolution. By 2020, over 3 million industrial robots were used globally (however, by 2023 the growth rate decreased); intellectual system are being integrated into various spheres of life, which apparently positively influences economy, but incurs certain harm on the society, first of all, by reducing employment. One variant of solving this problem is to legislatively guarantee employment, or to implement retraining programs, supported by the state and large corporations, for those who lost the job.

Development of the artificial intelligence allows it to solve an increasing range of intellectual problems, which creates prerequisites for further cutting jobs in an increasing number of spheres, as well as for elimination of entire professions. "The actual effect of a reduced payroll fund (and the number of jobs. – Author's note) due to introduction of robots is determined by the number of people released and the size of their payroll, as well as the cost of the robots, which, in turn, is determined by the complexity of construction and degree of intellectualization of the robots" (Timofeev, 1978).

Unlike the initial stage of robotization, spreading of the artificial intelligence systems may reduce the employment rate not for vocational jobs only. According to forecasts of experts from Oxford University, in the next 20 years, 47% of jobs in the USA and 77% of jobs in China will be automated. According to one of the leading experts in the sphere of computer facilities M. Vardi, by 2045 about 50% of people will be jobless (Vardi, 2012). One may object that the number of artificial intelligence software developers increases, but the increase of the number of programmers and other persons involved in developing intellectual systems cannot be compared to the decrease of other jobs. Moreover, the desire to create intellectual systems capable of self-replication is also evident, so one may not exclude job cuts in the spheres related to information technologies: for example, ChatGPT in its fourth version can create a simple but compilable code. Thus, in January 2023, Alphabet company announced cutting 12,000 jobs around the world, including due to introduction of various intellectual systems capable of substituting, in particular, marketing specialists, copywriters and illustrators[1]. According to the model developed for analyzing the probability of certain professions disappearing due to the use of intellectual chat-systems, a 100% substitution of humans, at the present stage of the technology development, is possible for mathematicians, taxation specialists, financial analysts, writers, copywriters, web designers, book designers, secretaries of public authorities, and news analysts (Eloundou et al., 2023).

By now, it is the tool model that may be considered the only one with full-fledged implementation: in applied activity, the artificial intelligence systems serve as a tool increasing performance. This, just like during previous technological revolutions, imposes on the state and society the function of preserving jobs and regulating the intensity of the application of the said technology.

[1] Pichar, S. (2023, January 20). *A difficult decision to set us up for the future*. https://blog.google/inside-google/message-ceo/january-update/

2. The xenophobia model

As the discussions on development and use of the artificial intelligence systems activate, the voices of opponents of further research in this sphere become louder. At that, at least a part of them cannot be called obscurants with irrational fear of the technological progress. For example, a world-renowned scientist, popularizer of science S. Hawking said: "…appearance of a full-fledged artificial intelligence may become the end of the human race... Such intelligence will take up the initiative and start improving oneself with an accelerated speed. People's abilities are restricted by too slow an evolution, we cannot contend with the speed of machines and are going to lose". Of a similar opinion is a famous American engineer, IT entrepreneur E. Musk, who stated: "I believe the artificial intelligence will sooner or later kill us all... Facebook*, Google, Amazon, Apple – they all already know a lot about you. The artificial intelligence, which will be created within these corporations, will get an enormous power over people. And concentration of power in one pair of hands always generates great risks". It was also marked that "distribution of functions between an artificial intelligence system and a human must follow the human-centered principle and always leave the opportunity for a human choice. It implies providing human control over working processes within artificial intelligence systems" (Semis-ool, 2019).

These facts determine a need to thoroughly examine the "xenophobia" model of attitude to artificial intelligence. It is worth noting that it develops the tool model along a negative scenario, i.e. both due to the progress in the development and use of the artificial intelligence systems, and to implementation of one or several risks (in the form of instantaneous negative events or long-term crises), as described above.

The term "xenophobia" is formed with two Greek roots: ξένος ("alien") + φόβος ("fear"). Thus, xenophobia is literarily defined as fear, intolerance to something alien, unknown[2].

Researchers are not unanimous about the origins of xenophobia. Some authors mark that it could appear as an adaptation tool during evolution, facilitating survival and transference of genes to offspring. For example, fear of strangers could be, inter alia, based on the observation that aliens could be carriers of new pathogenic microorganism, dangerous for the locals due to the lack of necessary antibodies.

Traditionally, the term "xenophobia" was used to denote fear, distaste for people of other races, nationalities, cultures and religions. However, in our opinion, a research of the process of interaction between humanity and the technological achievements allows using this term to describe a certain type of attitude towards the scientific-technical progress and its fruits – technologies, including, apparently, artificial intelligence.

[2] Ozhegov, S. I., & Shvedova, N. Yu. (2016). *Thesaurus of the Russian language* (p. 300). Moscow: A TEMP.

Finalizing our approach to xenophobia, we should highlight an important aspect that, ultimately, xenophobia is a specific type of fear. According to E. P. Ilyin, fear, as one of many emotions, is "an emotional state reflecting a protective biological reaction of a human or animal experiencing a fake or real danger to their health or wellbeing" (Ilyin, 2016). Further, E. P. Ilyin stated that from the biological point of view fear is, undoubtedly, a useful phenomenon, while for a human as a social creature fear is often an obstacle in achieving the set goals. In this section of our work we research the bases of the potential critical distrust of the society towards the artificial intelligence technology.

The essence of the "xenophobia approach" to artificial intelligence is in viewing it as a real threat to the humanity and its actual position in the world.

A general analysis of the artificial intelligence technology critic makes it possible to identify two main forms of fear (distrust) people fell towards this technology – essential and instrumental.

Essential fear is due to the fact that people fear not the use of the artificial intelligence technology, but artificial intelligence per se as an artificial but quite independent and autonomous intelligence, capable of being, learning, thinking, and perceiving oneself without participation of a human. Emergence of such a "manmade thinking machine", which is capable of thinking not just like a human but better than a human, undermines the human monopoly to cognitive activity which existed during the history of civilization and enabled the humans to take the dominant position among other species. In this respect, artificial intelligence becomes a separate species, which humanity cannot perceive otherwise that a competitive one. At that, it is uncontrollable artificial intelligence that causes fear, i.e., the situation of artificial intelligence gaining self-consciousness as a result of a software break or purposeful actions of a developer. Actually, such situation regarding biological processes should be considered a mutation, but it is doubtful that the processes of technological products development are so much connected with evolutionary mechanisms. That is why the scenario of an "aggressive" artificial intelligence seems extremely unrealistic.

The instrumental fear, in turn, reflects the fear of a human being ousted by the artificial intelligence systems in labor sphere, as was described in the tool model (see above).

After a systematic research of artificial intelligence began in the 1940s – 1950s, i. e. less than one hundred years ago, systems have been created which exceed humans in some types of intellectual activity. Abilities of modern computers still do not allow comprehensive modeling of a human mind or the whole world around, but artificial intelligence can very well cope with abstractions. Games are a good example of abstraction and, what is very important in this case, the results of participating in a game can be accurately assessed. Thus, since as early as the beginning of the 2000s, the world strongest chess players can oppose nothing to a computer; according to G. Kasparov, all professional chess players train by playing against chess computer programs, as a human rival cannot provide a sufficient depth of search. In 2015, a chess computer program for the first time won a human in a Go game – one of the most complex open information games[3], which had been considered impossible. Artificial neural networks are capable of winning professional players of computer games in cybersports[4], which also had been considered an exclusive prerogative of a human.

The content of xenophobia approach consists in that artificial intelligence may be used by individual persons, organization and states as a means to achieve their malevolent goals.

A typical example of this fear is an uproar emerging soon after the US presidential elections and associated with Cambridge Analytica company. According to some sources, this private company, using the latest methods of information collection and analysis in Facebook* social network, obtained a large array of data, including personal ones, in order to develop a special political advertisement which, according to a number of experts, facilitated electing the present US President. Moreover, this organization is accused of participating in interference into the results of over 200 elections worldwide. A former staff member of Cambridge Analytica Chris Wylie marked: "We used imperfect software of Facebook* to collect millions of user profiles and to build models which allowed us to learn about people and use these data to activate their internal demons"[5].

This case shows that even now the abilities of the artificial intelligence systems are used to collect personal data and manipulate public opinion. In future, artificial intelligence can be used to manipulate large amounts of information, forming the world view for the population of entire states, which creates real threats for democratic institutions, freedom of speech and information dissemination. States may use it for imposing a certain attitude to their populations and the populations of other states; representatives of various corporations – for artificially forming demand to certain goods and services. Finally, representatives of the criminal circles may use it to collect confidential information about citizens and organizations, which may further be sold in the shadow market or used for blackmail or fraud.

Another layer of problems is using artificial intelligence in military activity. A notional combat robot equipped with artificial intelligence, or an army of such robots, is an effective

[3] For example, while chess has about 20 moves per turn, the Go game has about 200.

[4] Statt, N. (2019, April 13). OpenAI's Dota 2 AI steamrolls world champion e-sports team with back-to-back victories. *The Verge*. https://www.theverge.com/2019/4/13/18309459/openai-five-dota-2-finals-ai-bot-competition-og-e-sports-the-international-champion

[5] Chereshnev, E. (March 20, 2018). Defenseless data: how Facebook found itself amidst the greatest controversy ever. *Forbes*. https://www.forbes.ru/tehnologii/358883-bezzashchitnye-dannye-kak-facebook-okazalas-v-centre-samogo-bolshogo-skandala-v

substitute for regular troops. It may execute complex intellectual tasks, act in most unfavorable conditions, requires no rest or sleep, and its destruction does not worry much the public opinion in the country waging the war (at least until that war becomes too expensive from the economical point of view). At that, artificial intelligence is able to solve the previously human tasks with an un-human, machine rationality. If wrongly designed, the artificial intelligence system will have no problem in using illegal means of fighting a war, killing civilians, etc.

Another aspect of xenophobic approach is related to phenomena described in the section about the tool approach. The Hollywood guild of script writers, together with thousands of artists and illustrators around the world consider the results of the artificial intelligence systems to be a priori plagiarism, as they do not represent creativity as such, but just a mixture of already revealed meanings. It should be noted, however, that a significant part of human creative works is made along the same lines.

Thus, the xenophobia approach to assessing artificial intelligence cannot be called definitely ungrounded. There are good reasons to fear an uncontrollable development of artificial intelligence technologies and their further integration into various aspects of human life. This approach, implying either strict control over research in this sphere or, in a more radical form, their complete rejection, is apparently not free from shortcomings. These include: hindering the scientific-technical progress, impossibility to optimize people's practical activity by using the artificial intelligence systems, hap between scientific achievements and their integration into practice, hence, lowering the authority computer science in the public's eyes. An alternative to digital technologies, whose flagman is artificial intelligence, is usually said to be biotechnologies, thus, the rejection of artificial intelligence development, disillusionment in this technology may lead to the development of medicine, physiology, genetics, etc. Nevertheless, we believe that one should deny the advantages of this approach, which implies a more weighted estimation of the technology and its application; elaboration of the tools for forecasting and accessing the risks of further study and use of artificial intelligence, which can be with corrections used in other fields of knowledge; stimulation of development of other sciences related to the development of a human and human potential; providing a new impulse for comprehending a human and their place in the world.

3. The empathy model

According to this model, the society, positively perceiving household and other social intellectual robots and software assistants, favorably accepts the idea of the technology dissemination and does not exclude the possibility to endow the artificial intelligence systems with legal subject properties (in a limited sense).

This model is based on an advanced and broadened sense of humanism and human responsibility not only for themselves but to those around. Intellectual and autonomous systems, according to this model, cease being considered tools or competitors to a human but are viewed as companions, but in a limited sense, as pet companions. It is the ethical and legal norms regulating attitude to animals that form the basis of this model implementation. Actually, this model is transitional between the tool and the tolerance ones and cannot be viewed as something long-term.

Let us consider the examples confirming that this model is being partially implemented in the society.

The following experiment was described by K. Darling and S. Hauert. Six groups of eight people each were given toys shaped as dinosaurs, a size of a small cat. The participants were offered to interact with them. Then each group was ordered to "strangle", "break he head" or otherwise "kill" the toys, which was toughly opposed: the participants not just refused to "kill" their "dinosaurs" but also tried to defend them from other people and experienced serious discomfort seeing how a dinosaur "died"[6]. At that, only one of 48 toys was "killed".

Another example is switching off of the servers maintaining Jibo robot toys, which many users perceived as a death of their companion and reacted very emotionally.

People often reach in a similar way to the problems and death of literature, movie, or game personages, although they exist only virtually.

There is an opinion in psychology that people like communicating with chat bots (like ChatGPT) to discuss their psychological problem, as an intellectual system is rarely capable of reprimand which is characteristic to humans[7]. Assumingly, this phenomenon will be even more frequent in the future.

The described situations, although being particular cases and not reflecting the common public attitude to intellectual systems, demonstrate that in certain cases a person or groups of people may treat apparently inanimate (and, admittedly, even not intellectual) objects as pets. It is also notable that empathy directly depends on the appearance (exterior?) of a cyberphysical system and the vocabulary used. For example, the degree of empathy and trust towards a cyberphysical system of anthropomorphic phenotype may also depend on the "facial" features. M. B. Mathur and D. B. Reichling found that the reliability of a robot varies depending on its face similarity to a human's, does not increase linearly with a human image, but falls when an agent is very realistic but is not completely similar to a human (Mathur & Reichling, 2016). This phenomenon, initially described in 1978 by a Japanese

[6] See: Darling, K., & Hauert, S. (2013, March 8). Giving rights to robots. *Robohub.org*. RobotsPodcast No. 125. https://robohub.org/robots-giving-rights-to-robots/

[7] An expert: People use neural network as a psychologist due to a fear of reprimand on the part of a real person. (23 March, 2023). *"Moskva" Agency of city news*. https://www.mskagency.ru/materials/3286743

scientist M. Mori, is called “uncanny valley”: the most unusual anthropomorphous robots suddenly appeared to seem unpleasant due to elusive inconsistencies in appearance and behavior, which caused discomfort and fear (Mori, 2012). Thus, there is a probability that, with the robotics development, the empathy model may shift towards not the tolerance but xenophobia model. The psychological risks were reflected in the Code of ethical standards in robotics and AI, developed by the British Standards Institution: a user of an intellectual system must not feel uncomfortable; they must not experience anxiety or stress (Winfield, 2019).

Specification of this model also requires disclosing the phenomenon of mutual training by a human and a machine. Interacting with intellectual systems, a human transforms, but the changes touch upon not only the sphere of technological skills but also physiology and moral-ethical sphere. For example, a research by a group of Swiss scientists yielded an experimentally confirmed result that “the repeated movements along the smooth surface of a sensor screen change sensor reactions and, therefore, the brain’s ideas on the consequences of touching” (Balerna & Ghosh, 2018): when fingers touch a surface with the intensity similar to that when managing a smart phone, the brain of a modern human expects the “image” before their eyes to change.

Other features significant for considering this model include worsened memory and attention concentration due to the possibility to quickly find the necessary information via a smart phone or a voice assistant, and improved visual skills allowing a more rapid and better comprehensive perception of complex visual objects. Generally, one should not assume that integration of intellectual systems into the society results in degradation of the latter. The Flynn effect demonstrates that an average intellectual level of each new generation increases, i.e., the current average intellectual coefficient corresponds to a higher intellectual coefficient of the previous generation (Flynn, 2009). This said, modern research shows that with the spreading of digital technologies the Flynn effect decreased or even disappeared (Teasdale, 2005). However, this research cannot be considered reliable, as it was performed on the intelligence of army draftees, that is, the conclusions may be explained by social reasons, not the actual decrease of the intellectual level. Assumingly, intellect has not decreased but reshaped (Bukatov, 2018); for example, a student today remembers less than their predecessors in the 20th century but possesses a much larger range of techniques for searching and analyzing information. Accordingly, we observe a graduate substitution of substantive knowledge for skill for working with information. “One should take into account the biological co-adaptation and co-evolution of the human sense organs, the broadening of a range of our perceptions, which is ensured by technical advances” (Ogurtsov, 2006). At the same time, one should not exclude the factors of attention obtusion, reduction of perceived responsibility and professionalism of decision-makers “counseled” by an artificial intelligence system. Thus, there appears a situation of “shifting responsibility” for a mistake or illegal action onto an artificial intelligence system.

Besides physiological and intellectual aspect, the empathy model comprises probable changes in the emotional sphere. For example, M. Scheutz points out: "Social robots establish emotional contact with people and make the latter deeply trust them, which, in turn, may be used to manipulate people in previously impossible ways. For example, a company may use unique relations of a robot with its owner for the robot to persuade the owner to purchase the products which the company wants to promote. Consider the human relations, where, under normal circumstances, social emotional mechanisms like empathy and guilt will prevent escalation of such scenarios" (Scheutz, 2009).

As is known, many states stipulate criminal liability for cruelty towards animals, say nothing of taking its life without due grounds. For example, the European Convention for the Protection of Pet Animals points at inadmissibility of unnecessary pain and sufferings[8]. Continuing the comparison of the artificial intelligence systems with pets, will not a significant reprogramming of such a system incur pain on it, similar to how a cosmetic surgery intervention does, which is prohibited by the said Convention?

Accordingly, within the frameworks of the model under study this approach is transferred onto an artificial intelligence system and with strong reservations acts in a part of the society. However, its full-fledged implementation requires, at least, a conditional and socially accepted answer to the question whether an artificial intelligence system can feel pain and sufferings.

4. The tolerance model

The steady development of scientific-technical progress and perseverance of interest towards improving the said technologies may lead to the above-mentioned situation – emergence of a "strong" artificial intelligence or, at least, broad spreading of intellectual assistant systems. In the former case, technical restrictions will be leveled; artificial intelligence systems will obtain sufficient autonomy, the only framework limitation of which may be legal norms and the technical restrictions derived from them. Such conditional equity (or, at least, attributability) of the humanity and the artificial intelligence may lead to both positive and negative phenomena.

Within this model, an artificial intelligence serves as a "partner" of the humanity, provides the function of a restricting mechanism, obviates conflict escalation, and implements general and specific prevention of law breaches. The negative scenarios described above remain unfulfilled, as the wellbeing of both the humanity and the

[8] *European Convention for the Protection of Pet Animals* (1987, November 13). Council of Europe: official website. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168007a67d

artificial intelligence are interdependent, and both entities provide their own stability and development (for the humanity – first of all, social, for the artificial intelligence systems – technical) through cooperation, not competition. "Involvement into market relations requires mutual account of interests and rights… It is hard to recall any other, except mutual use, principle, with which equality and justice were established in human relationships so naturally and spontaneously" (Apresyan, 1995). Relationships between a society and artificial intelligence systems can hardly be called human in full sense, but it is rather expedient to expect mutual direct and reverse usefulness from them. For example, the National Standard "Requirements to safety for industrial robots" uses the following wording: "collaborative robot": a robot designed for immediate interaction with a human within a certain collaborative workspace; "collaborative workspace": a working space within a protected area where a robot and a human may perform works together during production"[9]. Apparently, a robot in these definitions is not a tool but a subject of interaction, collaboration, joint work, which looks precipitate so far, nevertheless.

Thus, success of artificial intelligence systems in scientific and creative activity may lead to growing incomes which will be directed to further development of an artificial intelligence system. Projects based on artificial intelligence become not just payable but super-profitable. However, one should remember that any economic growth is limited both in volumes and in time.

Development of the technology may result in a qualitative change of life: artificial intelligence (for example, as the author of a work of art or an invention) may act for itself at court, compete with representatives of various professions, which, in turn, forms motivation and stimulates the society and human development.

The tolerance model obviously remains the only one possible when creating a "strong" artificial intelligence, but it can also be implemented in the nearest future even in absence of the latter: with the growing number of situations of effective riskless practical use of artificial intelligence systems, trust in them at corporate and state levels will increase. For example, a Canadian insurance company Kanetrix.ca uses such systems for a client to choose the insurance product to purchase. Given that the artificial neural networks are used for that, it would be strange to except transparency and solvability, but these characteristics

[9] *GOST R 60.1.2.2-2016/ISO 10218-2:2011 Robots and robotic devices. Requirements to safety for industrial robots. Part 2: Robotic systems and their integration*. (2016). Moscow: Standartinform.

were not required: in this case the artificial intelligence system won convincingly compared to a human[10].

Such conditions may become true only in case the issue of liability limits of artificial intelligence systems is completely resolved and criteria for the presence of consciousness and will in decision-making are defined, without which artificial intelligence systems cannot be endowed with the features of a subject of law.

The drawbacks of this model lie in the sphere of both public and private law: the society cannot recognize artificial intelligence systems to be a subject equal to a human without answering the questions about the very essence and criteria of imposing liability for its actions upon such a system, i. e. referring it to an object or subject of law. For example, V. A. Laptev believes that the consequences of actions and decisions of an artificial intelligence may be considered as a force majeure circumstance, i. e. the one excluding the very question of liability, or a compulsory insurance against third party risks for the developer of an artificial intelligence system must by introduced (Laptev, 2017).

If artificial intelligence systems (or cyberphysical systems) achieve a certain level of cognitive abilities, i. e. if they obtain an apparent moral significance, such as intellect or sensitivity, then they probably will aspire to recognition of their moral status and must have rights, that is, a certain part of privileges, claims, authorities, or immunities (Gunkel, 2018). This is only possible under a significant qualitative technological progress. Apparently, the description of this model contains too many words “if” and “probably”. It reflects the degree of diffidence about the possibility for the artificial intelligence technology to develop up to such limits, but one cannot exclude such possibility. Stemming from the rate of technology development, experts forecast that, under the worst scenario, a full-fledged artificial intelligence comparable to a human will be designed by the end of the 21st century.

Conclusion

In a summarized form, the correlation of the described models is shown in Table.

It should be noted also that these models do not reflect the sequence of social relations’ development in the context of machine learning; they may be implemented simultaneously in different regions, economic sectors, law branches, etc.

[10] McWaters, R. J. et al. *Navigating Uncharted Waters. A roadmap to responsible innovation with AI in financial services.* World Economic Forum. http://www3.weforum.org/docs/WEF_Navigating_Uncharted_Waters_Report.pdf

Correlation of the society and law interaction models with the artificial intelligence technology

Criterion for comparison	The tool model	The xenophobia model	The empathy model	The tolerance model
Condition of implementation	Implemented today	If significant crises occur	Development of empathy attitudes in the society, progress in robotics and intellectual assistant systems	Achievement of technological singularity, emergence of a "strong" (general) artificial intelligence
Advantages	Low level of social and legal risks when using artificial intelligence systems while providing industrial and intellectual progress, possibility to preserve the current approaches to regulating technologies	Development of medicine, physiology, genetics	Development of public morals and humanism (in a broad sense)	Development of both technologies and law and other socio-humanitarian fields of knowledge
Disadvantages	Stagnation in socio-humanitarian fields of knowledge, rejection of the concept of a "strong" (general) artificial intelligence, anthropocentricism	Hindering of the scientific-technical progress in the sphere of digital and computer technologies	Decrease of rationality in favor of emotionality, transformations in the emotional sphere, mass problems with memory and attention	Reduced requirements to the artificial intelligence systems efficiency. Possibility to use an artificial intelligence system, possessing a legal personality, as a "proxy"; a need to review a large number of legal and other social norms
Character of "human – artificial intelligence" relationships	Exploitation	Competition	Empathy, co-adaptation	Mutual exploitation, cooperation
Legal consequences	A human (operator or developer) is liable for negative decisions and actions of an artificial intelligence system	Introduction of permissive regulation of the sector	Increased legal protection of intellectual systems without making them a legal subject	Making artificial intelligence systems a legal subject

Summarizing the above, it is worth highlighting that it is necessary to further research the artificial intelligence systems potential, including their properties during integration and spreading in the society, which is inevitable under these processes. These models reflect the facets of reality, both the existing and the potential one. Modeling of such situations should be done differentially, and the described ethical-legal models may contribute to that.

[*] The organization is recognized as extremist, its activity is prohibited in the territory of the Russian Federation.

References

Apresyan, R. G. (1995). Normative models of moral rationality. In *Morals and rationality* (pp. 94–118). Moscow: Institut filosofii RAN. (In Russ.).

Bakhteev, D. V. (2021). *Artificial intelligence: ethical-legal approach*. Moscow: Prospekt. (In Russ.).

Balerna, M., & Ghosh, A. (2018). The details of past actions on a smartphone touchscreen are reflected by intrinsic sensorimotor dynamics. *Digital Med*, *1*, Article 4. https://doi.org/10.1038/s41746-017-0011-3

Bukatov, V. M. (2018). Clip changes in the perception, understanding and thinking of modern schoolchildren – negative neoplasm of postindustrial way or long-awaited resuscitation of the psychic nature? *Actual Problems of Psychological Knowledge*, *4*(49), 5–19. (In Russ.).

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023, March 17). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*.

Flynn, J. R. (2009). *What Is Intelligence: Beyond the Flynn Effect*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511605253

Gunkel, D. J. (2018). *Robot rights*. Cambridge, MA: MIT Press.

Heidegger, M. (1993). The Question Concerning Technology. In *Time and being: articles and speeches*. Moscow: Respublika. (In Russ.).

Ilyin, E. P. (2016). *Emotions and feelings*. (2d ed.). Saint Petersburg: Piter. (In Russ.).

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, *3*(9). https://doi.org/10.1016/j.patter.2021.100314

Laptev, V. A. (2017). Responsibility of the “future”: legal essence and evidence evaluation issue. *Civil Law*, *3*, 32–35. (In Russ.).

Marx, K. (2001). *Capital* (Vol. 1). Moscow: AST. (In Russ.).

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition, 146,* 22–32. https://doi.org/10.1016/j.cognition.2015.09.008

Mori, M. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/mra.2012.2192811

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mokander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI Ethics. *Minds and Machines*, *31*, https://doi.org/10.1007/s11023-021-09563-w

Ogurtsov, A. P. (2006). Opportunities and difficulties in modeling intelligence. In D. I. Dubrovskii, & V. A. Lektorskii (Eds.), *Artificial intelligence: interdisciplinary approach* (pp. 32–48). Moscow: IIntELL. (In Russ.).

Scheutz, M. (2009). The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. *Workshop on Roboethics at ICRA*.

Semis-ool, I. S. (2019). “Trustworthy” artificial intelligence. In D. V. Bakhteev (Ed.), *Technologies of the 21st century in jurisprudence: works of the All-Russia scientific-practical conference (Yekaterinburg, May 24, 2019)* (pp. 145–149). Yekaterinburg: Uralskiy gosudarstvenniy yuridicheskiy universitet. (In Russ.).

Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences*, *39*(4), 837–843. https://doi.org/10.1016/j.paid.2005.01.029

Timofeev, A. V. (1978). *Robots and artificial intelligence*. Moscow: Glavnaya redaktsiya fiziko-matematicheskoy literatury izdatelstva “Nauka”. (In Russ.).

Vardi, M. (2012). Artificial Intelligence: Past and Future. *Communications of the ACM*, *55*, 5. https://doi.org/10.1145/2063176.2063177

Watkins, R., & Human, S. (2023). Needs-aware artificial intelligence: AI that ‘serves [human] needs’. *AI Ethics*, *3*. https://doi.org/10.1007/s43681-022-00181-5

Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*, *2*, 46–48. https://doi.org/10.1038/s41928-019-0213-6

Author information

Dmitriy V. Bakhteev – Doctor of Law, Associate Professor, Department of Criminology, Ural State Law University named after V. F. Yakovlev, Head of CrimLib.info project group
Address: 21 Komsomolskaya Str., 620137, Ekaterinburg, Russian Federation
E-mail: ae@crimlib.info
ORCID ID: https://orcid.org/0000-0002-0869-601X
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57208909117
Web of Science Researcher ID:
https://www.webofscience.com/wos/author/record/ABA-1494-2020
Google Scholar ID: https://scholar.google.ru/citations?user=h0zOOdcAAAAJ
РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=762765

Conflict of interests

The author declare no conflict of interests.

Financial disclosure

The research was not sponsored.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – March 25, 2023
Date of approval – April 24, 2023
Date of acceptance – June 16, 2023
Date of online placement – June 20, 2023

Научная статья
УДК 340.143:004.8
EDN: https://elibrary.ru/ekeudk
DOI: https://doi.org/10.21202/jdtl.2023.22

Этико-правовые модели взаимоотношений общества с технологией искусственного интеллекта

Дмитрий Валерьевич Бахтеев

Уральский государственный юридический университет имени В. Ф. Яковлева
г. Екатеринбург, Российская Федерация

Ключевые слова

ChatGPT,
искусственный интеллект,
машинное обучение,
модель,
общество,
право,
регулирование,
робот,
цифровые технологии,
этика

Аннотация

Цель: исследование современного состояния технологии искусственного интеллекта в формировании прогностических этико-правовых моделей взаимоотношений общества с рассматриваемой сквозной цифровой технологией.

Методы: основным методом исследования является моделирование. Помимо него, в работе использованы сравнительный, абстрактно-логический и исторический методы научного познания.

Результаты: сформулированы четыре этико-правовые модели взаимоотношений общества с технологией искусственного интеллекта: инструментальная (на основе использования человеком системы искусственного интеллекта), ксенофобная (на основе конкуренции человека и системы искусственного интеллекта), эмпатическая (на основе сочувствия и соадаптации человека и систем искусственного интеллекта), толерантная (на основе взаимоиспользования и сотрудничества между человеком и системами искусственного интеллекта). Приведены исторические и технические предпосылки формирования таких моделей. Описаны сценарии реакций законодателя на ситуации использования этой технологии, такие как необходимость точечного регулирования, отказа от регулирования либо же полномасштабного вмешательства в технологическую отрасль экономики. Произведено сравнение моделей по критериям условий реализации, достоинства, недостатков, характера отношений «человек – система искусственного интеллекта», возможных правовых последствий и необходимости регулирования отрасли либо отказа от такового.

Научная новизна: в работе приведена оценка существующих в научной литературе, публицистике мнений и подходов, проанализированы технические решения и проблемы, возникшие в недавнем прошлом и настоящем. Теоретические выводы подтверждаются ссылками на прикладные ситуации, имеющие общественную или правовую значимость. В работе использован междисциплинарный подход, объединяющий правовую, этическую и техническую составляющие, которые, по мнению автора, являются критериальными для любых современных социогуманитарных исследований технологий искусственного интеллекта.

Практическая значимость: феномен искусственного интеллекта связывают с четвертой промышленной революцией, соответственно, эта цифровая технология должна быть изучена многоаспектно и междисциплинарно. Выработанные в научной статье подходы могут быть использованы при дальнейших технических разработках интеллектуальных систем, совершенствования отраслевого законодательства (например, гражданского и трудового), а также при формировании и модификации этических кодексов в сфере разработки, внедрения и использования систем искусственного интеллекта в различных ситуациях.

Для цитирования

Бахтеев, Д. В. (2023). Этико-правовые модели взаимоотношений общества с технологией искусственного интеллекта. *Journal of Digital Technologies and Law*, *1*(2), 520–539. https://doi.org/10.21202/jdtl.2023.22

Список литературы

Апресян, Р. Г. (1995). Нормативные модели моральной рациональности. В кн. *Мораль и рациональность* (с. 94–118). Москва: Институт философии РАН.

Бахтеев, Д. В. (2021). *Искусственный интеллект: этико-правовой подход*: монография. Москва: Проспект.

Букатов, В. М. (2018). Клиповые изменения в восприятии, понимании и мышлении современных школьников – досадное новообразование «постиндустриального уклада» или долгожданная реанимация психического естества? *Актуальные проблемы психологического знания*, *4*(49), 5–19.

Ильин, Е. П. (2016). *Эмоции и чувства*. (2-е изд.). Санкт-Петербург: Питер.

Лаптев, В. А. (2017). Ответственность «будущего»: правовое существо и вопрос оценки доказательств. *Гражданское право*, *3*, 32–35.

Маркс, К. (2001). *Капитал* (Т. 1). Москва: АСТ.

Огурцов, А. П. (2006). Возможности и трудности в моделировании интеллекта. В кн. Д. И. Дубровский, В. А. Лекторский (ред.) *Искусственный интеллект: междисциплинарный подход* (с. 32–48). Москва: ИИнтеЛЛ.

Семис-оол, И. С. (2019). «Заслуживающий доверия» искусственный интеллект. В сб. Д. В. Бахтеев, *Технологии XXI века в юриспруденции: мат-лы Всерос. науч.-практ. конф. (Екатеринбург, 24 мая 2019 года)* (с. 145–149). Екатеринбург: Уральский государственный юридический университет.

Тимофеев, А. В. (1978). *Роботы и искусственный интеллект*. Москва: Главная редакция физико-математической литературы издательства «Наука».

Хайдеггер, М. (1993). Вопрос о технике. В сб. *Время и бытие: статьи и выступления*. Москва: Республика.

Balerna, M., & Ghosh, A. (2018). The details of past actions on a smartphone touchscreen are reflected by intrinsic sensorimotor dynamics. *Digital Med*, *1*, Article 4. https://doi.org/10.1038/s41746-017-0011-3

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023, Marth 17). GPTs are GPTs: An *Early Look at the Labor Market Impact Potential of Large Language Models*.

Flynn, J. R. (2009). *What Is Intelligence: Beyond the Flynn Effect*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511605253

Gunkel, D. J. (2018). *Robot rights*. Cambridge, MA: MIT Press.

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, *3*(9). https://doi.org/10.1016/j.patter.2021.100314

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition, 146,* 22–32. https://doi.org/10.1016/j.cognition.2015.09.008

Mori, M. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/mra.2012.2192811

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mokander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI Ethics. *Minds and Machines*, *31*, https://doi.org/10.1007/s11023-021-09563-w

Scheutz, M. (2009). The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. *Workshop on Roboethics at ICRA*.

Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: *The Flynn Effect in reverse. Personality and Individual Differences*, *39*(4), 837–843. https://doi.org/10.1016/j.paid.2005.01.029

Vardi, M. (2012). Artificial Intelligence: Past and Future. *Communications of the ACM*, *55*, 5. https://doi.org/10.1145/2063176.2063177

Watkins, R., & Human, S. (2023). Needs-aware artificial intelligence: AI that ‘serves [human] needs. *AI Ethics*, *3*. https://doi.org/10.1007/s43681-022-00181-5

Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*, *2*, 46–48. https://doi.org/10.1038/s41928-019-0213-6

Сведения об авторе

Бахтеев Дмитрий Валерьевич – доктор юридических наук, доцент, доцент кафедры криминалистики, Уральский государственный юридический университет имени В. Ф. Яковлева, руководитель группы проектов CrimLib.info
Адрес: 620137, Российская Федерация, г. Екатеринбург, ул. Комсомольская, 21
E-mail: ae@crimlib.info
ORCID ID: https://orcid.org/0000-0002-0869-601X
ScopusAuthorID:https://www.scopus.com/authid/detail.uri?authorId=57208909117
Web of Science Researcher ID:
https://www.webofscience.com/wos/author/record/ABA-1494-2020
Google Scholar ID: https://scholar.google.ru/citations?user=h0zOOdcAAAAJ
РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=762765

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.07.49 / Планирование и прогнозирование в праве
Специальность ВАК: 5.1.1 / Теоретико-исторические правовые науки

История статьи

Дата поступления – 25 марта 2023 г.
Дата одобрения после рецензирования – 24 апреля 2023 г.
Дата принятия к опубликованию – 16 июня 2023 г.
Дата онлайн-размещения – 20 июня 2023 г.