Research article
DOI: https://doi.org/10.21202/jdtl.2023.41

Ethics of Artificial Intelligence and Robotics: Key Issues and Modern Ways to Solve Them

Nidhika Yadav

Independent Researcher
New Delhi, India

Keywords

artificial intelligence,
ChatGPT,
cyber security,
data privacy,
digital technologies,
ethics,
law,
robot,
robotics,
safety

Abstract

Objective: modern achievements in the development and dissemination of digital technologies have attracted the attention of scholars and practitioners to the discussion of key ethical issues related to artificial intelligence and robotics. Hence, this study presents the most relevant of these issues, posing new challenges for legal scholars and practitioners to develop the regulation of artificial intelligence and robotics in terms of technology moralization.

Methods: the research used practice- and risk-oriented approaches, complemented by multidisciplinary analysis of documents (European principles and codes of ethics) and studies, including those devoted to various problems of artificial intelligence and robotics.

Results: the article identifies key ethical issues in the field of artificial intelligence and robotics. It is established that the key ethical issues involved can be solved if they are legally formalized and implemented at the international level. The algorithm proposed by the author, based on the analysis of the digital technologies application, will allow improving the moral actions of technologies in the process of their decision making.

Scientific novelty: the article presents the latest ethical problems that concern scientists and practitioners in the field of artificial intelligence and robotics, and the methods of their solution by ethical and legal means aimed at moralizing technology and increasing its responsibility.

Practical significance: all solutions presented in the article have practical significance and are ready for wide implementation at the international level. Their formalization in normative form and subsequent compliance will reduce the harm that artificial intelligence may cause in applied fields, including robotics using artificial intelligence. Regulatory, including legislative, decisions must therefore be taken as soon as possible to ensure that artificial intelligence and robotics become reliable tools for these systems to be used at work, at home, and in other areas such as shopping centers, stores, schools, universities, etc.

For citation

Yadav, N. (2023). Ethics of artificial intelligence and robotics: key issues and modern ways to solve them. *Journal of Digital Technologies and Law, 1*(4), 955–972. https://doi.org/10.21202/jdtl.2023.41

Content

Introduction

Artificial Intelligence (AI) (Rich et al., 2009) has gained a lot of momentum all of its own recently. The progress in AI was immense in the past few years. The progress is record-breaking in each of its application areas (Saveliev et al., 2021). Further, many segments of recent developments in AI were, firstly, never thought of and, secondly, many things were never imagined as well (Xiao-Fan et al., 2023; Kumar et al., 2023). ChatGPT is one such

example wherein even the makers of this AI product were stunned in surprise by the peaks of development they made. ChatGPT is just one such example; the progress is seen in many domains, not just in generative texts, but in many parts of AI research. Some of these are discussed in this paper.

Artificial Intelligence (AI) is a term coined for a branch of science that deals with mimicking the working of the human brain to accomplish difficult work that humans perform. The aim was to ease human work and perform like humans or to provide auxiliary help in tasks presented to humans. The subject AI (Rich et al., 2009; Stahl, 2021; Mueller, 2012) is constituted of an amalgamation of many fields such as computer science, mathematics, cognitive sciences, and psychology to mention a few. The inception of AI was in the early 50's and it gained the interest of researchers since then (Turing, 2009; Schank, 1991). The subject was developed to build algorithms and/or machines that can do work or perform like humans and in some cases where humans could not do work, AI-based intelligent machines could do work for humans, for example, mining for metals in Earth's crust (Lesandrini et al., 2023; Memarian & Doleck, 2023).

The aim of AI was never to harm humans. The aim of AI was never to replace humans and the aim of AI was never to take over the world of humans. The aim of AI was to augment humans, to help humans, to help scientific discoveries, to help humans in case of natural calamities, to help humans in pandemic-like situations, and so on. Other applications where AI seems to insightfully help humans are in day-to-day medical care centers, mechanical engineering companies, driving in semi and fully-automated cars, in bot-based deliveries in dire flood-like situations, in intelligent robotics in various segments of application such as textile engineering to even in running advancements in aircraft, traffic management, and space programs, to mention a few.

But now the development of AI has been at its greatest peak and it can do the work humans are doing in almost all fields. So many jobs are being replaced by AI these days, but that should not fear people. New jobs are created at the same rate as well. However, the statistics for new jobs created versus jobs lost is yet to come out to common people. This is how development has been progressing since the industrial revolution. Getting into a new domain of work because AI has taken away many workflows is not easy. We all remember when the textile engineering department was computerized then so many jobs were taken away. The affected engineers had to take new jobs. Once an area becomes obsolete, a new area of development starts. When mobile phones came, landline phones became obsolete. Today the same mobile phones provide employment to many people all around the globe. More on solutions to job loss is discussed in the coming sections.

The aim of this paper is to present the ethical aspects of AI, its ethical uses, and proposed solutions to problems that arise with the growth of AI and especially AI with Robotics. Section 2 describes the why behind studying AI and Robotics ethical issues. Section 3 discusses Robotics with AI. Section 4 presents the solutions for key ethical problems in AI and Robotics, and Section 5 gives the conclusion of the paper.

1. Preliminary notes on Ethics in AI and Robotics

Many people think of AI as a threat to its useful applications. However, as with other scientific disciplines, AI is a well-grown subject on its own. The need for laying out the rules for Ethics in AI and Robotics is discussed in this paper. It must be known that just like other subjects of science, AI too needs to have certain constraints in its development, discovery, and deployment. This is nothing new, the well-developed subjects such as chemistry, physics, biology, and so on, all have guidelines of use. One example can be some lab-grown fungi, being leaked out and creating havoc in nature; this can be the result of a lack of caretaking and missed protocols of the lab in which fungi were stored for some practical works. Hence, guardrails are laid out in all fields of science and hence should be employed in AI and Robotics as well. Another example is a new car using nuclear energy to run the engines. If not tested well in all extreme temperatures and conditions it can create many mishaps.

The difference between other scientific disciplines and AI is that AI grew from some scientific disciplines primarily consisting of computer science and mathematics, in which there are no protocols for the delivery of products. Hence, it all started with the delivery of AI products without scientific and social guidelines for deployment. Another reason is AI is still a newbie when compared to other subjects; this must be accepted while framing the constraints of products produced with AI.

The good news is there are no major regular faults with AI and Robotics yet. Some problems that occurred a little while back were in autonomous cars and recommender system of a tech giant. The future may develop more AI and Robotics products but those should be ethical, there are the responsibilities of AI. AI has helped humans reach the level of development that exists today. So many devices and products we use, from washing machine to dishwasher to online vacation ticket booking, are rich in one form of AI or another. They are to be used in the future as well. The use of AI cannot be stopped; hence, AI and Robotics need to follow ethics to be used in larger potential.

AI is now an independent subject of its own. What we require now is that AI has its own instruction manuals, just like all physics experiments have pre-defined instructions, likewise all AI and Robotics experiments must have instruction sets. This must be done for each AI application. It may take some time to draw the instruction set, and the hope is, it shall be useful for we are defining the rules scientifically. This compliance of rules shall be tested well before the release of each AI product, be it in Robotics or just in NLP (Rich et al., 2009) tasks that do the recommendation or play a song based on your liking. Hence, proper guidelines for all kinds of AI labs need to be set up.

2. Robotics

The definition of Robotics is extensive and characterization is still bleak. Typically, robots are entities and machines that define mechanical movements and can be with or without any kind of Artificial Intelligence. Robotics is the study of robots (Agarwal & Stoff, 2023). However, this paper shall be focused on smart Robotics (Joachim et al., 2021) equipped with capabilities to run and work on AI-based concepts. The use of robotics can include a wide range from mere mechanical jobs, to robots for medical surgery, to even humanoid robots ready for space discoveries and space travel on their own, but still to be guided in one way or another by human targets.

There are other concerns about the definition of Robots, as in a machine that is remotely guided to perform some tasks. There are discussions on this as well (De Felice et al., 2022), whether such a machine is called an intelligent robot. The answer varies but we consider the answer to be yes, for it is not using a human brain, but signals sent from a distance through the internet or other sources. An example of such devices is IoT-based bots. Hence, these products increase the definition of Robotics, indeed intelligent robotics. They mimic the working of the brain, even if it is through a communication channel. This extends the laws of Ethics in Robotics to be applied to all such devices as well.

Why is Robotics and its analysis important to be considered in AI Ethics discussion? This is because Robotics is the future in which AI is heading to. Robots should always be monitored and altered by humans under whose supervision the assigned task is performed. Here comes ethics for Robotics that use AI. The Ethics for Robotics constitute not only the management of work undertaken but also how the AI-powered Robotics affect its environment actors around the task area.

2.1. Responsible robotics

Responsible robotics (Memarian & Doleck, 2023) is the robotics which aims to help humans in performing tasks that can vary from day-to-day household chores to even complicated surgeries, and even to journeys in space for scientific discoveries. But being a major part of AI, robotics should not harm humans and their environments. This should be the aim with which responsible robotics should be built, this needs to be the foundation stone of building intelligent AI-based Robotics. It must be noted that responsible robotics cannot hither to the other side of wrong impacts given the inputs are modeled into outputs. For this, the ethics protocols need to be made.

Why we need robots is a different question. They are required to do things to help humans or things which humans cannot do. Robots work on programming languages to even bytes and bits-based data. Intelligent robotics needs to understand problems before it can solve them.

Whether we use Robotics (Brooks, 1991) with AI, many kinds of robotics are now in use in industries, MNC companies, and manufacturing. They can be used in rescue operations from a mine in case of some need, to colonize another planet or space object before humans reach a new place in outer space that is not safe for humans. For the safety of humans, robotics needs to be sent to other outer space locations before humans. Yes, people do say "There is no Planet B", that is true now, but mankind progresses with hope, and with hope, humans have reached this advancement in all sectors of development. There is no harm in trying to make AI and Robots that can help the health and well-being of Planet A, which is the Earth while fostering other space bodies for life and needs.

Other areas where AI-based Robots can help are fire-resistant robots to help in extreme forest fires which the firefighters cannot extinguish fire. Robots can take out metals and reusables from garbage and to automatically distinguish single-use plastic from other kinds of garbage. How much can humans do alone? There are jobs not meant for humans, let us accept that. There come AI-based machines and Robots to use. With their use comes the responsibility of humans to make these machines safe for use. Once again, AI is still in its youth and in the nascent stage of development; hence, there should be nothing to fear, as long as ethics are applied to all new or old use cases for AI applications.

3. Solutions for Ethics in AI and Robotics

Once the software is set and a machine always follows the rules inscribed in its software, one can be confident that responsible robotics cannot hither towards the wrong side. There are two characterizations of AI in terms of its use and the outputs it produces. These are given as follows (Kumar et al., 2023):

1. The good AI

The good AI is the beneficial AI, which is developed to accomplish some tasks. It can be tested using black box and white box testing as well. These kinds of AI applications require fixed inputs and produce fixed outputs that are well predictable. These may use typical pattern recognition tools (Duda & Hart, 2006) such as decision tree-based solutions to predict leukemia using a support vector machine. These algorithms cannot do harm to anyone and they typically have predefined output guidelines. Other uses are in defense, aircraft flying, rocket launch, to even healthcare prediction, to operations. Hence, these algorithms are termed as good AI.

2. The bad AI

The AI algorithm can harm by use of bias, privacy sacrifice, bad language, self-esteem or even hurting in any form, and more comes under bad AI. These things materialize in AI as beautifying apps that affect teenagers' self-esteem, video editing by adults, fake profiles, fake images, AI-based agents playing games, fake and manipulated news, hacking with AI, and copywriting issues, to mention some. These can lead to self-harm apart from losses in personal and professional lives.

All major problems that we need solutions for are products of bad AI. Good AI is primarily ethical and does not provide any harmful problems. These typically use machine learning (Bishop, 2006, Palladino, 2023) algorithms to find solutions and work on the problems. Still, to be on the safer side, let the guidelines be made for all kinds of AI, good AI, bad AI or simple AI, or AGI, the general intelligence-based AI. For any of these AI and Robotics with AI need to follow both black box testing and white box testing. The testing can to some extent prevent wrong outputs. Black box testing can be performed by any good computer-literate person, while white box testing is more comprehensive testing that involves getting into the codebase to get the why of an encountered error. This solves the problem from the root. Hence, white box testing should be encouraged in ethical AI and its use in Robotics and IoT.

3.1. Privacy

AI tools such as “Help me write”, “smart compose” and “smart reply” can raise privacy concerns (Zhang et al., 2021; Stahl & Wright, 2018), due to the impression that your emails are being read by AI and can be used in the wrong ways. Privacy does not end with emails. Privacy can extend to data privacy, where data entered by a person is at risk of leaking, and personal privacy where the profile of the AI hardware/software user is at risk. The data can be data about the user’s personal choices or about their company, family, or properties. Moreover, any kind of attack on privacy is wrong, and AI-based tools and apps should be saved from privacy attacks or virus threats. AI apps should be made in such a way as to avoid attacks by viruses. Further, AI apps themselves can generate data from users’ web browsing or other activities; all these should be taken care of before using the AI tool.

The solution to this problem is self-awareness and installing relevant patches on the system that use AI, installing anti-malware and antivirus on systems. For other apps that use AI-based auto-fillers, a toolkit maker must be contacted to appliy the code of conduct for AI tools. This is the only way to gain security over our data.

3.2. Ownership

With the advancements in AI comes the problem of ownership. AI initially was used by artists to help artists, songwriters, scriptwriters, and movie makers but now it has taken over all these domains. Who taught AI all this? The answer is humans. AI-based software was trained on hundreds and thousands of human-created art, photographs, chats and texts. But when AI software creates something new, say a new image file, it may be created using thousands of man-made items This means the credit for new images made by AI to some extent lies in the hands of humans who made the base image, and who made AI learn painting or singing, as the case may be.

The solution to this can be understood as follows. There are traditional artists and AI and digital arts-based artists. Let us understand all this with the help of restricting art to painters; similar analogy would apply to other artists, be they scriptwriters or musicians, to mention a few.

3.2.1. Traditional artists

Traditional artists stick to age-old ways along with newer approaches to make art. Many artists create artwork with traditional colors and styles and may choose acrylic paints while some use tough oil paints. Many artists work with glass or clay. These are all examples of traditional ways of creating art. For them, this art is not only their profession but also their livelihood, passion, and source of love and worship. Most artists work as a hobby and for generating side income; these artists may not be as badly hit as those who rely completely on art for a living.

As with newer developments in AI, new needs, and developments in technology, old ways of making art seem to become obsolete. The newer kinds of art are:

1. Faster to make,
2. Typically cheaper, and
3. More lucrative (Scimeca & Bonfiglio, 2023).

However, most people still prefer the original arts. Traditional kinds take more effort than the competitive digital arts. For many, getting in a new technique is not an easy task to do, and hence they stick to older and typical ways to create art. Initially, digital arts are not easy to grasp either but once an artist is a master of digital art it will not take much hurdles to expertise in it.

3.2.2. Tech and AI artists

The Tech and AI artists are the artists who use some form of technology or AI to make artwork of any kind, be it music, script writing, or painting. These can use tools such as digital art software, 2D or 3D printed art, or may be powered by AI and Robotic hands. These artists take training to learn the toolkits; once learned the artists are free to explore the infinite world of arts that can be made with these toolkits. The AI-powered toolkits are even more interesting to make, though an artist may need to write a few lines of code to generate such masterpieces in music and arts. Here, the artist may be working in collaboration or on their own, and at times an artist works on artwork from scratch and many times they edit the work of other artists to enhance them.

Some of these arts can be made on top of art which are not owned by the group of artists who are making the new art. Then legally the original art owner possesses the rights to his/her art. They own the copywriting of the art and still, there are no stringent laws to take care of this. Here both suffer: the digital artist who has put his or her AI-based art knowledge on someone else's artwork, and the original traditional artist(s) for not getting the

acknowledgment as deserved. Till the time a law comes to this front, an acknowledgment must be paid to the base artist.

The pressure on artists is enormous, both on traditional artists and digital and AI-based artists. The concerns include license of use, earning, owning the rights, publicity, and certificate of authenticity.

3.2.3. Solutions

The AI and Robotics based arts are trained on thousands and more of human-made artworks, photographs, scripts, etc. These training data is used to make newer artwork. But it is not always clear how to find the original makers of artwork from AI-generated art. Sometimes the contribution is just to make a colored ball in the new AI art and many times it is the whole face that is used and maybe processed with a makeup filter. The solution lies in how much of art was taken from original training data, and how much was learned as a teacher-learner relationship. The solution can be legally formulated as follows:

1. The AI artwork has more than 80% of similarity with the original art in training data. Then the original artist and the new AI artist can get into some negotiations on sharing the profit or barring the sale of the new artwork, as the base artist wants. The original artist can buy the work the digital artist made, based on hours spent by the digital artist on editing.

2. In case of less than 80% of similarity, the original artist can claim a share of the profit on selling the artwork and cannot stop the digital artist from selling and purchasing the artwork.

3. In case the original artist wants to work together with a digital-AI artist, they can share the profits and the collaboration can lead to new heights in creating arts never seen before.

Such solutions may make the marketplace for arts ownership a harmonious place to work in.

3.3. Towards Responsible Robotics

The problems with responsible robotics are (Stahl et al., 2023):

1. Competing robots. As Robotics with AI grows, there may appear problems such as the encounter of robots made by two counties. The encounter needs to be peaceful and Robots should not challenge each other but work in harmony without creating global tensions with conflict of interests.

2. Fully Autonomous Robots

These Robots are the decision makers and are fully self-aware as part of Robotics units that materialize the use of AGI (Artificial General Intelligence). AGI-based Robots can

be self-aware and have their own thought processes. These are robots in making and to the best of our knowledge no true AGI-based Robot is out of use or work. Hence, special constraints need to be on AGI-based robots.

3. Malfunction AI-based Robots

These kinds of problems can be a real threat to the proper functioning of the process for which the robot was made. The only way this can be handled is by taking the robot out of work and installing some patches necessary to correct the working. This needs to be followed by proper testing and then re-deployment of the same or new robot at work.

The solution is to make universal machines which are non-competing in the delivery of outcomes and work under the Ethics of the welfare of humans. This makes it mandatory for the talks at the global level for AI and Robotics neutrality, compliance, and laws of use to attain mutually beneficial outputs. Hence, countries around the world should sign in legal documents wherein mutual cooperation is guaranteed both in AI and Robotics. The deal should include robotic entities made by countries, in defense, government, public and private sectors to be declared for use with a global approval.

3.3.1. AI-based Autonomous Cars

An autonomous, or AI-based car needs to not only drive from point A to point B, but must follow traffic rules, as well as drive on roads, get parked, maintain speed limits on different roads, not jump on humans and animals on the way. Only after these things are completed, an autonomous car can be deployed on roads. This needs to be a law. In the past few years an autonomous car bumped on the road into another entity and called it a mistake. As per what AI Ethics should make it a law, such cars should not be employed on roads to run, till all AI-based tests for Ethics are mandated for it.

One must note, no one law can be sufficient for Robotics. Let us understand ethics for autonomous cars with general Robotics ethics implementation. Each field of application of Robotics is different, hence first the definitions need to be made and then applications need to be filled in that states what role AI based Robots shall do, and then the legal team must fill in what all the entities and actors are which the AI Robot can touch in this process, then safety of operation must be made for each department. The ethics team must sit with the Robotics application team and set out guidelines for the safe use of Robotics. Once the company making robots is ready to deploy the robots, the proof of concept should be done by the AI Ethics Standards team. This needs to be through and robots should not be allowed to function till this robotic Proof of Concept is completed (Leeson & Coyne, 2005).

3.3.2. AI, Robotics, and Employment

With AI comes the automation of current fields, which may lead to some job losses. Training and retraining are some solutions. But this is not guaranteed in all areas, and some people cannot retrain, given their age or disability. In such cases, an employer should keep in the account of whom to fire and whom to keep. These are ethical concerns and this is a solution to the side effects of AI doing human work (Duan et al., 2019). There are huge potentials humans can reach, there are so many new fields to explore with both government and private funding. But the transition from one field to a new field should be smooth and companies should help the talent of their company find new jobs of equitable scope, honor, designation, and remuneration. Apart from that, a guarantee of, say, five years of work should be provided. Once again, sensitive people should be reserved in jobs and should not be laid off as in this year's mass layoffs around the world.

There are huge lists of applications that AI and Robotics can perform. All this does not mean job losses and layoffs. Robots installed in cities and countries shall not take away jobs without giving new jobs. Just like making a space for a new pet in the house or a new chandelier does not take away your sleeping time, the same is with AI and Robotics gadgets. This is the only way ahead, to accept the change AI and Robotics have brought and to wait for new arenas to be built for new jobs. At the same time, all should be done carefully while following the constraints of AI which follows Ethics (Hellwig et al., 2019).

3.3.3. AI-based Language Models

There are huge applications of AI in Large Language Models (LLM), and the past year saw the growth of AI LLM models like never before (Ferreira & Lipoff, 2023, Glaser et al., 2019). One after the other models came to the limelight breaking each other's records in performance and problem-solving and some in both. The problems that come with LLMs are the prediction of outputs, bias, privacy issues, wrong content, and bad language, to mention a few (Díaz-Rodríguez et al., 2023). The way forward for LLMs should require white box-style evaluations of AI software, as was discussed at the beginning of this section. White box testing should help predict the next output with a given input combination. This should minimize problems like those that happened in a tech giant two years ago when its wrong recommendation system affected many groups of people.

Conclusions

Once the software is set and well tested, a machine always follows the rules inscribed in its software, and one can be confident that responsible Robotics with AI cannot hither towards the wrong side. In this paper, we discussed many ethical issues existing in the current

scenario. Further, novel solutions were provided for these ethical issues. Moreover, the combination of Robotics with AI was discussed for its potential use and for its ethics for upcoming combinations of Robotics with AI. Many more areas of ethics need devised solutions, which is part of the future work of the author.

References

Agarwal, A., & Stoff, B. (2023). Ethics of using generative pretrained transformer and artificial intelligence systems for patient prior authorizations. *Journal of the American Academy of Dermatology*, *20*, 23–34. https://doi.org/10.1016/j.jaad.2023.04.030

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer. https://clck.ru/37AU6Z

Brooks, R. A. (1991). New approaches to robotics. *Science*, *253*(5025), 1227–1232.

Critchlow, A. J. (1985). *Introduction to robotics*. Macmillan Pub Co.

De Felice, F., Petrillo, A., De Luca, C., & Baffo, I. (2022). Artificial Intelligence or Augmented Intelligence? Impact on our lives, rights and ethics. *Procedia Computer Science*, *200*, 1846–1856. https://doi.org/10.1016/j.procs.2022.01.385

Hellwig, J., Huggett, S., & Siebert, M. (2019). *Artificial Intelligence: How knowledge is created, transferred, and used*. Elsevier. https://doi.org/10.17632/7ydfs62gd6.2

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges, and research agenda. *International Journal of Information Management*, *48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Duda, R. O., & Hart, P. E. (2006). *Pattern classification*. John Wiley & Sons.

Ferreira, A. L., & Lipoff, J. B. (2023). The complex ethics of applying ChatGPT and language model artificial intelligence in dermatology. *Journal of the American Academy of Dermatology*, *89*(4), e157-e158. https://doi.org/10.1016/j.jaad.2023.05.054

Glaser, J. I., Benjamin, A. S., Farhoodi, R., & Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. *Progress in Neurobiology*, *175*, 126–137. https://doi.org/10.1016/j.pneurobio.2019.01.008

Joachim von Braun, Margaret S. Archer, Gregory M. Reichberg, & Marcelo Sánchez Sorond. (2021). AI, Robotics, and Humanity: Opportunities, Risks, and Implications for Ethics and Policy. In J. von Braun et al. (Eds.), *Robotics, AI, and Humanity*. https://doi.org/10.1007/978-3-030-54173-6_1

Kumar, P., Chauhan, S., & Kumar, L. A. (2023). Artificial Intelligence in Healthcare: Review, Ethics, Trust Challenges & Future Research Directions. *Engineering Applications of Artificial Intelligence*, *120*, 105894. https://doi.org/10.1016/j.engappai.2023.105894

Leeson, P. T., & Coyne, C. J. (2005). The economics of computer hacking. *JL Econ. & Pol'y*, *1*, 511.

Lesandrini, J., Idris, M. Y., & Reis, D. S. (2023). The Ethics of Artificial Intelligence and Machine Learning. *Journal of Radiology Nursing*, *42*(3), 265–266. https://doi.org/10.1016/j.jradnu.2023.05.001

Memarian, D., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, *5*, 100152, https://doi.org/10.1016/j.caeai.2023.100152

Mueller, V. C. (2012). Introduction: philosophy and theory of artificial intelligence. *Minds and Machines*, *22*(2), 67–69. https://doi.org/10.1007/s11023-012-9278-y

Palladino, N. (2023). A 'biased' emerging governance regime for artificial intelligence? How AI ethics get skewed moving from principles to practices. *Telecommunications Policy*, *47*(5), 102479. https://doi.org/10.1016/j.telpol.2022.102479

Rich, E., Knight, K., & Nair, S. (2009). *Artificial Intelligence*. Tata McGraw Hill.

Saveliev, A. M., Zhurenkov, D. A., Poikin, A. E., & Berkutova, T. A. (2021). Ethics of Artificial Intelligence and Post-non-classical Scientific Rationality. *IFAC-PapersOnLine*, *54*(13), 397–401. https://doi.org/10.1016/j.ifacol.2021.10.480

Schank, R. C. (1991). Where's the AI? *AI magazine*, *12*(4), 38. https://doi.org/10.1609/aimag.v12i4.917

Scimeca, M., & Bonfiglio, R. (2023). Comment on redefining authorship in the era of artificial intelligence: balancing ethics, transparency, and progress. *ESMO Open*, *8*(5), 101635. https://doi.org/10.1016/j.esmoop.2023.101635

Stahl, B. C., & Wright, D. (2018). Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Security & Privacy*, *16*(3), 26–33. https://doi.org/10.1109/msp.2018.2701164

Stahl, B. C., Brooks, L., Hatzakis, T., Santiago, N., & Wright, D. (2023). Exploring ethics and human rights in artificial intelligence – A Delphi study. *Technological Forecasting and Social Change*, *191*, 122502. https://doi.org/10.1016/j.techfore.2023.122502

Stahl, B. C. (2021). Perspectives on Artificial Intelligence. In *Artificial Intelligence for a Better Future. SpringerBriefs in Research and Innovation Governance* (pp. 7–17). Springer, Cham. https://doi.org/10.1007/978-3-030-69978-9_2

Turing, A. M. (2009). *Computing machinery and intelligence* (pp. 23–65). Springer Netherlands.

Xiao-Fan, L., Zhaoyang, W., Wei, Zh., Guoyu, L., & Gwo-Jen, H. (2023). Technological support to foster students' artificial intelligence ethics: An augmented reality-based contextualized dilemma discussion approach. *Computers & Education*, *201*, 104813. https://doi.org/10.1016/j.compedu.2023.104813

Zhang, Y., Wu, M., Yijun Tian, G., Zhang, G., & Lu, J. (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems*, *222*, 106994. https://doi.org/10.1016/j.knosys.2021.106994

Author information

Yadav Nidhika – Ph.D., Independent Researcher
Address: Chattarpur Enclave Phase II Delhi-110074, India
E-mail: yadavnidhika68@gmail.com
ORCID ID: https://orcid.org/0000-0002-0165-6453
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57195398958
WoS ID: https://www.webofscience.com/wos/author/record/2269009
Google Scholar ID: https://scholar.google.com/citations?user=fndffSwAAAAJ

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

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

Article history

Date of receipt – July 23, 2023
Date of approval – October 25, 2023
Date of acceptance – November 30, 2023
Date of online placement – December 15, 2023

Научная статья
УДК 34:004:17:296.41:004.8
EDN: https://elibrary.ru/mdiefv
DOI: https://doi.org/10.21202/jdtl.2023.41

Этика искусственного интеллекта и робототехники: ключевые проблемы и современные способы их решения

Нидика Ядав

Независимый исследователь
г. Нью-Дели, Индия

Ключевые слова

ChatGPT,
безопасность,
искусственный интеллект,
кибербезопасность,
конфиденциальность данных,
право,
робот,
робототехника,
цифровые технологии,
этика

Аннотация

Цель: современные достижения в области развития и распространения цифровых технологий привлекли внимание ученых и практиков к обсуждению ключевых этических проблем, связанных с искусственным интеллектом и робототехникой, в связи с чем в настоящем исследовании представлены результаты этой дискуссии и обозначены наиболее актуальные задачи, решение которых определяет пути совершенствования регулирования искусственного интеллекта и робототехники в части морализации технологий.

Методы: в процессе исследования использовались практико- и риск-ориентированный подходы, дополняемые мультидисциплинарным анализом документов (европейских принципов и кодексов этики) и исследований, в том числе посвященных различным проблемам искусственного интеллекта и робототехники.

Результаты: в статье обозначены ключевые этические проблемы в области искусственного интеллекта и робототехники; установлено, что затрагиваемые ключевые этические проблемы могут быть решены при условии, если они получат юридическое оформление и будут реализованы на международном уровне; предложенный автором алгоритм, основанный на анализе практики применения цифровых технологий, позволит усовершенствовать нравственные действия технологий при принятии ими решений.

Научная новизна: в данной статье представлены новейшие этические проблемы, которые волнуют ученых и практиков в области искусственного интеллекта и робототехники, и методы их решения этико-правовыми средствами, направленными на морализацию технологий и повышение ее ответственности.

Практическая значимость: все решения, представленные в статье, имеют практическое значение и готовы к широкому внедрению на международном уровне. Их оформление в нормативном виде и последующее соблюдение приведут к уменьшению вреда, который может нанести искусственный интеллект в прикладных областях, в том числе в робототехнике с использованием искусственного интеллекта. В связи с этим нормативные, в том числе законодательные, решения должны быть приняты как можно скорее, чтобы искусственный интеллект и робототехника приобрели статус надежных инструментов при использовании этих систем на работе, дома и в других сферах, таких как торговые центры, магазины, школы, университеты и т. д.

Для цитирования

Ядав, Н. (2023). Этика искусственного интеллекта и робототехники: ключевые проблемы и современные способы их решения. *Journal of Digital Technologies and Law*, *1*(4), 955–972. https://doi.org/10.21202/jdtl.2023.41

Список литературы

Agarwal, A., & Stoff, B. (2023). Ethics of using generative pretrained transformer and artificial intelligence systems for patient prior authorizations. *Journal of the American Academy of Dermatology*, *20*, 23–34. https://doi.org/10.1016/j.jaad.2023.04.030

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer. https://clck.ru/37AU6Z

Brooks, R. A. (1991). New approaches to robotics. *Science*, *253*(5025), 1227–1232.

Critchlow, A. J. (1985). *Introduction to robotics*. Macmillan Pub Co.

De Felice, F., Petrillo, A., De Luca, C., & Baffo, I. (2022). Artificial Intelligence or Augmented Intelligence? Impact on our lives, rights and ethics. *Procedia Computer Science*, *200*, 1846–1856. https://doi.org/10.1016/j.procs.2022.01.385

Hellwig, J., Huggett, S., & Siebert, M. (2019). *Artificial Intelligence: How knowledge is created, transferred, and used*. Elsevier. https://doi.org/10.17632/7ydfs62gd6.2

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges, and research agenda. *International Journal of Information Management*, *48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Duda, R. O., & Hart, P. E. (2006). *Pattern classification*. John Wiley & Sons.

Ferreira, A. L., & Lipoff, J. B. (2023). The complex ethics of applying ChatGPT and language model artificial intelligence in dermatology. *Journal of the American Academy of Dermatology*, *89*(4), e157-e158. https://doi.org/10.1016/j.jaad.2023.05.054

Glaser, J. I., Benjamin, A. S., Farhoodi, R., & Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. *Progress in Neurobiology*, *175*, 126–137. https://doi.org/10.1016/j.pneurobio.2019.01.008

Joachim von Braun, Margaret S. Archer, Gregory M. Reichberg, & Marcelo Sánchez Sorond. (2021). AI, Robotics, and Humanity: Opportunities, Risks, and Implications for Ethics and Policy. In J. von Braun et al. (Eds.), *Robotics, AI, and Humanity*. https://doi.org/10.1007/978-3-030-54173-6_1

Kumar, P., Chauhan, S., & Kumar, L. A. (2023). Artificial Intelligence in Healthcare: Review, Ethics, Trust Challenges & Future Research Directions. *Engineering Applications of Artificial Intelligence*, *120*, 105894. https://doi.org/10.1016/j.engappai.2023.105894

Leeson, P. T., & Coyne, C. J. (2005). The economics of computer hacking. *JL Econ. & Pol'y*, *1*, 511.

Lesandrini, J., Idris, M. Y., & Reis, D. S. (2023). The Ethics of Artificial Intelligence and Machine Learning. *Journal of Radiology Nursing*, *42*(3), 265–266. https://doi.org/10.1016/j.jradnu.2023.05.001

Memarian, D., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, *5*, 100152, https://doi.org/10.1016/j.caeai.2023.100152

Mueller, V. C. (2012). Introduction: philosophy and theory of artificial intelligence. *Minds and Machines*, *22*(2), 67–69. https://doi.org/10.1007/s11023-012-9278-y

Palladino, N. (2023). A 'biased' emerging governance regime for artificial intelligence? How AI ethics get skewed moving from principles to practices. *Telecommunications Policy*, *47*(5), 102479. https://doi.org/10.1016/j.telpol.2022.102479

Rich, E., Knight, K., & Nair, S. (2009). *Artificial Intelligence*. Tata McGraw Hill.

Saveliev, A. M., Zhurenkov, D. A., Poikin, A. E., & Berkutova, T. A. (2021). Ethics of Artificial Intelligence and Post-non-classical Scientific Rationality. *IFAC-PapersOnLine*, *54*(13), 397–401. https://doi.org/10.1016/j.ifacol.2021.10.480

Schank, R. C. (1991). Where's the AI? *AI magazine*, *12*(4), 38. https://doi.org/10.1609/aimag.v12i4.917

Scimeca, M., & Bonfiglio, R. (2023). Comment on redefining authorship in the era of artificial intelligence: balancing ethics, transparency, and progress. *ESMO Open*, *8*(5), 101635. https://doi.org/10.1016/j.esmoop.2023.101635

Stahl, B. C., & Wright, D. (2018). Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Security & Privacy*, *16*(3), 26–33. https://doi.org/10.1109/msp.2018.2701164

Stahl, B. C., Brooks, L., Hatzakis, T., Santiago, N., & Wright, D. (2023). Exploring ethics and human rights in artificial intelligence – A Delphi study. *Technological Forecasting and Social Change*, *191*, 122502. https://doi.org/10.1016/j.techfore.2023.122502

Stahl, B. C. (2021). Perspectives on Artificial Intelligence. In *Artificial Intelligence for a Better Future. SpringerBriefs in Research and Innovation Governance* (pp. 7–17). Springer, Cham. https://doi.org/10.1007/978-3-030-69978-9_2

Turing, A. M. (2009). *Computing machinery and intelligence* (pp. 23–65). Springer Netherlands.

Xiao-Fan, L., Zhaoyang, W., Wei, Zh., Guoyu, L., & Gwo-Jen, H. (2023). Technological support to foster students' artificial intelligence ethics: An augmented reality-based contextualized dilemma discussion approach. *Computers & Education*, *201*, 104813. https://doi.org/10.1016/j.compedu.2023.104813

Zhang, Y., Wu, M., Yijun Tian, G., Zhang, G., & Lu, J. (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems*, *222*, 106994. https://doi.org/10.1016/j.knosys.2021.106994

Информация об авторе

Ядав Нидика – PhD, независимый исследователь
Адрес: 110074, Индия, г. Дели, Чаттапур Анклав Фейз II
E-mail: yadavnidhika68@gmail.com
ORCID ID: https://orcid.org/0000-0002-0165-6453
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57195398958
WoS ID: https://www.webofscience.com/wos/author/record/2269009
Google Scholar ID: https://scholar.google.com/citations?user=fndffSwAAAAJ

Конфликт интересов

Автор сообщает 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: 10.41.51 / Охрана авторских прав
Специальность ВАК: 5.1.3 / Частно-правовые (цивилистические) науки

История статьи

Дата поступления – 23 июля 2023 г.
Дата одобрения после рецензирования – 25 октября 2023 г.
Дата принятия к опубликованию – 30 ноября 2023 г.
Дата онлайн-размещения – 15 декабря 2023 г.