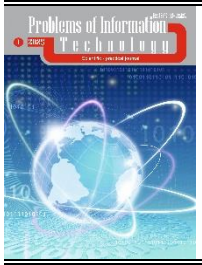

Available online at www.jpit.az

16 (1)
2025

Principles of ethical artificial intelligence

Zafar Jafarov[a], Atif Namazov[b], Javid Abbasli[c], Aygun Abdullayeva[d]

[a,b,c,d] Azerbaijan Technical University, H.Javid ave 25, AZ 1073 Baku, Azerbaijan

[a] zafar.cafarov@aztu.edu.az; [b] atif.namazov@aztu.edu.az; [c] cavid.abbasli@aztu.edu.az; [d] aygun.abdullayeva@aztu.edu.az

[a] https://orcid.org/0009-0001-9327-1779; [b] https://orcid.org/0009-0008-2252-7328 ; [c] https://orcid.org/0009-0005-8982-0164 ;

[d] https://orcid.org/0009-0004-4252-6382 ;

ARTICLE INFO

Keywords:

ethics
digital ethics
artificial intelligence
ethical principles

ABSTRACT

Terms such as algorithms, data, and artificial intelligence (AI) are becoming part of our daily lives. Undoubtedly, they are widely talked about and promise a more successful future. But what risks can losing control over their development pose? In order to respond to the challenges of accountability in the development of AI, many government agencies, private enterprises, and international organizations have published charters of practice, declarations of principles, and recommendations in this direction. They have demonstrated convergence around basic principles. However, principles are sometimes abstract and are not always defined in the same way all over the world. Much work remains to be done in this area. We need to think about the implementation of general ethical principles, ensure that they can be applied to each area of specific activity (education, science, healthcare, etc.), and the proposals arising from this thinking should be implemented in practice. It is important to involve citizens more in defining guidelines for the responsible use of AI and big data and to collect their informed opinions. Indeed, the application of artificial intelligence affects everyone and raises ethical and political questions that are the subject of public debate.

1. Introduction

The word "ethics" is a philosophical concept that studies what is right and wrong, social norms, values, and moral principles. This concept refers to the process of thinking, discussing, and analyzing people's ethical values and behaviors. Ethics determines the way we think and act about how people interact with each other in society, how they make decisions, and how they direct their lives (Anderson et al., 2007).

Ethics refers to a set of principles and standards that should be followed for right action, such as the principle of not harming others or rules that prohibit wrongdoing. Ethics is said to be prescriptive because it determines what is necessary and acceptable according to accepted values. Ethical standards describe accepted moral values. Ethics also refers to the philosophical subject that seeks to determine these standards and values.

The main principles of ethics are as follows:

- Normative ethics covers various concepts that we can call ethical behavior. It consists of three main schools of thought:
- Virtue ethics: describes the moral character of an action in terms of the goodwill that accompanies it. People speak of actions that are, for example, courageous, just, and generous. In this concept, the domains of human activity and moral characteristics are most important.
- Consequentalism: focuses on the effects of actions, their short- and long-term positive or negative consequences. Utilitarianism, which is

Received 12 September 2024, Received in revised form 18 November 2024, Accepted 29 November 2024
http://doi.org/10.25045/jpit.v16.i1.06

also considered a branch of consequentialism, begins with the fact that we aim to achieve outcomes that maximize the life chances and well-being of the most people.

- Deontology: examines whether an action is in accordance with a duty (often a professional duty) or a set norm. It tends to develop into specific codes for a profession. It is based on a framework designed to establish principles of behavior within a group.
- Meta-ethics examines the nature and foundations of ethics. This constantly feeds into normative ethics by renegotiating the proposed principles, the criteria for ethical evaluation of action.
- Applied ethics applies normative ethics in practice by comparing a specific situation with the principles derived from different schools of normative ethics. Ethical dilemmas are always resolved using applied ethics.
- Descriptive ethics, also known as comparative ethics, studies people's beliefs about morality. It contrasts with meta-ethics, which examines what ethical theories, terms, and concepts that define how people should behave are actually about.

2. Digital and artificial intelligence ethics

Digital ethics refers to a set of moral principles and values that guide behavior in cyberspace. As technology continues to develop and play an increasingly important role in our daily lives, it is important to consider the ethical implications of our actions in the digital environment.

Digital ethics covers a wide range of issues, including data privacy, online security, intellectual property, cyberbullying, digital addiction, and artificial intelligence. It is important to consider how our online activities affect others and the wider society. Digital ethics also involves examining the responsibilities of companies and individuals who create and use technology.

The main categories of ethics applied to digital technology are:

Project ethics focuses on the design phase of digital tools. Project ethics is directly related to technology, with all its technical complexity, and the knowledge of engineers and programmers. Therefore, project ethics specifically addresses the deontology (duty-based ethics) of all types of digital creators (developers, digital designers, project managers). Indeed, to the extent that data or algorithms can replicate or create human biases, reveal (or reproduce on a larger scale) new discriminations, and cause injustices, they have an ethical responsibility, starting from the project stage. “Project ethics” refers to the ethical rules and principles that must be followed in the processes of planning, executing, and terminating a project or task. This concept aims to ensure that a business or project is managed ethically and responsibly. Good ethics provide guidance for project managers and team members and prevent unethical behavior. Good ethics are essential for successful project management and for maintaining ethical values and responsibilities at all stages of the project. These principles ensure not only the technical success of the project, but also its implementation in accordance with social and ethical values.

“Use ethics” generally refers to the concept of using technology and digital tools. This concept includes a set of specific rules and values regarding how individuals, institutions, or communities use technology. Usage ethics requires an ethical assessment of how people use the technological resources at their disposal.

Social ethics examines the effects of digital technology on society at the macroscopic level. Ultimately, social ethics is an attempt to define ethical norms that are accepted at a social level, focusing on the general good of society. It aims to contribute to the development and well-being of society.

The term Artificial Intelligence (AI), recognized as an academic discipline by the Dartmouth Seminar, encompasses all conceptualizations of intelligent machines in terms of both operational and social outcomes. As the availability of information around us increases, people will increasingly rely on artificial intelligence systems to live, work, and play. Given the increasing accuracy and complexity of artificial intelligence systems, they will be used in an increasingly diverse range of sectors, including finance, pharmaceuticals, energy, manufacturing, education, transportation, and public services.

The first person to open the philosophy of artificial intelligence was the famous English logician and mathematician Alan Turing. In 1950, six years before the Dartmouth conference, Turing

published an article in the August issue of the philosophical journal Mind entitled "Computing Machinery and Intelligence." In this article, Turing asks: "Can machines think?" He carefully opens the question to philosophical discussion and rejects objections to the claim that machines can think.

Edward Fredkin, one of the heads of the Computer Science laboratories at MIT (Massachusetts Institute of Technology), used the following expressions in an interview with the BBC (British Broadcasting Corporation): "There have been three great events in history. The first is the creation of the universe. The second is the beginning of life. The third is the discovery of artificial intelligence."

"Just as electricity changed almost everything 100 years ago, I find it hard to think of an industry today that AI won't change in the next few years," says Andrew Yan-Tak Ng, a British-born American computer scientist and technology entrepreneur focused on machine learning and artificial intelligence, and the co-founder and CEO of Google Brain. He notes that electricity has transformed transportation, industry, agriculture, and healthcare. Artificial intelligence is having a similar impact. Two problems are holding back its adoption: a lack of information and a lack of expertise.

Renowned physicist Stephen William Hawking has noted that "Artificial intelligence can continue to evolve and even reinvent itself. Humans, limited by the extremely slow rate of biological evolution, cannot compete with such power."

The ethics of artificial intelligence is a branch of the ethics of digital technology that is specific to intelligent systems. Sometimes it is divided into concerns about the behavior of machines in the moral and digital ethics of the individual humans who design, develop, deploy, and use those systems. It also encompasses the issue of possible singularity through superintelligent artificial intelligence. The ethics of artificial intelligence is one of the ethical goals of seeking to identify and prevent abuses while promoting its benefits.

The ethics of artificial intelligence is a set of standards and values that apply to its development and use. As such, it is a limited field of ethics, but since artificial intelligence technologies can disrupt social relations and have very profound harmful consequences, this field of ethics is considered crucial and is in great development.

Finally, ethical AI is a part of public ethics (i.e., applied to public disputes that require a solution acceptable to those who disagree with it). This also applies to environmental ethics or bioethics.

3. Ethical problems of AI and Montreal declaration

Some of the main ethical issues related to AI are:

- Bias and discrimination: AI systems can perpetuate and reinforce existing bias and discrimination in society, which can lead to unfair outcomes for certain groups of people.
- Transparency and explainability: It is important to ensure that AI systems are transparent and explainable, meaning that their decisions and actions can be understood and scrutinized by humans.
- Privacy and data protection: AI systems often rely on large amounts of personal data, and it is important to ensure that this data is collected, used and stored in a way that respects individuals' privacy and data protection rights.
- Accountability and responsibility: It is important to establish clear lines of accountability and responsibility for AI systems, including who is responsible for their development, deployment and use.
- Safety and reliability: AI systems can have a significant impact on human life, and it is important to ensure that they are designed and tested to be safe and reliable.
- Human control and autonomy: AI systems should be designed to serve humans and respect human autonomy, rather than replace or control human decision-making.

The Three Laws of Robots are laws about the functions and rights of robots put forward by the American science fiction writer Isaac Asimov (Calo, 2017). These laws have been used repeatedly in popular culture and science fiction. They are also important in terms of modern robot ethics. In his article My Robots, he stated that he wrote the laws one by one in the story "The Chase", which was published in Astounding magazine in March 1942. In that article, he stated that he was willing to use these laws by other authors, provided that they were not used individually. In his book I, Robot, which contains this story, he also discusses the problems caused by these laws. The three laws are: A robot

may not injure a human being or allow a human being to come to harm; A robot must obey human orders except where this would conflict with the first law; A robot must protect its own existence as long as this does not conflict with the first two laws.

Isaac Asimov later saw the need for a new law due to the complexity of robots. In his 1985 novel Robots and the Empire, he wrote that advanced robots would prevent all of humanity from harming any human being. He called this the "Zenth Law of Robotics".

Addressing the ethical challenges of artificial intelligence requires collaboration between a variety of stakeholders, including AI developers, policymakers, ethicists, and civil society organizations. A number of organizations and initiatives have been established to promote AI ethics and to develop guidelines and standards for the responsible development and application of AI.

Christine Tappolet, a philosophy professor at the University of Montréal, and the Fondation du Recherche Québec, together with the Fondation du Recherche Québec, will form part of the Montreal Declaration for a Responsible Development of Artificial Intelligence. The participants of the Forum on the Socially Responsible Development of Artificial Intelligence, held in Montreal in November 2017, drafted a preamble to the declaration, inviting the public to contribute to the discussion on responsible AI. The preamble outlined the ethical challenges of artificial intelligence. The recommendations to propose a code of ethics for the development of artificial intelligence and, within it, guidelines for completing the digital transition, were developed after a year of research and studies, based on consultations with citizens, experts, politicians, industry stakeholders, non-governmental organizations, and professional societies.

4. Ethical principles of AI

The principles of ethical artificial intelligence (AI) are rules that are set to ensure that artificial intelligence systems operate in a fair, transparent, safe and human rights-respecting manner. The most important ethical AI principles are listed below:

1. Principle of well-being

The development and use of artificial intelligence systems should contribute to the advancement of the well-being of all living beings.

2. Principle of respect for autonomy

AI should be developed with the aim of increasing people's control over their environment and their own lives, taking into account their freedom.

3. Principle of privacy

Privacy and privacy should be protected from intrusion by artificial intelligence, data collection and archiving systems.

4. Principle of solidarity

AI development should not disrupt relationships between people and generations.

5. Principle of democratic participation

AI should meet the criteria of comprehensibility, validity and accessibility and be subject to democratic review, discussion and control.

6. Precautionary principle

Everyone contributing to the development of artificial intelligence should be aware of the negative consequences and take the necessary measures to prevent them.

7. Principle of internal diversity

The development and use of artificial intelligence should be consistent with the preservation of the social and cultural diversity of society and should not restrict lifestyle choices and personal experiences.

8. Principle of justice

The application of artificial intelligence should contribute to the creation of a just and egalitarian society.

9. Principle of responsibility

The development and use of artificial intelligence should not help reduce people's responsibility in decision-making.

10. Principle of sustainable development

The development and use of artificial intelligence should be carried out in a way that ensures the ecological sustainability of the planet.

The principles are not hierarchical. The last principle is as important as the first. Depending on the circumstances, one principle can be given more importance than another or one principle can be considered more relevant than another. It should also be emphasized that these are ethical principles, not legal norms, and not governance rules.

Recommendations on the ethics of artificial intelligence

Based on the Montreal ethical principles of artificial intelligence, the recommendations have been developed to propose guidelines for organizing the digital transition.

1. An independent organization.

An organization should be established to study and investigate the use of artificial intelligence and digital technology and its social impacts

2. Audit and certification

A policy should be implemented for the audit and certification of artificial intelligence that promotes responsible practices

3. Access to education that teaches understanding, critical thinking, respect and responsibility should be supported so that citizens have greater access to digital technologies, thus promoting active participation in a sustainable digital society.

4. Education and Ethics

The education of industry stakeholders on the use, design and development of artificial intelligence should be reviewed.

5. Inclusive development

A comprehensive strategy should be implemented using various existing institutional resources to promote the inclusive development of AI and prevent potential bias and discrimination in its application.

6. Democracy protection

A containment strategy should be implemented to prevent deception and political manipulation of citizens, plagiarism. It is important to create conditions for the healthy functioning of democratic institutions and informed citizens.

7. Environmental impact

A public/private sector strategy should be implemented in the development and use of AI that is consistent with environmental sustainability and allows for improvements in the resolution of environmental crises.

8. International development

A non-aggressive international development model that aims to cover the world should be implemented.

5. Conclusion and future works

The ethics of artificial intelligence addresses the ethical issues of artificial intelligence technologies and aims to adopt a more just, responsible and human-centered approach, taking into account their impact on people and societies. Laws and regulations are often not enough to ensure the ethical use of artificial intelligence. Users and businesses using artificial intelligence, as well as those who develop and provide such tools and technologies, are obliged to apply the ethics of artificial intelligence. Users and suppliers of artificial intelligence must take real steps to ensure that they use artificial intelligence ethically. This obligation should not be limited to making statements, but should be specific policies that are actively implemented.

This research paper discusses the ethical principles of artificial intelligence, the main ethical problems and ways to solve them. By examining the development of artificial intelligence technologies and their impact on society, the main principles of ethical artificial intelligence have been identified. The main conclusions include the following:

- Appropriate mechanisms should be developed to ensure that artificial intelligence is impartial, transparent and accountable.
- AI systems that respect human rights and privacy should be built.
- It is important to develop AI with social and environmental sustainability in mind.
- Legislative and regulatory frameworks for the application of ethical AI should be improved.
- If ethical principles of AI are not applied, its negative impacts on society may increase. Therefore, technology developers, policymakers and researchers should work together to ensure the development of this technology in accordance with ethical norms.

This research can be expanded in the following directions in the future:

- Applied Ethics: To further explore the ethical application of AI in various fields (health, education, law, etc.).
- Regulation and Legislation: To develop an ethical code and international legislative mechanisms for AI.
- Bias Reduction: To develop new methodologies to detect and minimize bias in AI systems.
- Human-Machine Interaction: To further analyze the impact of artificial intelligence on human decision-making and to establish ethical frameworks.
- Ecological Sustainability: To explore ways to reduce the energy consumption and environmental impact of artificial intelligence technologies.

Future research in these areas will ensure more ethical and responsible development of artificial intelligence and create conditions for its more beneficial use for society.

References

Anderson, M., & Anderson, S. L. (2007). Ethics for synthetic beings. AI & Society, 22(3), 313-323. www.doi.org/10.1007/s00146-007-0146-3

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In The Cambridge Handbook of Artificial Intelligence (pp. 316-334). Cambridge University Press. www.doi.org/10.1017/CBO9781139046855.020

Brundage, M. (2014). Limitations and risks of machine ethics. Journal of Experimental & Theoretical Artificial Intelligence, 26(3), 355-372. www.doi.org/10.1080/0952813X.2014.895111

Bryson, J. J. (2018). Patiency is not a virtue: The design of intelligent systems and systems of ethics. Ethics and Information Technology, 20(1), 15-26. www.doi.org/10.1007/s10676-018-9448-6

Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. UC Davis Law Review, 51, 399-435. www.doi.org/10.2139/ssrn.3015350

Coeckelbergh, M. (2020). AI ethics. MIT Press. www.doi.org/10.7551/mitpress/12230.001.0001

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). www.doi.org/10.1162/99608f92.8cd550d1

Gunkel, D. J. (2012). The machine question: Critical perspectives on AI, robots, and ethics. MIT Press. www.doi.org/10.7551/mitpress/9780262017435.001.0001

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399. www.doi.org/10.1038/s42256-019-0088-2

Lin, P., Abney, K., & Bekey, G. A. (2011). Robot ethics: The ethical and social implications of robotics. MIT Press. www.doi.org/10.7551/mitpress/9780262526005.001.0001

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. IEEE Intelligent Systems, 21(4), 18-21. www.doi.org/10.1109/MIS.2006.80

Müller, V. C. (2020). Ethics of artificial intelligence and robotics. In The Stanford Encyclopedia of Philosophy. www.doi.org/10.5840/stanford/202032445

Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. AI Magazine, 36(4), 105-114. www.doi.org/10.1609/aimag.v36i4.2577

Sharkey, A. J. C., & Sharkey, N. E. (2010). Granny and the robots: Ethical issues in robot care for the elderly. Ethics and Information Technology, 14(1), 27-40. www.doi.org/10.1007/s10676-010-9234-6

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. ACM Journal of Data and Information Quality, 11(4), 1-9. www.doi.org/10.1145/3236006

Tegmark, M. (2017). Life 3.0: Being human in the age of artificial intelligence. Knopf. www.doi.org/10.1038/d41586-018-05868-2

Turilli, M., & Floridi, L. (2009). The ethics of information transparency. Ethics and Information Technology, 11(2), 105-112. www.doi.org/10.1007/s10676-009-9187-9

Vallor, S. (2016). Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press. www.doi.org/10.1093/oso/9780190498511.001.0001

Wallach, W., & Allen, C. (2008). Moral machines: Teaching robots right from wrong. Oxford University Press. www.doi.org/10.1093/acprof:oso/9780195374049.001.0001

Winfield, A. F. T., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180085. www.doi.org/10.1098/rsta.2018.0085

Yampolskiy, R. V. (2013). Artificial intelligence safety engineering: Why machine ethics is a wrong approach. In Philosophy and Theory of Artificial Intelligence (pp. 389-396). Springer. www.doi.org/10.1007/978-3-642-31674-6_28

Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking artificial intelligence principles. www.doi.org/10.1109/ICDM.2018.00138