

Journal of Clinical Practice and Medical Case Report

Genesis-JCPMCR-2(1)-21
Volume 2 | Issue 1
Open Access
ISSN: 3048-8206

Ethics and Responsibility in The Development and Implementation of Artificial Intelligence

Tricia Bogossian*

Master in Work Management for Quality of the Built Environment - Multidisciplinary

***Corresponding author:** Tricia Bogossian, Master in Work Management for Quality of the Built Environment Multidisciplinary

Citation: Bogossian T, Ethics And Responsibility in The Development and Implementation of Artificial Intelligence. J Clin Pract Med Case Rep. 2(1):1-10

Received: January 20, 2025 | **Published:** February 9, 2025.

Copyright© 2025 genesis pub by Bogossian T. CC BY-NC-ND 4.0 DEED. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License. This allows others distribute, remix, tweak, and build upon the work, even commercially, as long as they credit the authors for the original creation.

Abstract

Artificial intelligence represents a significant advance in contemporary society, offering innovative solutions for different areas. However, its development requires attention to ethical and responsibility issues to avoid negative impacts, such as algorithmic biases, privacy risks and lack of transparency in automated decisions. Given this scenario, this study aims to discuss ethics and responsibility in the development and implementation of artificial intelligence. In order to explain the concepts related to the subject, a theoretical survey was carried out, using bibliographic research in books, articles and other scientific research with physical or virtual access that deals with the studied topic. It was concluded that regulation and the establishment of ethical standards are fundamental to guide its responsible use, and audit mechanisms, continuous monitoring and professional training are essential to ensure its reliability. In this way, artificial intelligence can be applied in an ethical and transparent manner, contributing to social and economic progress without neglecting the principles of equity, responsibility and respect for human rights.

Keywords

Artificial Intelligence; Ethics; Responsibility.

Introduction

Artificial intelligence (AI) has become one of the most transformative technologies of Nurse at the State Health Department of RJ and Maternity of UFRJ. Judicial Proposal of CONPEJ, Judicial Expert in Nursing. Master in Work Management for Quality of the Built Environment. Specialist in Adult Intensive Care Nursing and Neonatal Nursing from UERJ and Occupational Nursing (UFRJ).

Our time, impacting different areas, such as healthcare, finance, education and security. Its rapid advancement enables the automation of processes, the analysis of large volumes of data and autonomous decision-making, becoming an essential tool for optimizing services and solving complex problems. However, along with the benefits provided, ethical and social challenges arise that require in-depth reflection on the limits and responsibilities in the creation and application of these technologies.

The main ethical challenges in the development and implementation of AI include the transparency of algorithms, data privacy and security, algorithmic biases, and responsibility for automated decisions. The lack of clear regulation and control mechanisms can generate negative consequences, such as discrimination, violation of fundamental rights, and unpredictable impacts on the labor market and society. Therefore, ensuring that AI is developed and used in an ethical and responsible manner is one of the greatest contemporary challenges.

The need to discuss ethics and responsibility in the development of AI is justified due to its growing impact on daily life and social relationships. The indiscriminate and uncritical use of this technology can lead to serious harm, such as the amplification of social inequalities and the manipulation of information. Furthermore, the absence of well-established ethical guidelines can compromise society's trust in these innovations, making it difficult for them to be accepted and adopted in a safe and beneficial manner.

Given this scenario, this study aims to discuss ethics and responsibility in the development and implementation of artificial intelligence. In order to explain the concepts related to the theme, a theoretical survey was carried out, using bibliographic research in books, articles and other scientific research with physical or virtual access that dealt with the theme studied.

Definition of artificial intelligence (AI)

The history of technology is remarkably impressive. Its emergence can be traced back to the third industrial revolution in the 1950s and continues to this day. With the advent of computing on a global scale in the 2000s, following the fourth industrial revolution that introduced cyber-physical systems, technological advances have been relentless. Technology today is a fundamental element that shapes contemporary societies, and it is impossible to talk about evolution without mentioning the crucial role of technological tools in this process.

In this scenario, other new technologies emerge, including artificial intelligence. According to computer scientist John McCarthy, artificial intelligence is the science and engineering of making intelligent machines—especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI need not be limited to methods that are biologically observable [1].

Artificial intelligence is a result of both human advancement and technological evolution – it is increasingly becoming an integral part of the development of societies in this era. It is therefore important to understand how artificial intelligence plays a role in catalyzing global technological advances. AI is a field of study that has played an increasingly significant role in our society, driving innovation across a range of sectors and raising important questions about the nature of machine intelligence and autonomy [2].

A classic definition of AI, put forward by John McCarthy in 1956, describes AI as “the science and engineering of making machines intelligent.” This suggests that the goal of AI is to create systems that can perform tasks that normally require human intelligence, such as learning, reasoning, problem-solving, and decision-making. However, this definition often raises the question of how we define intelligence itself [3].

Furthermore, the definition of AI has evolved over time to include not only the ability to perform intelligent tasks, but also autonomy and the ability to learn from data. AI now includes systems that can adapt their behavior based on past experiences and improve their performance over time, such as machine learning systems and artificial neural networks.

[3] presents a definition of artificial intelligence as the study of systems that exhibit behaviors that can be considered intelligent by external observers. However, he acknowledges that this definition does not fully encompass the entire field of AI, particularly when it comes to more complex systems. At the same time, the author also suggests that artificial intelligence involves the application of methods inspired by the intelligent behavior of humans and other animals to solve problems of great complexity.

On the other hand, [4] offers an alternative definition of AI as a subfield of computer science that focuses on the automation of intelligent behaviors. He justifies the adoption of this definition as follows:

This definition is particularly appropriate for this book because it reinforces our belief that artificial intelligence is part of computer science and as such should be based on sound theoretical and applied principles in the field. These principles include the data structures used to represent knowledge, the algorithms needed to apply that knowledge, and the programming languages and methods used to implement it [4].

[5] noting that all the good ideas in physics seem to have already been developed by Galileo, Newton, Einstein and others, they note that “AI still has room for new Einsteins.” And they conclude:

Today, artificial intelligence encompasses a wide range of subfields, from general-purpose areas such as learning and perception to specific tasks such as playing chess, proving mathematical theorems, composing poems, and diagnosing diseases. AI systematizes and automates intellectual tasks, so it is potentially relevant to any sphere of human intellectual activity. In this sense, it is a truly universal field [5].

Although they do not define it – in fact, they write that “AI is interesting, but we haven’t said what it is yet” – they compare definitions from different authors and explain that they are sometimes related to thought and reasoning processes, sometimes to behavior, sometimes to the success of the system in terms of fidelity to human action, sometimes to “success in relation to an ideal conception of intelligence

that we will call rationality". A rational system "'does the right thing' with the data it has" [5]. Strictly speaking, artificial intelligence suggests seeking an answer to the following question: can computers be made to perform human-like tasks, or can machines be made to understand things, in other words, adopt "intelligent" behavior? [6].

It is an artificial intelligence research task that deals with a set of methods by which a computer emulates some human capabilities, including problem solving, natural language understanding, vision and robotics, expert systems and knowledge acquisition, knowledge representation methodologies.

Although there is no consensus on the meaning of this expression, it can be established that artificial intelligence describes the ability of machines to think to a certain extent or, better, to imitate human thinking, learning to use generalizations commonly used by humans, their usual solutions.

With this ability to imitate human thought, AI involves the use of cognitive functions such as language, planning, memory and perception, all of which are performed artificially and studied by both information technology and computer science. It is based on knowledge of statistics and probability, logic and linguistics. Based on this knowledge, it models the processes of human intelligence with the help of computational resources. [2, 22].

The National Council of Justice, in Resolution No. 332/220, which provides for the use of artificial intelligence in the Brazilian judiciary, defined the artificial intelligence model (article 3, II):

[...] set of data and computational algorithms, designed from mathematical models, whose objective is to offer intelligent results, associated or comparable to certain aspects of thought, knowledge or human activity [7].

Finally, AI consists of "a computer system designed to rationally model human decision-making, attempting to translate the functioning of the human brain into algorithms." It models human reasoning and performs intellectual tasks. The combination of several technologies allows "a machine to understand, learn, identify, or complete a human activity [8].

The evolution of AI

The impacts of AI vary depending on the type and area of application, and can result in both job replacement and job complementation in different areas. These characteristics are distinctive to AI, as it optimizes processes and increases productivity at various stages of the production chain, resulting in more benefits with less effort. For example, in agriculture, the automation of seeders and harvesters with systems such as GPS and autopilot help distribute fertilizers across large plantations, keeping operating costs low. This type of AI replaces human labor, contributing to unemployment, especially on large plantations. However, due to the high acceptance of the technology due to its effectiveness, this process is expected to increase even more. Global investment in smart agricultural technologies and systems is expected to triple in revenue by 2025, reaching \$15.3 billion. In addition, specific spending on AI technologies and solutions is expected to increase from \$1 billion in 2020 to \$4 billion in 2026, representing a compound annual growth rate of 25.5% [9].

On the other hand, AI in recommendation algorithms, such as those used by Netflix to suggest movies and series based on user preferences, does not replace human work, as it would be unfeasible to do it manually for each user account. In this case, AI creates a new process that benefits the final product and provides employment for those who manage the algorithm. Therefore, different types of AI have different impacts.

Another instance occurs when a process that should be performed manually ends up not being done due to various reasons, such as high cost, inefficient execution, or incompleteness. One example of this is the monitoring of tax fraud related to household taxation in France, which was addressed using an experimental AI algorithm as a solution. This system was implemented in regions where household taxation is a significant part of revenue collection, such as Ville Rennes, where 70% of taxes come from households, as well as in eight other French departments, out of a total of 101. AI analyses drone footage to identify residential pools and identify those that have not been properly declared and taxed. In these eight departments alone, more than 20,000 pools have not been declared, representing 5.7% of those currently legal. Taxing these pools could result in around €10 million in additional tax revenue for 2022, with estimates of reaching €40 million in 2023. This further highlight that different types of AI have different impacts [10].

The same AI system can have different impacts in different areas, such as a facial recognition app. When used on a phone to unlock the screen, the innovation simply replaces the previous process of typing a password on the keypad, thus changing the role of the technology. However, when a company that employs a doorman to control entry to the establishment adopts an automatic facial recognition system to allow registered people to enter, the AI replaces the human role. Thus, a technology with the same function of recognizing human faces, when implemented in different ways, in different areas, can generate different consequences. In addition, the technology itself may face other problems; for example, the same facial recognition system may be used for surveillance and monitoring, raising ethical questions that will be discussed later in this article.

The impacts of AI also vary depending on the location and demographic group in question. A simple example is the use of AI in agriculture mentioned above: its implementation in a country with little agricultural activity will not have as significant results as in a country like Brazil, which has vast tracts of productive land. However, Brazil is not yet well prepared for this technology due to low investment in AI compared to other countries. Thus, even though European countries have less productive territories, they are more exposed to AI technologies in agriculture than Brazil. The growth of AI in the agricultural sector will have a much greater impact in developed and technologically advanced countries, such as the United States, which is currently the largest user, and the European Union, which has the highest growth forecast [10].

A more specific example is China, a global leader in AI adoption. There, there is potential to fully digitize supermarkets, from customer service to inventory control, which could increase efficiency and profit margins, reduce operating costs, satisfy customers and provide competitive advantages. However, this process is lagging behind compared to other countries, despite the Chinese food retail market growing by

around 3% annually between 2015 and 2020. China's top 20 food retailers account for just 15% of the market, while in the West the top five account for between 40% and 60%. This fragmentation results in lower efficiency and an inability to adapt quickly to changing trends and consumer decisions, which explains why the Chinese have been unsuccessful or lagging behind in automating this specific area [11].

Thus, geographic location also influences the scope of AI's impact. These discrepancies in terms of implications, along with other AI-related criteria, will be analyzed in this chapter. Different AI models will be examined, taking into account their characteristics such as location, industry, and demographic group, focusing on the most prominent criteria and the patterns, if any, that can be identified among them. Finally, the prognosis of advances in AI and related technologies will be discussed, which, due to their diverse impacts, may represent a critical turning point in economic history.

Ethics and Artificial Intelligence

According to [12], the debate on Artificial Intelligence (AI) raises important ethical questions, becoming "a field of research and forces in which promises and disputes about conservation, revolution and ways of proceeding are in constant conflict". The growing presence of AI in society requires reflections on its application, impacts and risks, especially with regard to prejudices and discrimination reproduced by algorithmic systems.

According to [13], the data that feeds AIs is not neutral and may be loaded with biases and prejudices, perpetuating these distortions during the machine learning process. Since algorithms learn from the data provided, any bias present in this set of information can influence the results generated, often without programmers being fully aware of it. This problem becomes even more critical when the data used to train AI models does not reflect the diversity that exists in society.

A notable example of this occurred in 2016, when an AI was implemented in a beauty pageant to select the most attractive contestants, with the promise of an impartial judgment, without interference from sociocultural factors. The system, trained through machine learning, was supposed to analyze and make decisions based on an extensive database. However, at the end of the contest, it was found that the AI had predominantly chosen white people as the most beautiful.

Analysis by experts revealed that the problem was in the data used to train the machine. Most of the images provided were of white people, many of them from Hollywood backgrounds, which created a biased model. This shows that the machine intelligence is directly related to the quality and representativeness of the data it is fed. If diversity had been considered in the training process, the results would probably have been more balanced [14].

Data Privacy and Security

Data privacy and security are fundamental ethical challenges in the implementation of AI. With the increasing use of automated systems in different areas, it becomes essential to ensure that information is stored and processed securely, preventing unauthorized access, leaks and misuse [15]

The large amount of sensitive data involved, such as personal, financial and corporate information,

requires the adoption of strict measures to ensure the confidentiality and integrity of the information. In addition, compliance with laws and regulations, such as the General Data Protection Law (LGPD) in Brazil and the General Data Protection Regulation (GDPR) in the European Union, is essential to ensure the responsible use of these technologies.

According to [16], the integration of artificial intelligence raises significant issues regarding data privacy, requiring a careful review of security and protection practices to ensure the confidentiality and integrity of information. Companies and institutions that use AI must invest in advanced encryption, robust authentication protocols, and regular audits to minimize risks and ensure the reliability of systems.

Algorithmic bias

Algorithmic bias refers to the tendency of AI systems to reproduce biases present in the data used for their training. This phenomenon can result in discrimination and inequality in a variety of contexts, such as hiring processes, credit granting, legal decisions, and medical diagnoses [17].

For example, an AI system trained on historical credit data may perpetuate discriminatory patterns, denying financing based on unfair criteria. Similarly, an algorithm used for recruitment may favor certain candidate profiles while unjustifiably excluding others.

For Galiana [18], Algorithmic bias is a complex problem that needs to be addressed with a multidisciplinary approach, involving professionals from the areas of technology, law and civil society. Implementing effective solutions to minimize the possibility of discrimination in automated decision-making processes is essential to ensure justice, equity and human rights.

Furthermore, [19] warn that the presence of algorithmic bias can result in distorted interpretations of information, compromising the accuracy and impartiality of automated analyses. Therefore, identifying and mitigating these biases requires continuous monitoring of AI systems, improvement of the databases used, and greater transparency in the algorithms' decision criteria.

An effective approach to reducing the impacts of algorithmic bias includes frequent audits, diversifying training data, and engaging digital ethicists to ensure that automated decisions are fair and impartial.

Accountability and transparency

Accountability and transparency are fundamental principles in the application of artificial intelligence. Since automated systems make decisions that directly impact individuals and organizations, it is essential that their criteria are understood and can be audited.

A lack of transparency can create uncertainty, making it difficult to identify errors and reducing trust in AI systems. Therefore, ensuring that automated processes are explainable and that their decisions can be justified is crucial to avoid failures that compromise the reliability of these technologies.

According to [20], it is essential to ensure transparency in automated processes, allowing professionals and users to understand and audit the decisions made by AI systems, thus promoting responsibility and

trust in the use of these technologies.

Therefore, the development and application of AI must follow clear governance guidelines, establishing oversight and accountability mechanisms to prevent algorithmic errors from resulting in negative impacts on society.

Existing regulations

Regulating artificial intelligence is essential to ensuring the responsible use of this technology. Several laws and regulations have been created to guide data protection and ethics in the development of automated systems, aiming to minimize risks and protect users' rights.

The General Data Protection Law (LGPD) in Brazil and the General Data Protection Regulation (GDPR) in the European Union are examples of regulations that establish strict guidelines for the treatment of sensitive information, imposing penalties for the misuse of data. In addition to these laws, international bodies have developed recommendations on AI governance, such as the European Commission's report on ethics in artificial intelligence.

According to [21], compliance with regulations such as the LGPD is essential to guarantee the protection of personal and business data, ensuring trust in the use of AI.

Furthermore, governments and institutions need to work together to constantly update existing regulations, ensuring that new AI applications are properly regulated and aligned with current ethical and legal principles.

Role of ethical and professional standards

In addition to government regulations, ethical and professional standards play a key role in guiding the responsible use of AI. Organizations and regulatory bodies establish guidelines that must be followed to ensure the integrity of information generated by automated systems.

The use of AI must be guided by transparency, accountability and a commitment to impartiality. Professionals who interact with these technologies need to be prepared to interpret and validate the results presented by the algorithms, ensuring that decisions are well-founded and fair.

As highlighted by [22] ethical and professional standards play a fundamental role in ensuring the integrity and reliability of information generated by AI systems, thus promoting transparency and responsibility in the use of these technologies.

To ensure the ethical and efficient use of artificial intelligence, it is essential that there is ongoing training of the professionals involved, frequent auditing of automated systems and strict adherence to the guidelines established by regulatory bodies. In this way, AI can become a powerful tool for optimizing processes and facilitating decision-making, without compromising ethics and equity in society.

Conclusion

Artificial intelligence represents a significant advance in contemporary society, offering innovative solutions for various sectors. However, its development and implementation require careful consideration of ethical and responsibility issues in order to avoid negative impacts such as algorithmic biases, privacy risks and lack of transparency in automated decisions. Reflecting on these challenges is essential to ensure that AI is used in a fair and equitable manner, promoting social benefits without compromising fundamental rights.

In this scenario, regulation and the establishment of ethical standards are essential to guide the responsible use of AI. The implementation of audit mechanisms, continuous monitoring and professional training are essential measures to ensure the reliability of these technologies. In this way, artificial intelligence can be applied in an ethical and transparent manner, contributing to social and economic progress without neglecting the principles of equity, responsibility and respect for human rights.

References

1. Villas M. (2017) Artificial Intelligence and Industry 4.0. São Paulo: TI INSIDE Online.
2. Peixoto FH, Silva RZM. (2019) Artificial Intelligence and Law. Curitiba: Alteridade.
3. Coppin B. (2015) Artificial intelligence. Rio de Janeiro: LTC.
4. Luger GF. (2013) Artificial intelligence. São Paulo: Pearson Education Brazil.
5. Russel SJ, NORVIG P. (2004) Artificial intelligence. Rio de Janeiro.
6. Direne A. (2022) Artificial Intelligence Overview. Curitiba: Universidade Federal do Paraná.
7. Brazil. (2020) National Council of Justice. Resolution No. 332, of August 21, 2020. Provides for ethics, transparency and governance in the production and use of Artificial Intelligence in the Judiciary and other measures. Brasília, DF: CNJ.
8. Cheliga TTV. (2020) Artificial intelligence: legal aspects. Salvador: Juspodivm.
9. Covarrubias JZL, Enriquez OAM, Guerrero MG. (2022) Regulatory approaches for Artificial Intelligence (IA). Chilean J Law. 49:(3):31-62.
10. De Souza GC, Roveroni AJ. (2023) Artificial Intelligence (AI): The crucial role of regulation. Ibero-American Journal of Humanities, Sciences and Education. 9(10):1982-93.
11. Jacobsen G, Dias BM. (2023) Smart dispute resolution: Artificial intelligence to reduce litigation. Suprema. J Consti Stud. 3(1):391-414.
12. Barbosa XC. (2020) Brief Introduction to the History of Artificial Intelligence. Jamaxi. 4(1):90-97.
13. Garcia ACB. (2020) Ethics and Artificial Intelligence. Computing Brazil. 43:14-22.
14. Bostrom N, Yudkowsky E. (2018) The ethics of artificial intelligence. In: YAMPOLSKIY, R. V. (Ed.). Artificial intelligence safety and security. Boca Raton: Chapman and Hall/CRC. 57-69.
15. Etzioni A, Etzioni O. (2017) Incorporating ethics into artificial intelligence. J Ethics. 21:403-418.
16. Oliveira AB, Motta PR. (2020) Accounting and artificial intelligence: an analysis of the ethical implications. J Account Org. 14(1):1-10.
17. Frazão A. (2018) Algorithms and artificial intelligence. Brasília, DF: Jota.
18. Galiana LI, Gudino LC, González PM. (2024) Ethics and artificial intelligence. Span Clin J. 224(3):178-186.
19. Russell S, Hauert S, Altman R, Veloso M. (2015) Robotics: Ethics of artificial intelligence. Nature. 521(7553):415-16.
20. Liao SM. (2020) (Ed.) Ethics of artificial intelligence. Oxford: Oxford University Press.
21. Pereira LA, Oliveira CR. (2020) The General Data Protection Law and its influence on the use of artificial

intelligence in accounting. J Tec Soc. 16(2):45-56.

22. Silva AP, Coelho AZ, Feferbaum M, Silveira Acrd. (2023) Ethics, Governance and Artificial Intelligence. São Paulo: Almedina Brazil.