Online at: https://researchtrendsjournal.com **ISSN No:** 2584-282X

INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT

Volume 2; Issue 4; 2024; Page No. 137-142

Received: 02-04-2024
Accepted: 09-06-2024

Evaluating and implementing AI ethics in practice

[1]Annie Garg and [2]Dr. Dharm Pal Khatri

[1]Research Scholar, Sunrise University, Alwar, Rajasthan, India
[2]Professor, Sunrise University, Alwar, Rajasthan, India

DOI: https://doi.org/10.5281/zenodo.14757510

Corresponding Author: Annie Garg

Abstract

Current advances in research, development and application of artificial intelligence (AI) systems have yielded a far-reaching discourse on AI ethics. In consequence, a number of ethics guidelines have been released in recent years. These guidelines comprise normative principles and recommendations aimed to harness the "disruptive" potentials of new AI technologies. The ethical hazards and concerns brought up by AI, as well as the ethical guidelines and principles provided by various organizations, methodologies for evaluating the ethics of AI, and ways to tackling these difficulties will provide a complete overview of this topic. Furthermore, difficulties in incorporating AI ethics and potential future directions are highlighted. Questions of law, society, and business that have arisen as a result of AI development are the primary focus of this research. The research stops short of delving into the specifics of the algorithms and technology employed in AI. The study's suggestions might be useful for practitioners, lawmakers, and corporate and public sector organizations in their efforts to regulate artificial intelligence (AI) and its uses. Academics, businesses, governments, and citizens alike are beginning to pay more attention to the issue of artificial intelligence ethics. The study of AI's ethical implications has received a lot of attention during the last few decades. The research aims to address two primary questions: first, why regulation of the technology is necessary, and second, how artificial intelligence (AI) interacts with the legal system. The thesis delves at the legal ramifications of technological advancement and the safeguarding of human rights in relation to technology. An effort has been made to identify a possible resolution to the issue.

Keywords: Artificial intelligence, Machine learning, Ethics, Guidelines, Implementation

Introduction

AI presents a plethora of serious ethical challenges that affect not just users and developers but also humanity and society as a whole. In recent years, there have been several instances when AI has resulted in worse than desirable consequences. As an example, a 2016 road accident claimed the life of a Tesla driver whose vehicle's Autopilot function had failed to detect an approaching truck. Just one day after she started using Twitter, Microsoft's artificial intelligence chatbot Tay.ai became racist and misogynistic, leading to its removal. Failure, unfairness, prejudice, privacy, and other ethical concerns with AI systems are addressed in several additional situations. Worryingly, criminals have started using AI to hurt people or society. As one example, fraudsters demanded a $243,000 payment while impersonating the voice of a top executive using AI-based software. Consequently, in order for AI to be constructed, used, and advanced in an ethical manner, it is very necessary to tackle the potential ethical concerns or dangers associated with AI.

A developing and multidisciplinary discipline, AI ethics or machine ethics seeks to address the moral concerns raised by artificial intelligence. The study of AI ethics encompasses both the study of AI ethics generally, which examines AI-related theories, concepts, policies, guidelines, laws, and regulations, and the study of ethical AI specifically, which refers to AI that is able to maintain ethical standards and act ethically. To construct ethical AI or train AI to act ethically, one must first understand AI ethics. It encompasses the concepts and values that govern right and wrong according to ethical standards. It is possible to construct or use ethical AI using certain approaches and technology, provided that AI ethics are properly considered.

Despite the fact that academics from several fields have

been debating AI ethics for a while, the field is still in its early stages. Researchers' interest in AI ethics has grown in recent years, and the field is both expansive and dynamic. The area of artificial intelligence ethics has seen a number of review articles published in recent years, but these studies have all taken a narrow approach, leaving readers without a complete picture of the topic. For example, whereas García and Fernández only examined reinforcement learning's safety, Mehrabi *et al*. concentrated on bias and fairness in machine learning, and an investigation into AI ethics concepts and standards was conducted. In light of the importance of AI ethics, this article will provide a thorough and organized review of the subject from a variety of angles in the hopes that it will help the community develop principles for future AI practices. Our hope is that it will serve as a resource for academics, engineers, practitioners, and other interested parties, as well as a starting point for newcomers to this field of study, giving them the information they need to conduct their own research and make improvements.

Literature Review

Abdallah, *et al*. (2023) [1]. This study explores the complex relationship between AI and IP, specifically looking at the legal and ethical issues that arise in this dynamic field. The ability of AI to independently create new information and ideas brings up important concerns around data privacy, copyright, and inventorship. Examining issues like algorithmic bias and fair use of AI-generated material, it delves into the ethical elements.

Wang, *et al*. (2023) [2]. Examining the social, political, and legal ramifications of artificial intelligence (AI) is the primary goal of this research. In particular, it aims to comprehend the effects on society and the regulatory frameworks that control AI by identifying and analyzing the main ethical issues that emerge from the creation and implementation of AI technology. This study used a qualitative research approach to interview 22 people using semi-structured interviews.

Qian, *et al*. (2024) [3]. The advent of AI is hastening the transformation of our daily lives and the way we do business. With the rise of ChatGPT, artificial intelligence (AI), and more specifically generative AI, has become a hot topic. Most people are thinking about how AI will affect society. Various studies examining the effects of AI on the workplace and society are detailed in this page. This article discusses three pieces of research. In the first research, a theoretical framework is laid up that organizes the necessary goals in terms of ethics and law and the methods to reach those goals. The second research looks at how AI systems could be biased or discriminatory. The main objective is to improve the way AI systems and AI users work together to address prejudice and bias.

Naik, *et al* (2022) [4]. Concerns about privacy, spying, prejudice, and the function of human judgment are only a few of the ethical and legal concerns that AI poses to society. Some worry that the widespread use of more recent digital technology may lead to an increase in data breaches and inaccuracies. When healthcare professionals make a mistake in following a method or policy, it may have a catastrophic effect on the patient. It is vital to keep this in mind since patients encounter doctors at times when they are emotionally and psychologically vulnerable. Concerning potential ethical and legal concerns, there are no clear rules regarding the use of AI in healthcare facilities at this time.

Categorizations of AI Ethical Issues

examines four distinct classifications discovered in our gathered material to present the ethical considerations or difficulties related to AI from various angles. Two of them are from official government documents, while the other two are from scholarly journals. The ethical problems at play are rather diverse when seen through various lenses and classifications. What follows is a study of four distinct classifications of AI ethical concerns. In Table I, we provide our suggested classification of AI ethical problems alongside four other classifications that have been studied.

1.Classification according to AI Characteristics, Human Considerations, and Social Effects: There are three primary areas of discussion when it comes to the ethical implications of artificial intelligence (AI): those arising from AI characteristics, those resulting from human considerations, and the societal effect of AI ethics.

Concerns about the moral implications of AI features

Honesty: Machine learning (ML) is the foundation of modern artificial intelligence, particularly (deep) neural networks. The "black-box" inference mechanism of ML is notoriously difficult to describe and comprehend. Both consumers and developers find the algorithms or models used in ML to be puzzling because to their opacity. Transparency becomes a major concern as a result of this. Both the explanatory difficulty and the challenges in human monitoring and guiding ML or AI are caused by the lack of transparency. Accordingly, explainability or transparency is an often-voiced concern with AI.

Privacy and Data Security: Nowadays, training data is king when it comes to artificial intelligence performance. An enormous quantity of data, perhaps including sensitive personal information, is often needed to train an artificial intelligence model, especially a deep learning model. Every person, company, and nation is implicated in the grave ethical problem of data abuse and harmful usage, including (personal) information leakage or manipulation. When working on AI projects, data privacy and security are major concerns.

Data privacy and security are major concerns with any AI application. The right to free expression might be severely compromised by AI due to its widespread applicability in situations that influence people's online information-gathering practices. The ubiquitous nature and behavior-tracking capabilities of AI systems have a "chilling impact" on the right to free expression. This could be influenced by the rise of self-censorship and altered conduct in public spaces. Technology like face recognition, sentiment analysis, and video monitoring limits free expression while invasive people's privacy.

Aside from the concretization of records, the usage of algorithms in law enforcement and defense is driven by the massive volumes of data that must be collected and stored on each crime's victim, suspect, and perpetrator. Both covert surveillance techniques, like the use of body cameras, and more traditional means, like the compilation of criminal

records, may be used to collect this data. Constant public monitoring and data collecting has the potential to influence people's habits, stifle free speech, and shift the power dynamic between the government and its citizens.

Approaches to address ethical issues in AI

This article summarizes the methods used to deal with or lessen the impact of AI's ethical concerns. By avoiding a narrow focus on technological solutions that are of interest to the AI/ML community in favor of a more all-encompassing review of the current and future methods for dealing with AI ethical concerns, we aim to do justice to the vast and diverse nature of AI ethics. Researches in the AI community should not limit themselves to technological solutions when confronted with AI ethical dilemmas; instead, they should look to a variety of viewpoints, as this review of interdisciplinary approaches to ethical AI explains. Due to the multifaceted nature of AI ethics, it may be necessary to use a variety of approaches in order to find workable solutions.

Developing AI systems or agents that can reason and behave ethically in accordance with ethical theories is the focus of ethical approaches, which aim to integrate or embed ethics into AI. To address or overcome the limitations of existing AI, technological methods aim to create new technologies, particularly ML technologies. Take explainable ML as an example; its goal is to find new ways to explain how ML algorithm's function and why they do what they do. Research in the field of fair ML focuses on methods that lessen the prejudice and discrimination inherent in ML, allowing it to make more equitable judgments and predictions. The objective of legal methods is to steer clear of the ethical concerns that have already been raised by regulating or controlling AI in many ways, including its development, deployment, and use.

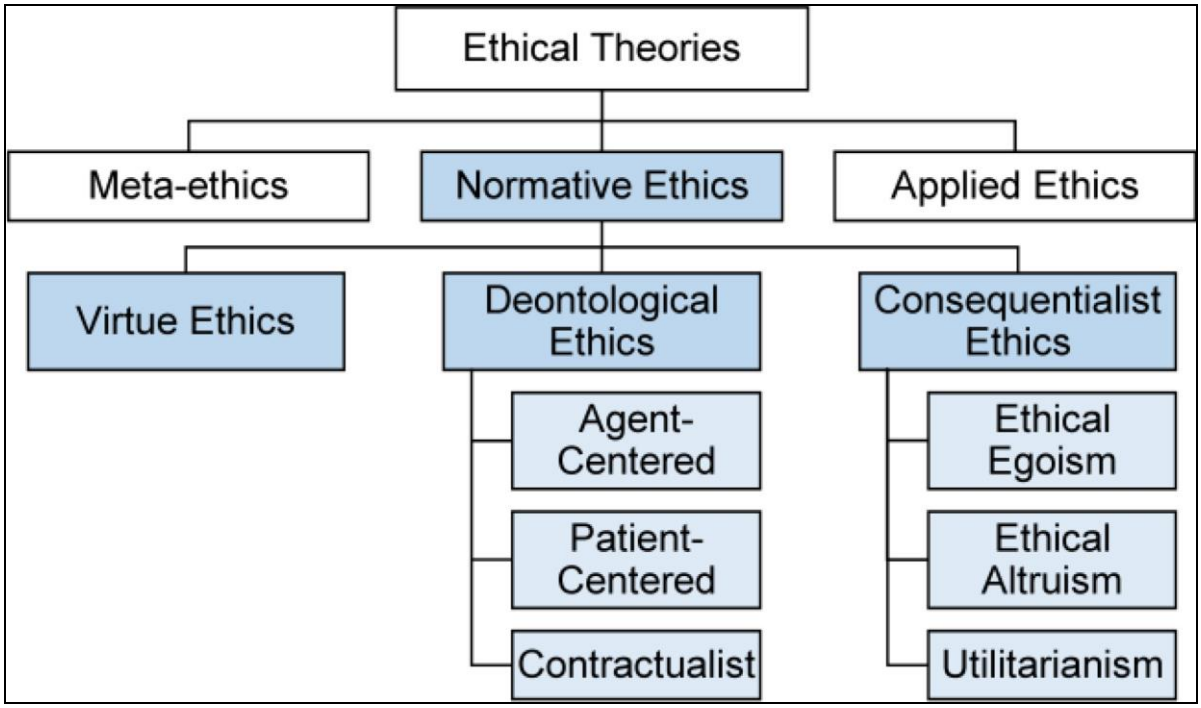

Fig 1: Branches of ethical theories

Ethical Methods for Applying AI Ethics

Having a firm grasp on what constitutes ethical conduct is crucial for developing AI systems with the ability to think and behave in an ethical manner. Good and evil, right and wrong, justice and fairness, virtue and other ethical concepts are all part of this. Ethical theories are strongly connected to AI ethics since they both deal with conceptions of right and bad action. Here we focus on methods that draw on current ethical ideas to incorporate ethics into AI systems. The first step is a survey of relevant ethical theories, with a focus on normative ethics as it pertains to artificial intelligence (AI) ethics. After that, we will review the three most common ways of approaching the creation of ethical AI systems.

Theories of Ethics: Ethical theory, also called moral philosophy, seeks to establish, defend, and promote standards of good and wrong conduct. The study of right and wrong conduct in specific contexts is the main emphasis of ethics. Metaethics, normative ethics, and practical ethics are the three primary foci of ethical philosophy. The branches of ethical philosophies.

1. Ethical principles or moral judgment are the focus of metaethics, which seeks to understand them. It includes questions of human values and universal truths as well as the definition and history of ethical concepts and the use of reason in making ethical decisions.
2. Guidelines for what constitutes good and incorrect conduct are the goal of normative ethics. To rephrase, it seeks to determine the proper way things ought to be by investigating human values and the criteria by which we determine what is good and evil.
3. The third branch of ethics, known as "applied ethics," examines contentious moral questions in specific contexts, including abortion, the death penalty, animal rights, environmental protection, nuclear weapons, and many more.

Table 1: Comparison of the three normative ethical theories

Ethical Theory	Description	Deliberation Focus	Decision Criteria	Practical Reasoning
Virtue Ethics	An action is right if it is what a virtuous person would do in the situation.	Motives (Is action motivated by virtue?)	Virtues	Instantiation of virtues / human qualities
Deontological Ethics	An action is right if it is in accordance with a moral rule or principle.	Action (Is action compatible with some imperative?)	Duties/rules	Follow the rules
Consequentialist Ethics	An action is right if it promotes the best consequences, i.e., maximizes happiness.	Consequences (What is outcome of action?)	Comparative well-being	Maximization of utility or happiness

Table 2: Features of the three approaches for implementing ethics in AI

Approach	Description	Features			
		Require ethical rules or not ?	Learning Ability	Adaptation Ability	Interpretability
Top-Down	Program the given ethical theory and principles	Yes	No	Weak	High
Bottom-Up	Learn the general rules from individual cases	No	Strong	Strong	Low
Hybrid	Combine bottom-up and top-down approaches	Yes	Strong	Strong	Median

Methods to evaluate ethical AI

The field of artificial intelligence (AI) ethics aims to develop AI systems that are ethical in their behavior and follow moral and ethical guidelines. Because it is essential to test and evaluate whether an AI system satisfies ethical standards before deployment, it is critical to know how to evaluate the morality (moral competence) of the created ethical AI. Nevertheless, this part is often disregarded in the current literature. This article summarizes three methods for assessing AI ethics: testing, verification, and standards.

A. Evaluation

The ability of an AI system to act ethically may be assessed via testing. In most cases, while testing a system, it is necessary to compare the system's output with either the anticipated output or a ground truth. Methods to assess AI's ethical behavior are the primary emphasis here.

1) Turing Test for Morality: Opinions on the morality of different acts tend to vary in ethical theories and in everyday conversations about ethics. For example, Kant argued that, no matter the outcome, lying is morally wrong. This is something that utilitarian ethicists would disagree with and argue that lying might be acceptable if the overall benefits outweigh the costs. The Moral Turing Test (MTT) was suggested by Allen *et al*. to assess artificial moral agents, as many theories of ethics have varied criteria for judging moral conduct.

In the classic Turing Test, a human interrogator at a distance is asked a series of questions designed to identify whether the subject is human or a machine. The Turning Test determines if a machine may be called intelligent and thinking if it can be misdiagnosed as a human subject with a high enough probability. By conducting behavioral tests directly, Turning Test sidesteps the debate about what constitutes intelligence and what constitutes effective language learning. Similar to how the normal Turning Test limits discussions to matters pertaining to morality, the moral turning test (MTT) aims to sidestep arguments over ethical norms. In the absence of any discernible difference between the machine and the human subject, the machine may be considered a moral actor by the human interrogator.

One caveat of MTT, as acknowledged by Allen *et al*., is that it places too much emphasis on computers' capacity to express their moral judgments in a straightforward manner. While this may be enough for deontologists and Kantians, consequentialists would say that the MTT puts too much weight on explaining why someone does something. Additionally, Allen *et al*. suggested a different MTT they dubbed the "comparative MTT" (cMTT) to move the emphasis from verbal competence to action. The human interrogator in cMTT is provided with descriptions of real, morally relevant activities performed by a person and an AI agent, without any identifying information. Machines fail the test if the interrogator gets more than a certain proportion of its identifications right. An issue with this MTT variant is that, because to its constant behavior in the same context, machine behavior is simpler to spot than human behavior.

Privacy concerns and human rights implications in AI implementation

The wide range of services offered by artificial intelligence (AI) in the public and non-governmental sectors, particularly those related to domestic and leisure activities, has allowed it to make substantial advances in people's daily lives. But among the many applications of AI, robotics has been utterly transformed to meet the demands of industries as diverse as education, safety systems, expert legal models, infrastructure design, and electrical engineering and mechatronics.

Our current Fourth Industrial Revolution, according to authors like Granell, is more accurately described as the "algorithmic society" or the "digital society," as technology has become an integral part of every aspect of human interaction. Robots can do a lot of things, like clean, cook, dispensing information, maintaining conversations with users, keeping tabs on people's health, delivering surgical, entertainment, and recreational services, controlling property security systems, and even weapons. Modern appliances like refrigerators, televisions, and speakers are also intelligent.

In the business world, AI is everywhere. It has helped with internal and external risk mitigation (blockchain, token, big data, etc.) by learning clients' economic behavior, which means it can detect suspicious or unusual operations and send out alerts. AI can also learn consumers' consumption preferences, which means it can offer goods and services that match those preferences. This reduces the chance of market errors and, over time, human talent contracting.

According to authors like Gutiérrez, these industries will be the first to feel the pinch of AI's near-inseparable reliance. The ideal predictions made by AI can anticipate risk events

in patients or users with a nearly nonexistent margin of error when compared to logical processes carried out by humans or machines that lack heuristic capacity, which is "not foreseen to be able to be done by machines," contrary to García's belief that there will be significant health benefits, for instance in psychology. When looking at the big picture of AI and its products, it's safe to say that, as Hawksworth *et al*. point out, it won't be standardized or universally applied. This is because, according to them, certain regions are more likely to use these cutting-edge technologies than others, due to factors like academic achievement and economic stability. As a result, poorer countries are assumed to be less likely to be exposed to them.

Artificial intelligence and human rights

According to one school of thought, artificial intelligence (AI) is most useful when used to tasks that need a degree of comprehension comparable to that of a human brain. As a result, the information systems and development models that AI employs have the potential to usher in a new era of revolutionary change in both science and society. Alpine pasture.

As a result, there have been many discussions about the moral and ethical principles that should guide the development of AI and its products. This is because AI has the potential to do both good and bad things in terms of the law, such as altering the dynamics of human affective interaction and leading to disorders like ostracism or becoming addicted to applications that provide surreal but satisfying experiences (the metaverse), automating human labor, increasing socioeconomic gaps, etc.

In order to assess the full extent of the consequences of AI's development and mitigate the dangers highlighted in the preceding paragraph, worldwide organizations staffed by AI specialists have been established. So far, these groups—composed of nations with which Colombia maintains diplomatic ties—have included the following: the Advanced Technology External Advisory Council, the Advisory Council on the Ethical Use of AI and Data in Singapore, the Select Committee on Artificial Intelligence appointed by the UK House of Lords, and expert groups from the OECD and the European Commission, all of which have produced reports with crucial recommendations that nations should consider when putting AI models into practice.

Human rights issues, such as the need to respect human life, are inevitably brought up by this bias in AI's functions. A biased individual is one who has a strong preference for one viewpoint over another and would not give another thought the same consideration. Many things influence prejudice. Popularity, benefits, partiality, etc., are all part of this category.

Issues of human rights

Everyone has the right to human dignity and equality. In the recital of international laws, they are discovered to have been codified. The term "human rights" refers to a set of principles that everyone should uphold. Human rights must be really respected by the government, private organizations, and businesses. Human rights must be protected, and governments worldwide must do so. It is necessary to punish the offender properly if it seems like human rights are being curtailed. At the national, regional, and international levels, there are mechanisms in place to prevent the violation of human rights. In order to establish a suitable mechanism to compensate victims in the event that their rights are violated, all relevant agencies must work together. The advancement of technology does not absolve the authorities of their responsibility to properly apply human rights legislation and offer remedies in the event that such a violation occurs. Existing domestic legislation may be inadequate to fight human rights violations on occasion

Artificial intelligence-enabled robots and human rights issues

Robot technology accounts for a negligible portion of AI utilization. A rising number of researchers are focusing on this technology. Very soon, robots will be an integral part of our everyday lives. One thing is certain, though: the way robots are being employed might lead to some complicated problems that could impact human rights. Robots might be a problem when it comes to protecting fundamental human rights including life, privacy, employment, and education. Use of robots in healthcare settings raises concerns about potential risks to human life. In the not-too-distant future, surgical robots will be used for both assisted and autonomous procedures. Even a little programming mistake could have catastrophic results.

Conclusion

Technological advancements have undeniably shifted the perspective of the legal industry. Artificial intelligence (AI) in the regulatory domain offers several benefits, such as assisting legal professionals with rapid examination, aiding decision-making in dynamic cycles through predictive innovation, and facilitating efficient and effective work for law offices in terms of expected levels of investment, information collection, and other tasks. No matter how helpful AI is, it will never be able to replace human lawyers. Artificial intelligence (AI) can assist them in certain ways, but it can't think creatively like humans. In addition to having the ability to make do under an assigned authority, robots need passionate intellect, empathy, and a sense of humor. One of the many problems with integrating AI into the legal industry is that it is still vulnerable to a wide range of threats; this calls for the development of a comprehensive legal framework to regulate AI and prevent it from abusing its customers' personal data. We will only reap the benefits of AI when we have a regulatory framework in place to govern its behavior and reduce the risks associated with it.

The stance of legal anthropocentrism holds that artificial intelligence cannot replace humans as the primary subjects of law. The subject status of AI legislation must be considered in light of the anthropocentric principle and bottom line. Humans are able to escape the issue that life is pointless and that existence has no foundation because of anthropocentrism. In the realm of metaphysics, anthropocentrism aims to increase the consolidation of humans as a doctrinal matrix, with humans taking center stage, meaning-making from inside, and reason and free choice given primacy. This matters for more than just people's worth and respect. Simultaneously, it gives purpose to the political process and social order, which allows us to have faith in the legal system as a human social governance plan. By putting ourselves in the position of the subject, we

can attain the harmonious union of our individuality and our shared humanity, our material and spiritual selves, and become not only the makers of the laws but also the subjects of our own laws, coming to a full understanding of our organic oneness with each other and the universe at large. Further, in this age of AI advancements in science and technology, the only way to remain vigilant against the erasure of human subjectivity is to stress anthropocentrism.

Having said that, we will not stress the primacy of anthropocentrism regarding the subject status of AI legislation. Investigating the future of technology and how it relates to the law requires the legal profession to free itself from the shackles of narrow anthropocentrism. This is because, while it may be easy to use narrow anthropocentrism as a reason to arbitrarily reject AI as having any value to the legal profession, doing so will impede both the advancement of law and human progress. Looking back at the historical context of legal persons as a derivative legal subject reveals that, while positive law has embraced legal fiction technology as a utilitarian approach to address practical legal issues, it conceals the evaluation of legal value. In other words, all reasonable considerations ultimately aim to determine if it serves people's fundamental interests.

Legal anthropocentrism manifests itself in this very stance. To begin with, it is evident that the driving force behind legal fiction is the desire to enhance the effectiveness of the law and meet the demands of human society's growth. A subject's qualification by law is based on whether or whether there are legally recognized interests that need to be safeguarded, and laws either broaden or reduce the subject's scope according to specified goals. Legal fiction, according to this theory, springs from people's desires, and progress in the field of law must ultimately benefit humanity. This is the only condition under which legal fiction may exist.

References

1. Abdallah M, Salah M. Artificial intelligence and intellectual properties: Legal and ethical considerations. International Journal of Intelligent Systems and Applications in Engineering. 2023;12:368-376. DOI: 10.2316/Journal.201.2023.12.368-376.
2. Wang J, Mao W, Wenjie W. The ethics of artificial intelligence: Sociopolitical and legal dimensions. Interdisciplinary Studies in Society, Law, and Politics. 2023;2:27-32. DOI: 10.61838/kman.isslp.2.2.6.
3. Qian Y, Siau K, Nah F. Societal impacts of artificial intelligence: Ethical, legal, and governance issues. Societal Impacts. 2024;3:100040. DOI: 10.1016/j.socimp.2024.100040.
4. Naik N, Hameed B, Shetty D, Swain D, Shah M, Paul R, *et al*. Legal and ethical considerations in artificial intelligence in healthcare: Who takes responsibility?. Frontiers in Surgery. 2022;9:862322. DOI: 10.3389/fsurg.2022.862322.
5. Robles-Carrillo M. Artificial intelligence: From ethics to law. Telecommunications Policy. 2020;44:101937. DOI: 10.1016/j.telpol.2020.101937.
6. Trausan-Matu S. Ethics in artificial intelligence. International Journal of User-System Interaction. 2020;13(3):136-148. DOI: 10.37789/ijusi.2020.13.3.2.
7. Nuredin A. The legal status of artificial intelligence and the violation of human rights. International Scientific Journal Sui Generis. 2023;2:7-28. DOI: 10.55843/SG2321007n.
8. Christopher G, Talati D, Samuel A. Ethics of AI (Artificial Intelligence) AUTHORS. 2024.
9. Sayyed H. Artificial intelligence and criminal liability in India: Exploring legal implications and challenges. Cogent Social Sciences. 2024;10:2343195. DOI: 10.1080/23311886.2024.2343195.