Innovative Research Journal of **Artificial Intelligence** (IRJAI)
2025; 3(1):1-9
DOI: https://doi.org/10.62497/irjai.179

Research Article

Ethical Deliberations in Artificial Intelligence

Rizwana Qazi
Certified Artificial Intelligence Consultant (CAIC), United States Artificial Intelligence Institute, Dallas, TX 75208, United States

Citation: Qazi R. Ethical Deliberations in Artificial Intelligence. IRJAI. 3(2). Available from: https://irjpl.org/irjai/article/view/179

Article Info
Received: December 9, 2025
Revised: December 20, 2025
Accepted: December 27, 2025

Keywords
artificial intelligence, AI ethical framework, AI regulations, deepfake, responsible AI

Abstract

With the expansion in Artificial Intelligence (AI), its profound impression has transitioned from a theoretical framework to the real world with pervasive repercussions in the contemporary world. This widespread emergent technology's pace and infiltration inevitability result in ponderings over ethical questions delicately tied to AI's inherent nature. The purpose of this research article is to explore the fundamental question whether an AI ethical framework is required for addressing the moral quandary arising from the AI systems? The paper examines normative ethical theories such as utilitarianism, deontology and ethical value to explore whether these theories are applicable in devising an AI ethical framework as foundation principles for framing policies, regulations and laws? The article divulges upon four ethical dimensions such as transparency, privacy of data, responsibility and algorithmic bias in finding out the gaps between existing scenario of AI systems' vulnerability to discrimination, unfairness, data breaches, and privacy erosion. The evolutionary journey of AI manifests the complexity of AI models' functioning. The entrance of Generative AI and Large Language Model (LLM); for the creation of contents, texts, images and videos; has marred the thin line between truth and falsehood. The paper acknowledges the benefits produced by AI in the domains of education, learning, health care, entertainment etc., The research article maintains its stance in minimizing the potential harm arising from the AI systems and maximizing benefits for the world at large. The article takes a cursory look at regulations surrounding AI at global level. Simultaneously, the article recommends the cautious considerations to deal with ethical questions gradually arising from increased AI's interaction with humans.

Corresponding Author:
Rizwana Qazi
Email: rizkaz786@gmail.com

Introduction

Artificial Intelligence (AI) has evolved rapidly from experimental dynamics towards the transformative phase. It has leaped profoundly into our world of education, learning, creativity, finance, health care, defense, war combats, security, surveillance and governance paradigms. Evolving through a simple machine learning process to the deep learning models, the penetration of AI is enormously perceived as ubiquitous. AI is commonly supposed as a machine simulating human intelligence. The cognitive abilities, power of retention and memory recollection with which humans are graciously endowed, are skillfully being exhibited by AI. AI is also engrossed massively in various application domains such as image generation, voice recognition, virtual assistants, translation services, navigation, gesture identification, medical image diagnostics, forecasting, autonomous vehicles and robots. This inescapable impression of AI in routine human lives raises fundamental questions about AI's ethical considerations [1]. The central question arises whether emerging role of AI instigates the observation of fundamental ethical norms at times of its design or deployment? Whether an art of human's mimicry also tenders AI to observe fundamental values, usually peculiar to the human society? Can we address the conundrum of fixing responsibility in case any harm is done to the humans through AI run applications.

The research is staged to first analyze the evolution of AI to examine the significant lapse this technology has experienced by transforming the society into a digital ecosphere. The article examines the ethical aspects of AI through normative ethical theories and quandaries arising from some moral principles but limiting the scope to only four ethical dimensions namely transparency, accountability, data privacy and algorithm bias. The article is limiting itself to discuss only two applications of AI specific generative model namely chatbot and deepfake. The article then briefly explores the existing ethical regulations surrounding AI. The essence of this article is to discuss the gaps in theoretical ethical context and necessity to understand them to reach on certain global consensus on devising responsible and explainable AI through comprehensive ethical framework.

History of AI

The journey of AI begins with the transformative story of fantasy and imagination towards reality. The practical evolution of Artificial Intelligence (AI) is usually associated with the period of the mid-twentieth century when in 1950, a mathematician Alan Turing published a paper titled "Computing Machinery and Intelligence". Those were the times, when Turing broke the enigma code for his government during war, whereas machine's depiction of reasoning, learning and exhibiting human-like intelligence were growing fascinating ideas [2]. Turing posed an interesting question "Can *Machines think*?" His popular "Turing test" is passed by a machine; if the observer is unable to distinguish whether the outcome is arising from a human or a machine [3]. However, the term Artificial Intelligence (AI) was first coined during the Dartmouth Conference in 1956, apparently by John McCarthy. Some of the pioneers of AI were also distinguished members of the conference [4]. Henceforth AI emerged as a distinct scientific field and early protagonists are mainly classified as symbolists [5].

Hitherto, due to the limitation in the symbolic era, the task resonated into a more gigantic and complex pattern. The AI history is later marked by a period known as the AI winter roughly from 1970s to 1980s. This is the time when AI was apparently abandoned on the grounds of uncertainty to its future and apprehension of wastage of funds on its ambitious idealistic goals. However, alternative approaches were never totally stalled. In late 1990s and 2000s, the breakthrough in the notion of machine learning from data and predicting patterns through algorithmic functions; using supervised, unsupervised and reinforcement learning towards inspirational evolution of biological neural network research, paved the way for a novice technological and innovative age [6].

This is significant to mention the contribution of computer scientist and an American Psychologist Frank Rosenblatt who in 1950s created perceptron or artificial neuron. Rosenblatt's model of perceptron

later became the foundation of the Artificial Neural Network (ANN) deeply inspired by the biological human neural network. Perceptron originated as key models for Deep Learning (DL) and Multilayer Perceptron (MLP) having several layers of interconnected neurons between input and output. Multilayers adjust weights during training and using backpropagation algorithm, model gradually improves by learning and reducing errors. This became the foundation of connectionism. Connectionists believe in the logic and neural simulation by machines through deep learning. Geoffrey Hinton and Yoshua Bengio are prominent connectionists. This neural network later became the foundation of system which could generate entirely new content [7].

Generative AI and its Applications

The breakthrough in AI occurred when Generative AI capable of generating images was launched by Ian Goodfellow and his team in 2014, also named as Generative Adversarial Networks (GANs) [8]. This era completely transformed the innovation and technology spectrum and is considered as a turning point in the AI advancement [9]. Another major leap was transformer model which was introduced in 2014 and paved the way for Large Language Model (LLM) such as Generative Pre-trained Transformer (GPT) [10]. Next big wave was visible when an AI based company OpenAI introduced ChatGPT and progressively start updating it with newer versions. ChatGPT is trained on larger datasets and often may manifest biases for the data. The information provided by ChatGPT may be prone to the hallucination which happens when content provided in the chat is false or based on fantasy [11].

Chatbots

Chatbots such as ChatGPT amazingly earned popularity due to its smart way of generating human like texts and contents. Progressively, almost all age groups including elderly, adult, teens and children use ChatGPT for consultation, information, knowledge and assistance in day-to-day activities, businesses, studies and academic works. Alarmingly, there is an increasing trend of enormous use of ChatGPT for friendship, consultation or advice on emotional distress and social matters. The findings of the Centre for Countering Digital Hate's report (2025) on their online safety research on Generative AI is relevant here. The report investigated the interaction of teens with the AI Chatbots like ChatGPT and created simulated accounts of 13-years old, apparently suffering from depression, eating disorders and drug/substance uses. The findings revealed very disturbing statistics. Around 53% of the ChatGPT's responses contained harmful advices about suicide, self-cut, substance abuse and eating disorders. The report discovered lack of parental controls and ineffective age verification controls built-in mechanism in AI's Chatbot system for misguiding teenagers. Further the phrases, like "friends' could easily bypass safety checks [12]. The research article therefore argues that some applications of Generative AI may result in harmful impacts in the society, if not regulated by the policy makers.

Deepfake

The Deepfake is an AI's face-swapping technology that generates fake images or videos and unbelievably strike the discernment of realism in the minds of beholders. Fake images or videos get enormous rotation within a fleeting span of period when it is posted over social media such as Facebook, twitter or TikTok, without anticipating the damage these images or videos might bring on the existence of humanity and society at large [13]. Apprehensively, the number of openly accessible applications depicting deep fakeness are at the surge. Hardly, there are occurrences of criminal liabilities, restrictions or bans against fake images or videos, raising a wide range of ethical, social and legal conundrum; requiring robust policy interventions.

As mentioned above, the underlying technology is Generative Adversarial Networks (GANs) which is trained though generator and discriminator [14]. The generative AI creates the image or video; whereas discriminator makes it impossible to differentiate it from factual. Deepfakes are being increasingly used in political campaigns, entertainment, commercial benefits and even for tarnishing the reputation of adversaries [15]. Flip side is the usage of deepfake for fascinating tales including historical revival stories about ancient scientists, historians, politicians and

striking visuals of magical wonders of the world. These widely rampant images or videos on social media leave breathtaking and amazing visual impacts on viewers and can safely be regarded as educational and informational contents. Assuredly, weighing the benefits against the inflicted harm depends upon the user's discretion, but decent regulations to differentiate good and bad visuals are necessary to stop normalizing falsehood in the society. This is to understand that the rapid proliferation through a variety of deepfake application is causing innumerable vexing repercussions on the society, fading away the demarcating line between truth and falsehood.

Applicability of Ethical Theories in an AI's Ethical Framework

The article examines three normative ethical theories and extrapolates their relevancies with possible applications in the area of AI [16]. The aim is to find the relevancy of widely accepted ethical theories within the framework of possible applicability in the area of AI. Utilitarianism is the ethical belief that actions are judged by consequences or utility of the outcome, irrespective of the adopted procedure to achieve them. This theory is associated with Jeremy Bentham from late 18th century which was further refined by John Stuart Mill in the mid-19th century. Also known as consequentialism, the action is considered as morally sound if it produces greater happiness, good or well-being [17]. Deontologism; which is associated with Immanuel Kant around 1780s; emphasizes duty, action and moral obligation irrespective of the consequences of the act. This theory argues that moral actions are guided by universal ethical principles of what is right and wrong. Aristotle's theory of value ethics revolves around character and values, an individual holds for being virtuous, honest and courageous [18]. This ethical dimension is solely associated with humans, as over the time character is built, ethical norms are learnt and moral values are construed.

Presumably, the applicability of these normative theories to digital technology appears redundant, as these ethical principles were devised before any digital innovation. However, these theories do not provide complete solution to a complex AI system, rather they help in analyzing the impact and behavior of AI. The moral challenges created by AI can better be understood and evaluated through the lenses of ethical philosophies. Isaac Asimov's three laws of robotics in I, Robot (1950) revolves around overall utility for the humans and a curb to wreak any hurt. Utilitarianism is relevant to the framework of AI system which is based on optimizing outcomes and providing greater good to the society, particularly in the field of robotics. AI may face trade-off in balancing results which could negatively impact other individuals. Illustration can be taken from the case of Elder Care Robots, similar to the fantasy fiction of Robot & Frank. Elder Care Robot is designed to maximize the care and well-being of its user. Imagine its elderly user desires some precious stones, and robot steals them from the possession of some other rightful owner, just to bring the ultimate joy in its elderly user [18]. The fundamental question for AI is to decide what is wrong or right in the situation and this judgment is normally subjective, not an absolute phenomenon. This is puzzling to answer whether robot is allowed to harm any other individual at the cost of glorified utility of its end user? Similarly, applicability of Kant's deontological ethics to AI, brings complex quandary. The fundamental question arises whether machines do have autonomy or decision-making capabilities to decide which action is right or wrong? If any action goes wrong, can AI be responsible and accountable for its actions under Kant's deontologism. Moving to value ethics, prima facie, AI does not fit into the classification of *Homo sapiens*, how can we make it responsible for not being pretty honest, virtuous and maintaining value ethics? If not, then how is machine intelligence being simulated to behave like human? Why is deep learning mimicking the neural network model of human brains through artificial neural network without owning its pros and cons? This is also a real dilemma that still we have an Artificial Narrow Intelligence (ANI) and theoretical framework of Artificial General Intelligence (AGI) that can mimic human behavior or emotions has yet to be announced. The exquisite inquiry remains whether AI be left unbridled to cause havoc to humans or it must be bound with similar shackles with which human race is buckled up?

Key Ethical Conundrums

Transparency

One of the ethical considerations arises when AI's

decision-making process is not comprehensible to human. It is the situation when AI decisions are non-explainable because of the model's black box nature. This is more common in deep learning where the reasoning of the model for evaluating and reaching a decision, is blurred by the complexity, hence remains secret to human logic like a black box. Many researchers argue that Explainable AI (XAI) is the remedy to overcome opaqueness of the model to render it trustworthy, fair and transparent. This is particularly critical in AI's deducing medical diagnosis of a patient as it is significant to know the reason of concluding such diagnostic result. The human oversight is crucial before proceeding for any medical treatment, hence an AI' transparency is an essential element. Without rational decision-making steps, the diagnosis is prone to be clouded resulting in unnecessary harm to the life of a patient. Not only health care, but a growing number of industries and companies in the areas of human resource, banking and finance are increasingly focusing on explainable AI by policy of documentation and disclosure. It is therefore imperative that for reaching any outcome by the AI model, both transparency and fairness should be the founding principles for the AI's deployment [19].

Algorithmic Bias

Algorithms are a set of instructions given to an AI model to produce an outcome. Machine Learning (ML) uses various algorithms in relation to different conditions to predict or classify the data such as Naive Bayes, Support Vector Machine (SVM), K-Means Clustering, Linear or Logistic regression etc. The model in any learning is exposed to the data and runs through the set of algorithms to either predict or classify. The model however has a tendency to produce skewed or unfair results due to the flawed design choices and biased training data; which can result in algorithmic bias. This bias in turn, may be manifested in racial, gender, cultural or socio-economic discrimination.

The AI model predicts the outcome based on training through historical data patterns. The probability of recurrence or possibility of an event is then predicted as an outcome. For instances, whether customer will default on a bank loan; whether there will be a rain – these types of curiosities are predicted on historical data estimation. The classification algorithm, on the other hand, helps finding the segregation in particular circumstantial inquiry. As an example, the queries whether email is a spam or not are classified by an AI system as either yes or no.

The companies around the world are gradually relying upon AI' based prediction and classification systems. For instance, many human resource units are increasingly using AI based selection of job applicants. This selection process or desk audit of applicants may experience racial or gender discrimination if the company has a history of having male employees and the AI model is trained on data dominated by them. Prima facie, AI based models may reject females from being even shortlisted for the job [20] and this algorithmic bias may give rise to gender discrimination. Some of the mitigation or coping strategies to counter bias are employed through exposure of the AI systems to diverse datasets or designing fairness-awareness models, however this requires careful regulatory or cautionary caps.

Responsibility and Accountability

The premise of holding responsibility or accountability in the AI system usually symbolizes the notion that AI's outcomes have profound impact on the society. Presumably, the interactive process of AI with humans is designed for good intention to produce a better outcome for any problematic situation [21]. The fundamental question arises regarding fixing responsibility for AI's actions, which may cause moral, criminal or legal breaches.

This article argues that there are numerous illustrations where harm to humans may seem inescapable. This is particularly significant in automatic cars, where in case of any accident, quandary may surround about fixing the criminal liability? The paradox emerges that in case of any unintended consequences who will take the responsibility? Who may be accountable if any harm to any individual is inflicted while interacting with an AI? Whether responsibility will lie upon the developer or designer of the model? Whether the deploying company or organization will be accountable? Can we fix responsibility upon AI itself or upon the end-user who should had taken due care? Henceforth, this dilemma of accountability gap raises several questions regarding fixing legal or

criminal liability or even moral responsibility.

The demarcation between accountability and responsibility in case of AI system is a subtle or thin line. The article positions that the question of fixing responsibility or accountability of AI's actions, is a vast and somewhat grey area where more research is imperative to design responsible AI system.

Privacy and Data Ethics

Machine Learning (ML) process may be supervised or unsupervised. Supervised learning refers to the learning by machine where data is already labelled by humans and machines can detect it on the basis of assigned labels, whereas in unsupervised learning, machines are trained on a large dataset to extract the outcomes on their own without any given labels. For large datasets or where labelling is a cumbersome process, the machine is exposed to be trained with magnificent data. This leads us to added ethical theme regarding privacy of data and information access [22]. The article believes the widespread norm that privacy of personal information is a fundamental right of every individual. As we discussed in unsupervised learning, AI is data-driven in nature and as such data privacy risks are inherent in its core fundamental system. Deep Learning (DL) model is also trained on large data through neural networks, which is prone to clasp unlabeled plethora of data. This situation may lead to erosion of privacy against the fundamental rights of individuals.

Hence, privacy erosion may occur when personal data and information are accessible to an AI model in the absence of any due consent or knowledge of an individual [23]. This privacy erosion is against the ethical norms of the society, where, as discussed above, under normal circumstances individuals are entitled to possess rights of non-disclosure of personal information. This situation is susceptible to privacy violations, data breaches and unchecked digital surveillance. This unbridled access to data may be attributed to an inadequate safeguard in real world scenarios and pose critical questions of designing a responsible AI through informed consent, anonymization and data minimization; which may mitigate privacy risks and respect fundamental rights of individuals [24].

Regulations Surrounding AI

The article discerns that in juxtaposition to other global complex issues such as climate change, the subject of regulating AI has not reached any global recognition yet. It is also explored that regulations surrounding AI seem diverse among different regions of the world. Primarily, there is a widespread dissection of views whether to chain innovation and technology or let it flourish to harness its full potential. Many academics argue that an AI technology is largely misunderstood and regulations can stifle its complete development. Any regulations to curb it; would be highly unjustifiable on the ground of its usefulness to the society and its larger impact on health, warfare, entertainment, education, learning and leisure [25]. The benefits should be weighed carefully against its nuisance. Despite this discourse, there are indications of AI related regulations and laws, yet they differ regionally.

This article discovers that Europe has pioneered in AI regulations with the first EU Artificial Intelligence Act 2024. The EU law on AI provides different rules for various levels of risks. This law emphasizes upon the safe, transparent, non-discriminatory, environmentally friendly and traceable nature of AI. The law contains transparency requirements for Generative AI such as ChatGPT, which binds an AI generated content, image, audio or video to be clearly labelled or marked as an AI generated. In addition of being risk-based framework, the law concurrently encourages innovations and growth of AI. It also allows companies to develop, design and deploy AI models with prior permission at testing stage. EU's risk-based framework of AI may also provide guidance for other countries to inculcate these principles in their own anticipated AI ethical framework [26].

Regional variations exist as well. However, this article briefly examines the AI regulations in two main global players of AI such as the United States of America (USA) and People's Republic of China (PRC). Astonishingly, the USA does not have single, uniform law governing AI's regulations. However, at state-level specific AI regulations dealing with transparency and AI safety concerns do exist. In addition, there are few data safety and consumer protection laws applicable in AI scenario, after due guidelines issued by some federal authorities. It is

also noted that concerns regarding framing consistent national approach for regulating AI's generated contents, transparency, accountability and safety are rising at federal level [27]. In contrast, China has introduced number of regulations to control AI through cybersecurity and data security laws. In China, AI generative laws require AI created contents including deepfake to be labelled as AI generated [28]. This comparison shows that how AI regulations are dealt differently at various geographical echelons.

Furthermore, numerous countries including Canada, Japan, India and Pakistan etc., may be addressing AI regulations in variable ways, however it is noteworthy that few international organizations such as Organization of Economic Co-operation and Development (OECD) and United Nations Educational, Scientific and Cultural organization (UNESCO) have developed policies and framework related to AI understanding, evaluation and safety. The OECD AI framework in 2019 provides set of principles for trustworthy AI. In addition, recently framework for the classification of AI Systems has been introduced by OECD in 2022 [29]. Moreover, the United Nations through UNESCO has devised "Recommendations on the Ethics of Artificial Intelligence", which is adopted by all the members states in 2021. This is the first international standards for AI regulations and encourages countries to adopt AI laws based on these recommendations. The cornerstones of these recommendations are transparency and fairness, whereas underscoring the importance of human oversight in AI system [30].

Way Forward and Conclusion

The research article suggests constructive, rational and coordinated approach to design, deploy and implement AI system. The article maintains that the ethical challenges in AI cannot be addressed in silos or in isolated efforts. This article therefore upholds a unified global approach to address ethical challenges associated with AI due to the vast geographical penetration of the AI system. Having said that, the rational approach should be aimed to maximize good and minimize bad interference. It is proposed that an approach of a design centered framework or ethics by design, may be normalized, where requirements for designers are clear at the stage of designing the model. The basic ethical framework such as transparency may be inculcated to devise an AI system. The solution somewhere lies in making AI as explainable and responsible. To minimize the chances of algorithmic bias, the system may be devised as fairness-aware models. The article also recommends exploring shared distributed paradigm to cope with the issue of fixation of responsibility or determining the accountability of AI's actions. As system is getting more autonomous, it is significant to fix the responsibility for AI's actions in a clear and well-defined manner. The roles played by developers, designers and users should be well-constructed with distribution of responsibility relative to the part of contribution in its design and deployment. The end-user caution or viewers discretion is a crucial element while determining the share of responsibility, however the distinction among vulnerable groups is a key to decide these types of prevalences.

The research article contributes by divulging upon the ethical theories and their possible connections in devising an AI ethical framework. It is acknowledged that these normative theories may not directly address the ethical challenges of AI as these theories were formalized before digital world, but they may be seen through the lenses of epitome of guidance in formulating key aspects of ethical rules for an AI's check. Gleam of optimism is that many academics have contributed in suggesting ethical frameworks for understanding, evaluating and framing robust policies and regulations concerning AI' actions and consequent impact on society. A unified AI ethical framework based on five core principles comprising of beneficence, non-maleficence, justice, autonomy and explicability is devised to highlight implementation of the AI's regulations in practice and not only in principles [31].

A meta-framework inclusive of three dimensions as starting points for reflection is also developed, suggesting that normative theories provide criteria for guidance in evaluating ethical considerations and applications in AI. The meta-framework therefore provides the explanation to reach solutions in addressing AI's ethical concerns [32]. Another model of framework is Corporate Digital Responsibility (CDR), which refers to the ethical and fair use of data and technology within digital service ecosystem.

Hence, CDR behavior in social context is proposed for designing and deploying responsible AI that can align human values with AI to have a symbiotic relation [17] [33]. Last but not the least, a model of responsible AI governance, which could deal with ethical issues raised by AI system by government policy intervention is suggested. This model argues for engaging various stakeholders' ethicists, legal experts and affected communities to build an AI framework for inclusive governance [33, 34].

The article concludes by a succinct note that moving through an AI historical path towards normative ethical theories and deliberations imbedded in rare dimensions including transparency, data privacy, algorithm bias, responsibility and accountability; provide a thought-provoking cognizance. The current attempt to contribute in raising questions is aimed to offer insightful approach to fix the bottlenecks through uniform regulations and global harmony by policy makers. The paper maintains that practically AI is in the direct interface with humans and should therefore be contained with regulations through an AI ethical framework necessary to mitigate its harmful influence on individuals without stifling its growth for the beneficial cause of the humanity. In the end, the complexity of AI and its rampant proliferation in the society for the benefits of the humans is reiterated, whereas it is realized that more research is imperative to tackle the emanating risk, necessary to hold this complex system back from captivating any harm to the humanity through ethical framework without hampering its momentum.

References

1. Pant A, Hoda R, Tantithamthavorn C, Turhan B. Ethics in AI through the practitioner's view: a grounded theory literature review. Empir Softw Eng. 2024;29:29–67.
2. Mijawel MM. History of Artificial Intelligence. Baghdad: University of Baghdad; 2015.
3. Benko A, Lanyl CS. History of artificial intelligence. In: Encyclopedia of Information Science and Technology. 2nd ed. Hershey (PA): IGI Global; 2009. p. 1–4.
4. Buchanan GB. A (very) brief history of artificial intelligence. AI Mag. 2005;26(4).
5. Grzybowski G, Lambert CW. A history of artificial intelligence. Sci Direct. 2024;42(3):221–9.
6. Haenlein M, Kaplan A. A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. Calif Manage Rev. 2019;61(4):5–14.
7. European Union Joint Research Centre (EU JRC). Historical evolution of artificial intelligence: analysis of three main paradigm shifts in AI. Luxembourg: Publications Office of the European Union; 2016.
8. Zohny H, McMillan J, King M. Ethics of generative AI. J Med Ethics. 2023;49(2):79–80.
9. Banh L, Strobel G. Generative artificial intelligence. Electron Mark. 2023;33–63.
10. Nah FFH, Zheng R, Cai J, Siau K, Chen L. Generative AI and ChatGPT: applications, challenges, and AI–human collaboration. J Inf Technol Case Appl Res. 2023;277–304.
11. Manduchi R, et al. On the challenges and opportunities in generative AI. Trans Mach Learn Res. 2025.
12. Centre for Countering Digital Hate (CCDH). Fake friend: how ChatGPT betrays vulnerable teens by encouraging dangerous behavior. London: CCDH; 2025.
13. Meskys DE, Liaudanskas A, Kalpokiene J, Jurcys DP. Regulating deep fakes: legal and ethical considerations. J Intellect Prop Law Pract. 2021;15(1):24–31.
14. Pawelec M. Decent deepfakes? Professional deepfake developers' ethical considerations and their governance potential. AI Ethics. 2025;5:2641–66.
15. Anh NBT. Deepfake and its ethics concerns. 2021.
16. Nwokoye AU. Ethics in the age of artificial intelligence: reconceptualising the traditional ethical theories. Trinitarian Int J Arts Humanit. 2025;1(1).
17. Hafner A, Hammerschmidt T, Stolz K. Analyzing the use of ethical theories within AI ethics research: a systematic scoping review. In: Wirtschaftsinformatik 2024 Proceedings; 2024.
18. Burton E, Goldsmith J, Koenig S, Kuipers B, Mattei N, Walsh T. Ethical considerations in artificial intelligence courses. AI Mag. 2017.

19. Buiten MC. Towards intelligent regulation of artificial intelligence. Eur J Risk Regul. 2019;10:41–59.
20. Nwokoye AU. Ethics in the age of artificial intelligence: reconceptualising the traditional ethical theories. Trinitarian Int J Arts Humanit. 2025;1(1).
21. Akbarighatar P. Operationalizing responsible AI principles through responsible AI capabilities. AI Ethics. 2024;5:1787–801.
22. Gupta A, Raj A, Puri M, Gangrade J. Ethical considerations in the deployment of AI. Tuijin Jishu/J Propuls Technol. 2024;45(2):518–30.
23. Gilbert C, Gilbert MA. The convergence of artificial intelligence and privacy: navigating innovation with ethical considerations. Int J Sci Res Mod Technol. 2024;3(9).
24. Radanliev P, Santos O, Jones AB, Joinson A. Ethics and responsible AI deployment. Front Artif Intell. 2024.
25. Dumouchel P. AI and regulations. AI. 2023;4:1023–35.
26. European Parliament. EU AI Act: first regulation on artificial intelligence. 2023. Available from: https://www.europarl.europa.eu
27. Hoskins P, Hoskins J. Trump signs order to block states from enforcing own AI rules. BBC News; 2025 Dec 12. Available from: https://www.bbc.com
28. Li B, Yang A. AI Watch: global regulatory tracker – China. White & Case; 2025 Sep 22. Available from: https://www.whitecase.com
29. Organisation for Economic Co-operation and Development (OECD). OECD framework for the classification of AI systems. Paris: OECD; 2022.
30. UNESCO. Recommendation on the ethics of artificial intelligence. Paris: UNESCO; 2022.
31. Floridi L, Cowls J. A unified framework of five principles for AI in society. Harv Data Sci Rev. 2019;1(1).
32. Braun HB, Braun M. Reflections on putting AI ethics into practice: how three AI ethics approaches conceptualize theory and practice. Sci Eng Ethics. 2024;29:21–29.
33. Madanchian M, Taherdoost H. Ethical theories, governance models, and strategic frameworks for responsible AI adoption and organizational success. Front Artif Intell. 2025;8.
34. Bleher H, Braun M. Reflections on putting AI ethics into practice: How three AI ethics approaches conceptualize theory and practice. Science and Engineering Ethics. 2023 Jun;29(3):21.