

## Ethical Challenges and Policy Approaches in Artificial Intelligence

Sandesh Lamichhane<sup>1</sup>, Usman Aijaz N<sup>2</sup>, Tulika Suman<sup>3</sup>, MuniBhovi Pavan Kalyan<sup>4</sup>, Mustafa Anwar Hussain Khan<sup>5</sup>

<sup>1,4,5</sup> PG – Artificial Intelligence and Machine Learning with Minor in Data Science, Yenepoya University, Bangalore, Karnataka, India.

<sup>2</sup>Associate professor, Dept. of Computer Science and Information Technology, Yenepoya University, Bangalore, Karnataka, India.

<sup>3</sup>Assistant Professor, Computer Science and Information Technology, Yenepoya University, Bangalore, Karnataka, India.

**Email ID:** [san20lamichhane@gmail.com](mailto:san20lamichhane@gmail.com)<sup>1</sup>, [usmanaijaz.blr@yenepoya.edu.in](mailto:usmanaijaz.blr@yenepoya.edu.in)<sup>2</sup>, [tulikasuman20@gmail.com](mailto:tulikasuman20@gmail.com)<sup>3</sup>, [39197@yenepoya.edu.in](mailto:39197@yenepoya.edu.in)<sup>4</sup>, [mustafakhann00786@gmail.com](mailto:mustafakhann00786@gmail.com)<sup>5</sup>

### Abstract

Artificial intelligence (AI) is transforming various fields at a rapid rate. This phenomenon has raised major ethical issues, leading to the establishment of regulations by governments, organizations, and educational institutions. In spite of the increasing rate of AI-related activities, there have been contradictions in the concepts and regulations. There have been convergences in the fundamental concepts of AI ethics, including beneficence, non-maleficence, autonomy, justice, and explicability. Empirical studies have shown that there is a good understanding of AI ethics. However, the application of the numerous mitigating techniques that have been proposed in the literature remains low. Mixed-method studies of international regulations have shown major deficiencies in the regulations, especially from the Global South and underrepresented groups. Evaluations of regulations in different sectors, including healthcare, education, and governance, have shown that there are insufficient instruments and low levels of stakeholder engagement in the execution of AI ethics. New frameworks like EMMA and Ethics by Design have shown promising strategies for the integration of ethical monitoring in the life cycle of AI development. There is a major gap between the ethics of AI and the regulations that can practically be implemented.

**Keywords:** Artificial Intelligence, Ethics, Fairness, Governance, Policy, Sustainability, Transparency

### 1. Introduction

The fast-growing application of artificial intelligence (AI) in areas ranging from healthcare, education, finance, and the government has led to considerable debate on ethics and policy. With the growing application of AI in decision-making that was exclusively within the domain of human beings, the question of fairness, transparency, accountability, and social justice has come to the forefront. In this regard, many institutions and governments have proposed ethics guidelines to ensure the responsible development and application of AI technology. Although there are many guidelines on ethics, there is a certain lack of clarity and uniformity on the interpretation and application of ethics principles

from one industry and culture to another. Floridi and Cowls proposed beneficence, non-maleficence, autonomy, justice, and explicability as a possible foundation of common principles to establish AI ethics globally [1]. Critics argue that many frameworks are merely aspirational and lacking in the ability to be implemented in practical ways [2]. Furthermore, ethical issues vary from one sector of application to another. Artificial intelligence has the potential for use in the diagnosis and treatment of patients, while at the same time posing challenges with regard to bias and autonomy [3]. Likewise, the application of artificial intelligence software in the education sector has also led to a question about what

ethical standards need to be defined for children and whether it is necessary to lay guidelines for them [4]. With the ever-growing faith in the transformative power of artificial intelligence, there is an emerging need for the development of ethical, inclusive, and contextually sensitive frameworks that can be effectively operationalized from principles. The purpose of this study is to synthesize key findings from the literature on ethical AI frameworks, identify commonalities and gaps, and provide a way forward towards more practicable and universally applicable ethical AI frameworks.

## 2. Literature Review

This discourse in AI ethics has moved from abstract philosophical inquiry to more concrete and policy-based discourse. In this regard, earlier works in this area may have identified some of the core principles that govern AI ethics, such as fairness, accountability, and transparency; however, more recent works suggest that these principles have varying degrees of importance depending on the sector and/or culture in which they are being promoted and/or implemented. In this regard, Huang et al.'s [5] study that traces the development in AI ethics suggests that there is substantial support for high-level principles and a lack of globally consistent operational norms in law and/or industry. Some recent works have also highlighted that Western ethics have dominated the development and formulation of global standards in AI ethics and that other worldviews have been omitted from this discourse. In this regard, Roche [6] suggests that Euro-American values dominate AI public policy discourse and that other worldviews, such as Ubuntu in Africa and Islamic ethics, have been omitted from this discourse. This is in line with Stahl and Eke's [7] caution against the ethnocentrism of technologies such as ChatGPT that reflect and produce socio-technical biases that are based on Western values and principles. Sector-specific analyses have also identified some of the issues in current ethical standards in AI ethics. In this regard, Adams et al.'s [8] study identified ethical issues in AI in K-12 education that relate to issues in child data privacy and that current ethical principles are

inadequate to accommodate youth-specific issues and concerns. Similarly, Jeyaraman et al.'s [9] study highlighted ethical issues in healthcare that require delineating issues such as algorithm bias and accountability in healthcare and in critical care AI systems development. Governance and regulation of these challenges have shown considerable diversity. In a qualitative and quantitative synthesis of global AI ethics policy, Saheb and Saheb [10] have reported considerable variations in the approaches and levels of development of different countries. They have stated that, while in some aspects the approaches align with an idealized vision of a feminist perspective on the ethical development of AI, there is a considerable divergence in the overall regulation and ethics of the field. This divergence is even more profound at a deeper level, a fact that is supported by Stix [11] in his discussion of the different domains of AI ethics, including public regulation, with distinct boundaries. The interrelation of AI with big data and surveillance technologies is a major area of ethical concern. In their discussion of the use of AI in the COVID-19 pandemic, Anshari et al. [12] have reported a critical need for the regulation of the ethical aspects of the use of AI in the context of big data and surveillance. In a similar vein, the discussion of the ethics of the use of AI in the health sector by Karimian et al. [13] has focused on the need for greater clarity in the allocation of responsibility. Beyond these tenets, the new methods seek to connect other design paradigms with ethical theory. "Ethics by design" (EbD) is a methodology put forth by Brey and Dainow. They argue that the translation of abstract values into concrete technical tasks needs to happen as part of the development processes [14]. This is different from conventional ways of thinking about ethical theory and its application. What their contribution gives us is a clue to how ethics might be integrated with the life cycle of artificial intelligence (AI). In current discourse, issues of power and responsibility have remained ubiquitous. Stahl et al.'s contribution is to propose an "ecosystem approach" to AI ethics, which goes beyond the formulation of individual principles by exploring issues such as

corporate power, failures of regulation, and social inequality [15]. A key component of this is "human flourishing" as a measure by which AI success is to

be judged. Thus, AI success is not to be judged simply against a pre-defined norm, but against its effects on "human flourishing" as well.

**Table1 Comparison of Major AI Ethical Frameworks**

| Framework                     | Origin        | Key Principles                                                                    | Strengths                             | Weaknesses                              |
|-------------------------------|---------------|-----------------------------------------------------------------------------------|---------------------------------------|-----------------------------------------|
| Floridi & Cowls (2019)        | EU (Academic) | Beneficence, Non-maleficence, Autonomy, Justice, Explicability                    | Bioethics-based, widely cited         | Western-centric, abstract               |
| OECD AI Principles            | OECD          | Inclusive growth, Human-centered values, Transparency, Robustness, Accountability | Policy adoption by governments        | Non-binding, lacks enforcement          |
| AI4People                     | EU Commission | Same five principles as Floridi & Cowls                                           | Policy-aligned ethical roadmap        | Difficult to operationalize             |
| IEEE Ethically Aligned Design | IEEE          | Human Rights, Well-being, Data Agency, Effectiveness, Transparency                | Technical depth, industry focus       | Idealistic, limited uptake              |
| China's AI Guidelines (III)   | China         | Harmony, Responsibility, Security, Prosperity                                     | Reflects local sociopolitical context | Low external scrutiny, vague principles |

### 3. Methodology

In the present study, an integrative methodology is employed to explore the ethical dilemmas and policies related to Artificial Intelligence (AI). The objective of the present study is to offer a comprehensive synthesis and analysis of the existing scholarly research, policies, and institutional reports

in the field of AI ethics and governance in an attempt to develop a cohesive perspective on the overall discourse of emerging ethics and governance in the field of AI. This non-empirical study is focused on clarifying concepts, evaluating arguments, and exploring ethical issues and initiatives. The present investigation is not bound by any time constraints,

and all the literature cited is relevant to the theme. The literature survey is carried out through various platforms like Scopus, Web of Science, Google Scholar, and institutional repository websites. The keywords used in the survey include, but are not limited to, “AI ethics,” “algorithmic accountability,” “ethical governance,” “AI policy,” “responsible AI,” and “AI regulation.” The scope of the survey includes the overall ethical dilemmas in the field of AI, including fairness, transparency, privacy, accountability, human rights, and social impact. The overall methodology of the present study is based on the concept of thematic synthesis. The relevant ethical principles, dilemmas, and governance mechanisms were identified and synthesized in the overall analysis of the literature. The overall ethical dilemmas and mechanisms were synthesized and categorized based on the major thematic concepts. The overall thematic concepts include the overall foundation of AI ethics, sector-specific ethical dilemmas in the field of AI, variations in the regulation of AI across the world, and the overall significance of interdisciplinary approaches in the field of AI. A secondary part of the overall methodology is based on the concept of evaluating the overall valuation of ethics in the context of different frameworks and their overall potential in the field of AI. This methodology does not involve testing hypotheses or conducting data collection in the traditional sense. Rather, the methodology incorporates a critical approach with interlocutors and secondary sources in order to reveal important knowledge gaps, theoretical discrepancies, and policy misalignments. It enables the creation of a well-informed narrative that is as diverse and intricate as necessary with regard to AI ethics. The findings will be used in order to increase understanding, as well as inform the development of future ethical and legal issues in the area of AI.

#### 4. Results

Of the ethical AI frameworks discussed, there seems to be consensus on the basic concepts of ethics. Floridi and Cowls [1] have identified 47 ethical principles in six important AI texts, which they

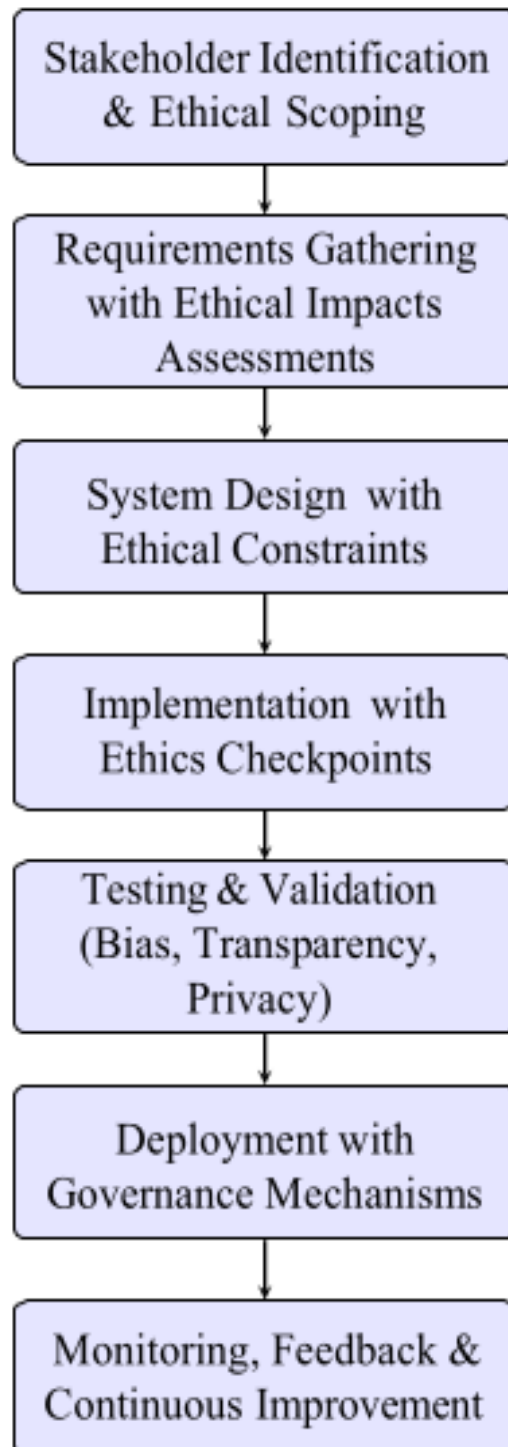
subsequently linked to four basic values in bioethics: beneficence, non-maleficence, autonomy, and justice. However, they argue that these measures are inadequate in dealing with the special issues raised by AI, including the lack of transparency and accountability in AI systems. In order to address these issues, they proposed a fifth criterion, “explicability,” which incorporates the concepts of intelligibility and responsibility. The strategy has achieved legitimacy in the form of important policy documents, including AI4People and the OECD, though this is limited in its acceptance in the Western world. In reality, there is a considerable “principle-practice gap” that surpasses theoretical projections. Stahl et al. [16] have found five “motifs” that firms apply in addressing AI ethics: “consciousness and reflection,” “technology mitigation,” “human oversight,” “ethical training,” and “trade-offs.” In many situations, while there is recognition of awareness and association with stakeholders, there are various ethical assessments where legal guidelines, such as GDPR, take precedence over system-oriented issues such as bias or even societal implications. Data anonymisation/encryption, for example, have surfaced as prominent themes, though these solutions only address surface issues through technological solutions, rather than addressing the ethical dilemmas. The awareness of ethics is responsive, and the implementation of ethics is arbitrary, highlighting the dichotomy between business and ethical considerations. Anshari et al. [17] conducted a study examining the ethical concerns of AI in social media and decision-making within the context of the COVID-19 pandemic. The analysis found data privacy, prejudice, and lack of permission as the most prominent ethical challenges. AI technologies, particularly those that analyse social data, frequently lack user understanding, resulting in violations of privacy and fairness. The research examined these difficulties through deontological and utilitarian lenses, emphasising the imperative of accountable data governance, informed consent procedures, and strategies to

mitigate algorithmic discrimination in both public and private AI applications. Saheb and Saheb [18] accomplished this at a macro policy level by compiling a total of 146 AI ethics standards globally. From their findings, five ethical criteria were identified as being discussed by the majority of the participant experts: transparency, fairness, accountability, non-maleficence, and privacy. The identified themes were then classified into the following categories: principles, personal data privacy, governance and oversight, processes, epistemology, and state action. Thus, it is safe to conclude that the study reveals a taxonomy of ethical considerations relevant to the public health context. To summarize, the study reveals a taxonomy of relevant ethical considerations for the context of public health. Although there is a global consensus on basic principles, there are also variations, especially with regard to Western and non-Western countries. This is a clear indication of the need for a more inclusive discourse on the ethics of AI in the near future. The findings of the diverse investigations highlight the areas of consensus and disagreement with regard to the ethics of AI, as well as the theoretical, organizational, and policy-related dimensions of the subject. This consensual advancement with regard to basic principles is noteworthy; however, it is also a fact that, over the past few years, there is a growing divergence between idealistic aspirations and reality. The investigation reveals recurring themes, especially with regard to the necessity for contextually relevant, enforceable, and inclusive ethics.

## 5. Discussion

Ethics in artificial intelligence should move beyond its principle-centric focus towards a dynamic, inclusive, and implementational ethic. This review paper aims to clarify that ethics is conducted on two levels, depending on how one chooses to define the framework: ethics is concerned with facilitating progress while maximizing AI's positive potentials,

such as those related to sustainable development or limiting its misuse, such as in autonomous weaponry or surveillance access. While these approaches, which have considerable impact on global policy agendas, to some extent reflect a Western-centric focus, failing to sufficiently integrate perspectives from AI-leader states such as China, or those of the Global South, a cross-cultural, pluralistic approach is required, according to Dignum et al., to avoid the risk of ethical imperialism in setting the global AI agenda [19]. Along with the development of principles, there is a parallel development of these principles in the field. The underlying social implications of this phenomenon are not sufficiently discussed and may result in a compliance-oriented corporate ethics, in which data privacy and risk reduction play a major role, while the overall social impact is neglected. This may lead to the following major tensions: (1) governance misalignment, in which technical solutions replace solutions within a broader framework of responsibility; (2) sectoral divide, in which the state or the corporation is focused on a particular ethical concern; and (3) temporal failures, in which the overall ethical risks of a decision are neglected in favor of compliance issues in the short term [20]. In the absence of any kind of incentive and legal mechanism, the phenomenon of “ethics washing” arises, in which a symbolic commitment to ethics is displayed instead of a real commitment to morality. The ethical issues have become even more pressing with the appearance of the COVID-19 crisis and the emergence of the related AI technologies in the field of medicine and public communication. The voluntary approaches were not immune to the challenges of the crisis and the related violation of ethical principles in favor of expediency. As discussed by Morley et al., trust and stakeholder education play a major role in the overall governance of AI technologies, in addition to the enforcement of norms [21].



**Figure 1 Ethics-by-Design: A Life-Cycle Approach for Embedding Ethics in AI Systems**

Activities such as audits, validation, and landscape assessments, alongside widespread, multi-stakeholder engagement, form the foundation of this domain. The ethical integration of artificial intelligence within society is not only concerned with the use of AI, but also with the process of creation, which is something that needs to accompany, and evolve with, AI progress within legal, technical, and organizational structures. There is a critical need to develop a global, multi-disciplinary understanding of ethics, which includes marginalized regions and populations, something that requires more involvement from low- and middle-income countries, alongside ethical processes that are geographically sensitive to sociopolitical realities, such as those emphasized by sambasivan et al. [22].

Furthermore, ethical AI governance needs to be considered as a 'living architecture,' one that not only expands and reconstitutes itself, but does so with regard to technological evolution, societal, cultural, and institutional developments. To achieve such a feat, there is a need for international collaboration, global policy, alongside research that is conducted through normative ethics, while recognizing the empirical effects of AI.

### Conclusion

The governance of ethics in artificial intelligence (AI) needs to move from foundational principles to operational, adaptive, and responsive frameworks that are enforceable and inclusive. While foundational principles are a good starting point for moral reasoning, the principles of beneficence, non-maleficence, autonomy, justice, and explicability are not sufficient. The results show that there are still gaps between the ethical standards and their implementation in practice in various fields, such as healthcare, education, and public administration. The gaps are manifested in the Western bias in the ethics debate in the global arena, the lack of enforcement of policies, and the reduction of ethics to a compliance checklist. Moreover, the lack of

access to shape AI policy leads to underrepresentation of marginalized groups and people in low- and middle-income countries, which raises questions about justice and parity. A multi-layered approach is recommended to deal with these challenges. The first layer is more difficult but relatively easier to implement, as it involves the direct incorporation of ethics into the AI development process using Ethics-by-Design and EMMA. The second layer recommends the use of interdisciplinary methods that bring together technologists, ethicists, policymakers, and stakeholders as a matter of course. The third layer demands that global and national bodies make inclusive policies that give voice to the Global South and re-evaluate ethical standards to better suit the realities of the field. AI governance needs to be considered as a "living architecture" that can adapt to technological progress, cultural differences, and social needs. This adaptability can be achieved through continuous research and dialogue, as well as the development of new legal frameworks, accountability, and training. Ethical AI governance needs to be considered as a dynamic and evolving concept. Architecture needs to adapt to incorporate new technologies, cultures, and social needs, requiring continuous research and dialogue in conjunction with legal frameworks to hold organizations accountable for their behavior and to enable ethical development.

### References

- [1]. L. Floridi and J. Cows, "A unified framework of five principles for AI in society," *Harvard Data Science Review*, vol. 1, no. 1, 2019. doi:10.1162/99608f92.8cd550d1.
- [2]. C. Stix, "Actionable principles for artificial intelligence policy: Three pathways," *Science and Engineering Ethics*, vol. 27, no. 1, 2021. doi:10.1007/s11948-020-00277-3.
- [3]. G. Karimian, E. Petelos, and S. M. A. A. Evers, "The ethical issues of the application of

artificial intelligence in healthcare: A systematic scoping review,” *AI and Ethics*, vol. 2, pp. 539–551, 2022. doi:10.1007/s43681-021-00131-7.

- [4]. C. Adams, P. Pente, G. Lerner, and G. Rockwell, “Ethical principles for artificial intelligence in K-12 education,” *Computers and Education: Artificial Intelligence*, vol. 4, p. 100131, 2023. doi:10.1016/j.caeai.2023.100131.
- [5]. J. Huang, A. S. Kesselheim, and C. T. Robertson, “Artificial intelligence and ethics: Sixteen challenges and opportunities,” *Nature Machine Intelligence*, vol. 3, no. 1, pp. 13–14, 2021.
- [6]. D. Roche, “Republic of codes: Codes of ethics and practices in AI,” *AI & Society*, vol. 36, pp. 531–542, 2021.
- [7]. B. C. Stahl and D. Eke, “Ethical aspects of artificial intelligence: An overview,” *AI and Ethics*, vol. 2, no. 1, pp. 1–10, 2021.
- [8]. R. Adams and L. M. Smith, “Artificial intelligence in K-12 education: Ethical concerns and policy implications,” *Educational Policy*, vol. 36, no. 4, pp. 573–594, 2022.
- [9]. M. Jeyaraman, “Artificial intelligence in critical care: Ethical considerations,” *BMJ Health Care Informatics*, vol. 27, no. 3, 2020.
- [10]. T. Saheb and M. Saheb, “Artificial intelligence and ethics in policy: A global perspective,” *Government Information Quarterly*, vol. 38, no. 2, 2021.
- [11]. C. Stix, “Three approaches to governing AI: Self-regulation, co-regulation, and public regulation,” *Science and Public Policy*, vol. 48, no. 2, pp. 262–271, 2021.
- [12]. M. Anshari and M. N. Almunawar, “Ethical use of AI and big data in health crises: COVID-19 case study,” *Health and Technology*, vol. 12, pp. 617–626, 2022.
- [13]. H. Karimian, “Artificial intelligence in healthcare: Ethical and cybersecurity challenges,” *Frontiers in Artificial Intelligence*, vol. 6, p. 112, 2023.
- [14]. P. Brey and B. Dainow, “Ethics by design: A conceptual framework,” *Ethics and Information Technology*, vol. 23, pp. 145–157, 2021.
- [15]. B. C. Stahl et al., “AI for human flourishing: An ecosystem approach to ethics and governance,” *AI & Society*, vol. 37, pp. 1–13, 2022.
- [16]. B. C. Stahl, J. Antoniou, M. Ryan, K. Macnish, and T. Jiya, “Organisational responses to the ethical issues of artificial intelligence,” *AI & Society*, vol. 37, no. 1, pp. 23–37, 2021. doi:10.1007/s00146-021-01148-6.
- [17]. M. Anshari, M. Hamdan, N. Ahmad, E. Ali, and H. Haidi, “COVID-19, artificial intelligence, ethical challenges and policy implications,” *AI & Society*, vol. 38, no. 3, pp. 707–720, 2023. doi:10.1007/s00146-022-01471-6.
- [18]. T. Saheb and T. Saheb, “Mapping ethical artificial intelligence policy landscape: A mixed method analysis,” *Science and Engineering Ethics*, vol. 30, no. 9, 2024. doi:10.1007/s11948-024-00472-6.
- [19]. V. Dignum, “Ethics in artificial intelligence: Introduction to the special issue,” *Ethics and Information Technology*, vol. 20, no. 1, pp. 1–3, 2018.
- [20]. B. D. Mittelstadt, “Principles alone cannot guarantee ethical AI,” *Nature Machine Intelligence*, vol. 1, no. 11, pp. 501–507, 2019.
- [21]. J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, “From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices,” *Science and Engineering Ethics*, vol. 26, no. 4, pp. 2141–2168, 2020.
- [22]. N. Sambasivan, N. Rangaswamy, E. Baumer, T. Doshi, and A. Nair, “Reimagining

algorithmic fairness in India and beyond,”  
Communications of the ACM, vol. 64, no. 4,  
pp. 40–43, 2021.