A Systematic Analysis and Taxonomy of Ethical AI Frameworks: Principles, Practices, and Future Directions

Harpreet Singh Dalla[1], Sonal Rattan[1,2*], and Gesu Thakur[3]

[1] AIT-CSE, Chandigarh University, Mohali, India
[2] Dept of UCRD, Chandigarh University, Mohali, India
[3] College of Smart Computing, COER University Roorkee, India
Harpreet.e13143@cumail.in, sonal.e15123@cumail.in, drgesuthakur@gmail.com
*Corresponding Author: sonal.e15123@cumail.in

Abstract. Artificial Intelligence (AI) has become an important technology in different sectors such as finance, healthcare, education, transportation and public administration. Various industries are achieving production efficiency, accuracy and innovation with AI's power of analyzing data, making predictions and taking automated decisions. Undoubtedly, AI has proved to be beneficial, however there are some ethical issues such as privacy risks, lack of transparency, bias, discrimination and accountability issues associated with it. Many governments, corporations and universities have developed various frameworks to ensure responsible AI design and use; however, these frameworks are either inconsistent and fragmented or difficult to implement. This research focuses on systematic review of various AI frameworks introduced by different organisations. Several AI frameworks from different sources have been analyzed before proposing a multi-level taxonomy to classify these frameworks based on stakeholders, lifecycle stages and principles. This paper also discusses challenges, trends and research gaps in accomplishing the successful implementation of AI ethics. The main goal is to help policymakers, researchers and practitioners to develop trustworthy AI frameworks which will increase transparency, governance and public interest in AI technologies.

Keywords: AI governance, ethical AI, Taxonomy, systematic review, AI Ethics Frameworks

1 Introduction

Artificial intelligence (AI) is when machines are designed to act like humans. The machines accomplish this by learning, thinking, solving problems, and understanding their surroundings. It covers many technologies that let computers do tasks that usually need human intelligence, like understanding different languages, studying information, or providing recommendations. Few examples of AI in day to day life include voice assistants, chatbots, and self-driving cars, where AI learns from data and uses it to make smart choices or take action.

A. Agnihotri et al. (eds.), *Proceedings of the Conference on Bridging Engineering Disciplines with AI and Machine Learning (BEDAIML 2026)*, Advances in Intelligent Systems Research 209,
https://doi.org/10.2991/978-94-6239-697-5_7

In recent years, artificial intelligence has gained a lot of attention in the technological world and it has created several opportunities worldwide allowing human connections through social media, facilitating healthcare diagnoses and providing operational efficiencies to different organizations through automated tasks. The benefits of AI are clear, however despite these benefits, it is important to focus on the ethical challenges associated with AI as it directly impacts both the individuals and the communities. Various AI technologies are directly impacting human lives, decisions and rights. Therefore, as AI advances further, it will become more important to ensure that it is developed and used in a responsible manner. In the absence of ethical guidelines, there are changes of unfair, harmful or even dangerous outcomes from various AI systems[1].

The increasing need to introduce ethical guidance in AI has led to the development of various AI frameworks that focus on responsible use of AI. These frameworks introduced by government and non-government organisations define values, ethical principles and government models. But these frameworks are extremely diverse in structure and difficult to implement. There is a lack of clear mechanism about how these frameworks can be enforced.

Previous research focussed on examining the ethical principles that constitutes AI, yet there is significant gap when it comes to comparing a variety of AI frameworks that exist on global scale. Without a common understanding, organisations are facing issues to align AI systems with ethical standards that can be universally accepted. Apart from this, the inconsistencies in these frameworks make the developers, practitioners and policy makers confused about what actually constitutes "ethical AI."

This paper focuses on solving the issues by systematically reviewing over 50 frameworks and developing a taxonomy that can help in classifying and analyzing these frameworks in an easy manner. These frameworks are taken from various corporate, government and academic sources to provide a detailed overview about how these frameworks are integrating ethics into AI governance.

The proposed taxonomy divides the selected frameworks into three dimensions that includes ethical principles, lifecycle stages and stakeholders roles. Considering these three dimensions, the objectives of this study are threefold:

1. To identify and analyze recurring principles and patterns across existing ethical AI frameworks.
2. To propose a structured taxonomy that captures differences in scope, purpose and practical implementation of these frameworks
3. To provide light on significant challenges, trends and research gaps in the successful practical implementations of AI ethics.

2 Background

2.1 Ethical AI

Ethical AI means developing, deploying and using AI technologies in a way that aligns with the key human values, legal obligations and social norms. The

main motive behind introducing AI ethics is to ensure that all the AI systems are designed, developed and implemented to provide benefits to humans while ensuring minimal harm. It focuses on key human values such as transparency, fairness, privacy and human rights.

Ethical AI is an old concept. It is based on previous ethical theories such as deontology, consequentialism and virtue ethics. Deontology means duty based ethics that focus on adhering fairness and justice. Consequentialism is an outcome based ethics that emphasizes on ensuring AI actions lead to fruitful results for humans which are equitable in nature. At last, the virtue ethics includes the character based ethics. Therefore, virtue ethics focus on the moral character of individuals involved in designing and developing AI. So, it guides the developers as well as users to advocate empathy, integrity and responsibility in development and implementation of AI technologies. These philosophical foundations act as moral compass that guides to balance technological progress with protection of human rights and technological innovation with equity. Despite the fact these theories are universally applicable, it is very difficult to apply them on real world AI systems. There is always a possibility of ethical trade-offs while applying these philosophical foundations[5].

2.2 Key Ethical Principles in AI

- Fairness and Non-discrimination: This principle ensures that all AI systems should treat all individuals equally. There should not be any discrimination based on age, religion, gender or socioeconomic status of a person. It is possible to achieve fairness in AI systems by identifying and getting rid of algorithm biasness via inclusive design and transparent data selection.
- Accountability and responsibility: It is important to have a clear mechanism to make sure all AI decisions and outcomes are being accomplished with assigned responsibility. In simple words, in any chance if any harm or error occurs due to any AI system, then human actors must take the responsibility which will help to prevent responsibility gaps.
- Privacy and Data Protection: AI decisions are usually based on vast datasets. Therefore, it is very important to protect the personal information of individuals. This ethical framework focuses on compliance with various data protection rules and regulations; for example, consent, advocate for data minimization, and secure storage.
- Human Autonomy and Oversight: It is important that human control remains intact. AI systems should augment human decision-making rather than completely replacing or unduly influence. Humans must be the final decision-maker, supervisor and can override these AI systems when necessary. This will ensure that the final accountability remains in the hands of humans, not these autonomous systems.
- Transparency and explainability: All the stakeholders must understand how AI models are built, trained and used. There should always be transparency about AI systems and all non-technical users must be explained about the AI systems and how it works.

- Safety and Reliability: this ethical framework ensures that AI systems must work in a consistent way. These systems should be consistently secure. The risks of adversarial attacks and malfunctions must be reduced. This can be accomplished through continuous testing, validation and monitoring of these systems.
- Society and Environment's well-being: the previous principles mainly focused on individuals. However, it is also important that AI should have an overall positive impact on society and the environment. AI systems should contribute to the progress of society and environment protections. Therefore, all the AI frameworks must consider implications of AI on inequality, employment and the ecological balance.

3 Literature Review

Boddington et al. (2024) performed a review on the rise of checkbox AI ethics[2]. The main focus was give on reviewing existing practical approaches to Ai ethics in the health industry. They searched for large databases and shortlisted the review studies to provide detailed analysis. After the analysis, it was found that many of the AI frameworks remained in declarative form rather than operational. These frameworks failed to demonstrate how to achieve AI governance in practical deployment.

Jobin et al. (2019) in their research conducted comparative analysis of different ethical AI guidelines[4]. The study focussed on reviewing 84 AI ethics documents. These documents were taken from different organisations across the globe. The research shows that transparency, privacy, responsibility, fairness and justice are the key recurring principles in these documents. The researchers emphasize the fact that there is global consensus on ethical AI principles, however noted a significant gap when it comes to practical implementation of AI ethics. Even the benchmarks that can help evaluate the AI compliance were also missing.

Floridi et al. (2020) conducted a deep review of "Ethics Guidelines for Trustworthy AI"by the European Commission[3]. They focussed on explaining how it is important to integrate ethical principles into different stages of the AI lifecycle. They emphasized the need to successfully transform the moral values into practical implementation while using these frameworks.

4 Methodology

In this research, a systematic literature review (SLR) is used to collect and analyze the information from different sources. The data is collected from various ethical frameworks that have been developed by different research institutions,governments and organizations. The main purpose of this research is to compare the frameworks and develop a taxonomy based on ethical principles, stakeholders roles and AI lifecycle stages. This systematic review focuses on papers published between 2015 and 2024 as during this period, a significant growth in AI ethics discussion was analyzed. Data is gathered from numerous academic

databases and documents from organizations. Academic resources used include papers published in IEEE, Google Scholar and policy documents by companies such as Google, IBM and Microsoft. Over 30 frameworks were selected with inclusion criteria focused on AI ethics. However, out of 30 frameworks, 10 frameworks are selected for final interpretations of results. On the other hand, any material that included informal reports or opinions were excluded from the research. Each framework is thoroughly reviewed to identify the key ethical principles in use, the AI lifecycle stages covered in it and the key stakeholders using that framework. All of these frameworks are analysed and grouped together to determine how they relate to one another. To ensure the research reliability, the sources are reviewed thoroughly for their credibility.

5 Result

After analysing the above table, it is evident that most of the AI frameworks focus on privacy, fairness, transparency and accountability. These ethical principles are considered as the foundation for developing responsible AI systems. EU, Google and Microsoft's AI frameworks particularly emphasise on these aspects. However, there are some organizations such as IEEE and UNESCO that expand the focus to sustainability, human rights and inclusiveness. These organizations show that there should be broader focus on social and humanitarian perspectives while considering ethics in AI. Another key finding is that the former organizations i.e. IEEE and UNESCO cover all stages of the AI lifecycle. It means that these organizations consider it important to implement throughout starting from planning to implementation and later on. On the other hand, China and UK AI Council's framework have limited focus on AI lifecycle stages limiting to design and deployment.

The third important aspect covered in the table is the stakeholder focus. Undoubtedly, a significant variety in scope can be observed in the EU, UK and the US's government-driven initiatives shows inclination towards public institutions and citizens. Therefore, these countries aim to safeguard society's interest while implementing technological advances. On the opposite sides, corporate companies such as Google and Microsoft primarily focus on industry leaders and developers. Another trend can be seen in research based and collaborative efforts by AI Now Institute and the Partnership on AI respectively. They promote cross-sector collaboration. Overall, it can be inferred that most of the frameworks align across similar ethical principles. But when it comes to stakeholder engagement and AI lifecycle coverage, there are significant differences.

6 Conclusion

In this paper, ethical AI frameworks by various governments, research organisations and corporate companies have been reviewed. The key finding of the research is that most of the frameworks share common grounds on key ethical principles important to be focused upon includes privacy, accountability,

Table 1. AI Framework Comparison

Framework	Ethical Principles Covered	AI Lifecycle Coverage Stage	Stakeholders
Microsoft Responsible AI	Transparency, fairness, privacy, accountability, fairness & inclusiveness	Data collection, design, development, model training, deployment	Developers, users, product teams
Google AI Principles	Privacy, accountability, fairness, social benefit	Design, development, deployment	Organization, developers, industry leaders
European Union High-Level Expert Group on AI Guidelines	Transparency, fairness, privacy, accountability, safety, human oversight	Design, development, deployment	Citizens, government, organizations
UK AI Council Principles	Transparency, fairness, privacy, accountability	Design and deployment stages	Government
UNESCO AI Recommendations	Accountability, inclusiveness, sustainability, human rights, transparency	All stages of AI lifecycle	Public institutions, government, researchers
AI Now Institute Framework	Labor rights, accountability, fairness, transparency	Development and implementation stages	Students, public, policymakers
White House AI Bill of Rights	Fairness, safety, transparency, privacy, accountability	Design, deployment, usage	Developers, citizens, government
China's Governance Principles for New Generation AI	Shared responsibility, fairness, transparency, harmony	Design, development, deployment	Corporations, government
IEEE Ethically Aligned Design	Beneficence, accountability, human rights, transparency	All stages of AI lifecycle	Policymakers, engineers, researchers
Partnership on AI Tenets	Collaboration, accountability, safety, fairness, transparency	Design, development, evaluation	Academia, non-profits, industry

fairness and transparency. The key difference between these frameworks is how these principles can be applied in real life. UNESCO and IEEE are two renowned organisations that cover all stages of the AI lifecycle while implementing ethics in AI. However, some organisations just focus on design or deployment stages. There is a large gap between ethical principles and their operational implmentation. Many of the discussed frameworks are broad in nature and difficult to make operational. The main reason behind it is the lack of methods, tools and standards to evaluate ethical performance.

The research findings also demonstrate that most of the frameworks are developed in the western countries. Therefore, these frameworks miss social and cultural values that are embraced by the rest of the parts of the world. To make sure that AI implementation is fair and ethical, there is a requirement for inclusive approaches and global cooperation. It is suggested that policymakers, governments and developers must work together to create universal standards for ethical AI and to make them practical.

References

1. Ahmad, S.F. et al.: Impact of artificial intelligence on human loss in decision making, laziness and safety in education. Humanities & social sciences communications. 10, 1, (2023).
2. Boddington, P., Herkert, J. R., & Williams, M. A. (2024). The Rise of Checkbox AI Ethics: A Review.
3. Floridi, L., Cowls, J., Beltrametti, M., et al. (2020). AI4People—An ethical framework.
4. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines.
5. Zhang, Y. et al.: Artificial Intelligence and Its Applications.