Exploring the Ethical Dimensions of Artificial Intelligence

Ms. M. Sri Soundharyaa, MCA., M.Phil., (Ph. D)[1], M Aswitha[2], A Abhilasha[3], S Dharani[4], S Sam Wesley[5]

[1]Assistant Professor, Department of Computer Science with Cyber SecuritySri Ramakrishna College of Arts and Science, Coimbatore

[2,3,4,5]Department of Computer Science with Cyber SecuritySri Ramakrishna College of Arts & Science, Coimbatore.

DOI: https://dx.doi.org/10.51244/IJRSI.2026.130200138

Received: 18 February 2026; Accepted: 23 February 2026; Published: 13 March 2026

ABSTRACT

Artificial intelligence (AI) has transformed many aspects of modern life and continues to do so at an accelerating pace. AI is being deployed across diverse sectors including autonomous driving, healthcare, media, finance, industrial robotics, and online services. While this broad integration into the economy and society has improved efficiency and generated substantial benefits, it has also altered social structures and raised significant ethical concerns. Issues such as privacy breaches, algorithmic discrimination, workforce displacement, and security risks associated with AI systems have become matters of public, organisational, and governmental concern. Accordingly, AI ethics—the study of the ethical issues surrounding AI—has emerged as a critical research area. This article provides a systematic overview of the field by summarising and analysing the ethical risks and issues raised by AI, reviewing ethical guidelines and principles published by major organisations, examining approaches to address these issues, and discussing methods to evaluate the ethics of AI systems. Additionally, the article discusses challenges in implementing ethics in AI and outlines future research perspectives. This work aims to offer a clear, accessible overview of AI ethics for researchers, practitioners, and newcomers to the field.

Indexterms: AI ethics, artificial intelligence, ethical issues, ethical principles, ethical guidelines.

INTRODUCTION

Artificial intelligence (AI) has developed rapidly and impressively over the past decade. Technologies such as machine learning (ML), natural language processing, and computer vision are expanding into virtually every domain of modern society. AI is increasingly undertaking tasks previously performed by humans, and its rapid growth is already transforming daily life and society in fundamental ways.

On the beneficial side, AI-powered chatbots can handle customer inquiries around the clock, improving satisfaction and increasing sales. AI enables doctors to treat patients in remote areas through telemedicine services. More broadly, the widespread deployment of AI has driven improvements in efficiency and reduced costs, benefiting economic growth, social development, and human well-being.

However, AI also presents significant ethical risks for users, developers, and society at large. In recent years, a number of high-profile incidents have illustrated these risks. In 2016, the driver of a Tesla vehicle was killed when its Autopilot system failed to detect an oncoming truck. Microsoft's AI chatbot, Tay.ai, was taken offline after it generated racist and sexist content within a day of being deployed on Twitter. In a financial crime case, criminals used AI software to impersonate a CEO's voice, resulting in a fraudulent transfer of $243,000. These and many similar examples illustrate failures, fairness problems, bias, privacy concerns, and ethical dilemmas that arise from AI systems. It is therefore essential to address the ethical concerns and risks associated with AI in order to ensure its responsible development and application.

AI ethics—also referred to as machine ethics—is an emerging field focused on identifying and addressing the ethical issues raised by AI. It involves studying ethical theories, guidelines, policies, principles, rules, and regulations related to AI, as well as establishing standards to ensure that AI systems operate in accordance with

moral values. By implementing sound ethical frameworks, we can develop AI that is designed and built to behave ethically.

Although interdisciplinary researchers have been discussing AI ethics for several years, the field is still in its early stages. It is broad and rapidly evolving, and it is attracting increasing attention from the research community. While a number of review articles have been published, they tend to focus on specific aspects of AI ethics, such as bias and fairness in machine learning, safety in reinforcement learning, or privacy and security in deep learning. There remains a need for comprehensive reviews that offer a holistic view of the field. This article addresses that gap by providing a systematic overview of AI ethics from multiple perspectives, offering guidance for the community to advance ethical AI practice. It is intended to inform scientists, researchers, engineers, practitioners, and newcomers to the discipline.

The main contributions of this article are as follows:

- We summarise the ethical issues and risks associated with AI and propose a new categorisation of AI ethical issues at the individual, societal, and environmental levels (Section III). This categorisation aids in recognising, understanding, and analysing ethical challenges in AI. We also map ethical issues to the stages of an AI system's lifecycle.

- Section IV presents a global overview of AI ethics guidelines and principles drawn from 146 guidelines released by various organisations and governments. These guidelines offer high-level direction for the planning, development, deployment, and use of AI.

- Section V reviews multidisciplinary approaches to solving AI ethical challenges, encompassing ethical, technological, and legal perspectives, and suggests concrete solutions beyond purely technical measures.

- Section VI reviews methods for assessing the ethics of AI systems. Testing whether an AI system meets ethical standards is an important and often overlooked dimension of AI ethics practice.

- Section VII highlights ongoing challenges in AI ethics and offers future research perspectives, including potential research questions and directions.

The remainder of this article is organised as follows. Section II describes the scope and methodology of the review. Section III provides a detailed summary of the ethical issues and risks of AI. Section IV reviews AI ethical guidelines and principles. Section V covers approaches to addressing ethical issues. Section VI discusses methods for evaluating the ethics of AI systems. Section VII outlines challenges and future perspectives. Section VIII concludes the article.

SCOPE AND METHODOLOGY

A. Scope

This review covers four interconnected aspects of AI ethics. The first is the identification and analysis of ethical issues and risks associated with AI, which forms the foundation of the field. The second is a review of the ethical guidelines and principles that guide the development and use of AI, reflecting the growing attention paid to these issues across academia, industry, and government. The third is an examination of approaches to solving ethical problems in AI, with a focus on ethical and technological approaches given their particular relevance to the AI research community. The fourth is a review of methods for evaluating the ethical quality of AI systems—an important but often overlooked aspect of AI ethics practice.

B. Methodology

This review draws on a wide range of documents including academic publications, organisational reports, government documents, grey literature, and news reports. The literature search was conducted in two phases. In the first phase, relevant literature was retrieved from Google Scholar, Web of Science, IEEE Xplore, ACM Digital Library, Science Direct, Springer Link, arXiv, and Google using keyword combinations drawn from the following terms:

Concept terms: ethics, ethical, responsibility, responsible, trustworthiness, trustworthy, transparent, explainable, fair, beneficial, robust, safe, private, sustainable.

Topic terms: issues, risks, guideline, principle, approach, method, evaluation, assessment, challenge.

Subject terms: artificial intelligence, AI, machine learning, ML, intelligent system, intelligent agent.

Literature published since 2010 was prioritised. In the second phase, reference lists of retrieved articles and other works by identified authors were examined to locate additional relevant sources.

For the ethical AI guidelines component, only documents available in English (or with official English translations) that were publicly accessible online were included. A complete list of the collected guidelines with URL links is provided in the Supplementary Materials.

Ethical Issues and Risks of AI

To tackle the ethical problems of AI, it is necessary first to recognise and understand the potential ethical issues or risks that AI may present. Only with this understanding can appropriate ethical guidelines, policies, principles, and rules be formulated. The ethical issues of AI refer generally to morally problematic outcomes related to AI systems that need to be identified and addressed. Issues such as lack of transparency, privacy violations, accountability gaps, bias and discrimination, safety and security concerns, and the potential for criminal or malicious use have been identified across a range of applications and studies.

This section reviews existing categorisations of AI ethical issues and proposes a new three-level categorisation that classifies AI ethical issues at the individual, societal, and environmental levels. This categorisation is intended to be comprehensive and accessible, aiding practitioners in understanding and analysing the ethical challenges raised by AI. The section also maps ethical issues to the stages of an AI system's lifecycle to help identify where in the development and deployment process particular issues are most likely to arise.

A. Categorisation of Ethical Issues

Table 1 below presents the proposed categorisation of AI ethical issues across three levels: individual, societal, and environmental. This framework comprehensively covers the ethical issues identified in the existing literature and is designed to be intuitive and easy to apply in practice.

Table 1. Categorisation of Ethical Issues and Risks of AI

Individual Level	Societal Level	Environmental Level
Safety	Fairness & Justice	Natural Resources
Privacy & Data Protection	Responsibility & Accountability	Energy
Freedom & Autonomy	Transparency	Environmental Pollution
Human Dignity	Surveillance & Datafication	Sustainability
	Controllability of AI	
	Democracy and Civil Rights	
	Job Replacement	
	Human Relationship	

At the individual level, key ethical issues include safety (the risk that AI systems may cause physical or psychological harm), privacy and data protection (the collection, processing, and misuse of personal data), freedom and autonomy (the potential for AI to limit human decision-making and self-determination), and human dignity (the risk of dehumanising treatment arising from AI-mediated interactions or decisions).

At the societal level, issues include fairness and justice (the risk of AI perpetuating or amplifying social inequalities), responsibility and accountability (uncertainty about who bears responsibility when AI causes harm), transparency (the difficulty of understanding how AI systems make decisions), surveillance and datafication (the use of AI to monitor and quantify human behaviour), controllability of AI (the challenge of keeping AI systems under meaningful human oversight), democracy and civil rights (the potential for AI to undermine democratic processes or civil liberties), job replacement (the displacement of workers by automated systems), and human relationship (the impact of AI on social bonds and interpersonal interaction).

At the environmental level, issues include the consumption of natural resources, the energy demands of large-scale AI training and inference, environmental pollution arising from hardware production and disposal, and the long-term sustainability of AI development trajectories.

Ethical Guidelines and Principles for Ai

As concerns about the ethical issues of AI have grown, many organisations—spanning academia, industry, government, and civil society—have begun to develop and publish guidelines and principles intended to guide the responsible development and use of AI. This section provides an overview of the global landscape of AI ethical guidelines and principles, drawing on an analysis of 146 guidelines published since 2015.

The guidelines reviewed represent contributions from a diverse range of actors across numerous countries including Australia, Canada, China, Denmark, Finland, France, Germany, Ireland, Japan, the Netherlands, Norway, Russia, Singapore, South Korea, Spain, Sweden, Switzerland, the United Kingdom, the United States, and several international bodies. The volume of guidelines published has increased substantially each year, reflecting growing institutional attention to AI ethics.

These guidelines address the full lifecycle of AI, from planning and design through development, production, and deployment. While the specific principles articulated vary across organisations and sectors, several themes recur consistently: transparency and explainability, fairness and non-discrimination, privacy and data protection, safety and robustness, accountability and responsibility, human oversight and control, and beneficial outcomes for individuals and society.

Despite the breadth of activity in this area, a consensus on AI ethics has not yet been achieved, and no single set of guidelines has been universally adopted. Different organisations and sectors continue to hold different views on which principles should take priority and how they should be operationalised. This remains an important challenge for the field (see Section VII).

The complete list of the 146 guidelines reviewed, with URL links, is provided in the Supplementary Materials to this article.

Approaches to Addressing Ethical Issues in AI

Addressing the ethical issues identified in Section III requires drawing on multiple disciplines and perspectives. This section reviews three broad categories of approaches: ethical, technological, and legal. While all three are important, this review gives particular attention to ethical and technological approaches, as these are most directly relevant to AI researchers and practitioners.

Ethical Approaches

Ethical approaches to AI involve applying established moral frameworks to the design, development, and deployment of AI systems. Key frameworks include virtue ethics, deontological ethics, and consequentialism. Virtue ethics focuses on cultivating morally admirable traits in AI designers and deployers. Deontological approaches seek to identify clear moral rules and duties that AI systems must respect.

Consequentialist approaches evaluate AI actions based on their outcomes, aiming to maximise benefit and minimise harm. In practice, AI ethics draws on all of these frameworks, often requiring that they be balanced or integrated depending on the context of application.

Technological Approaches

Technological approaches seek to embed ethical requirements directly into the design and architecture of AI systems. Key areas of work include: fairness-aware machine learning (developing algorithms that minimise bias and discriminatory outcomes), explainable AI (developing methods to make AI decisions interpretable to users and stakeholders), privacy-preserving AI (incorporating techniques such as differential privacy and federated learning to protect personal data), and safety and robustness engineering (building AI systems that behave predictably and resist adversarial manipulation). Significant progress has been made in each of these areas, although important challenges remain, particularly in integrating multiple ethical requirements simultaneously.

Legal and Regulatory Approaches

Legal approaches involve the development of regulations, laws, and governance frameworks that establish binding requirements for AI development and deployment. Examples include the European Union's General Data Protection Regulation (GDPR), which imposes requirements relevant to automated decision-making, and the EU's proposed AI Act, which establishes a risk-based regulatory framework for AI. Legal approaches complement ethical and technological efforts by providing enforceable standards and mechanisms for accountability. However, the pace of AI development often outstrips the ability of legal systems to adapt, and international coordination remains limited.

Methods To Evaluate Ethical Ai

The aim of AI ethics is to create AI systems that behave ethically and follow moral principles. Evaluating or measuring the ethicality and moral competence of designed AI systems is crucial. Such systems need to be tested to ensure they meet ethical requirements before deployment. However, this aspect is frequently overlooked in the existing literature. This section reviews three categories of approaches for evaluating the ethics of AI: testing, verification, and standards.

Testing

Testing is a common method for assessing the ethical capabilities of an AI system. In general terms, testing involves comparing a system's outputs against a ground truth or expected result. Applied to ethical AI, testing is used to assess how well a system performs on morally relevant criteria.

1) The Moral Turing Test (MTT): Proposed by Allen et al., the MTT is designed to evaluate whether an artificial moral agent can exhibit moral reasoning comparable to that of humans. Analogous to the original Turing Test, in which a machine passes if it cannot be distinguished from a human in conversation, the MTT restricts the evaluation to moral questions. If a human interrogator cannot distinguish the AI's moral judgements from those of a human, the AI is considered a moral agent. The MTT has been criticised, however, for focusing too heavily on the articulation of moral reasoning rather than on moral action. To address this, Allen et al. introduced the Comparative Moral Turing Test (CMTT), in which interrogators compare morally significant actions by humans and machines without being told which is which. If the AI's actions are consistently judged as less moral, it fails the test. Wallach and Allen regard the CMTT as a practical step toward evaluating AI morality in the absence of universally accepted criteria.

2) Expert and Non-expert Tests: In addition to the MTT, researchers have sought to assess AI systems' moral competence by comparing system outputs against assessments provided by either domain experts in normative ethics or non-expert participants representing public moral intuitions. Expert tests apply established ethical standards, while non-expert tests align system performance with the moral judgements of ordinary citizens.

Verification

Formal verification involves demonstrating that an AI system behaves correctly according to specified properties. As discussed by Seshia et al., a typical formal verification process checks a model of the system against a model of its environment and a set of required properties, yielding a binary yes/no result. A negative result typically includes a counterexample showing how the system fails to satisfy the property. Arnold and Scheutz proposed "design verification" as a means of assessing an AI system's moral competence, noting that MTT-based evaluations may be susceptible to deception, poor reasoning, and low moral performance.

Regardless of how an AI carries out moral reasoning, it is essential that its moral actions align with the goals of ethical design.

Standards

A number of industry standards have been proposed to guide the development and use of AI and to provide frameworks for assessing AI products:

- In 2014, the Australian Computer Society published a Professional Code of Conduct for information and communication technology professionals, identifying six key ethical values and associated conduct requirements.
- In 2018, the ACM updated its Code of Ethics and Professional Conduct to reflect changes in computing practice since 1992. The code expresses the profession's ethical commitments and is intended to guide the behaviour of all computing professionals. Each principle is supported by guidelines to assist practitioners in understanding and applying it.
- The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems has approved the development of the IEEE P7000™ standards series, covering topics from data collection and privacy to algorithmic bias.
- ISO/IEC JTC 1/SC 42 is a joint ISO and IEC committee responsible for standardising AI. The committee is developing a broad suite of standards addressing foundational AI concepts, big data, AI trustworthiness, use cases, applications, and ethical and societal concerns.

As concern about AI ethics has grown, interest in standards to shape AI design, deployment, and assessment has increased correspondingly. However, a significant gap still exists between published standards and actual practice. Currently, only some large corporations—such as IBM and Microsoft—have established their own internal standards, frameworks, and guidelines. Smaller organisations, which often lack the necessary resources, face considerable challenges in applying these principles. Developing well-structured, accessible standards and promoting their practical adoption therefore remain important priorities.

Challenges And Future Perspectives

A. Challenges in AI Ethical Guidelines and Principles

As discussed in Section IV, many guidelines have been proposed by various organisations, companies, and governments, and these reveal considerable variation in the principles they articulate. There is currently no set of guidelines that has been widely approved and adopted across different sectors. In other words, different organisations and companies—even within the same field—hold different views on AI ethics. A consensus on common principles and values for AI has not yet been achieved, and it remains unclear what baseline ethical standards all AI systems should be required to satisfy. Moreover, different ethical principles may be appropriate depending on the domain of AI application, yet domain-specific ethical discussions are rarely seen in the literature. Establishing basic, widely shared ethical principles through collaboration among organisations, sectors, and governments—and then refining them for specific application areas—is therefore a critical priority.

B. Challenges in Implementing Ethics in AI

Implementing ethics in AI faces multiple challenges across ethical, technological, and organisational dimensions. Within ethical frameworks, significant difficulties arise: virtue ethics struggles because the motives behind AI actions are opaque and virtuous traits are difficult to define or measure; deontological ethics faces difficulties in identifying appropriate moral rules and resolving conflicts between them; consequentialism is challenged by the difficulty of predicting and quantifying the outcomes of AI actions, especially in opaque models such as large neural networks. Coordinating differing ethical standards across cultures and organisations adds further complexity.

From a technological perspective, most current AI ethics research addresses isolated aspects of the problem—such as fairness or explain—without integrating multiple ethical principles, which often conflict with one another. Evaluating AI ethics remains fundamentally difficult because ethical judgements are qualitative and culturally situated, making precise and objective assessment nearly impossible.

C. Future Perspectives

Future research in ethical AI should pursue several directions. First, a flexible, multi-theory approach is needed, enabling AI systems to understand and navigate different ethical frameworks depending on context, as humans do. Combining normative ethics with domain-specific ethical requirements will be essential for user acceptance and practical relevance. Second, developing AI and machine learning models that can be guided simultaneously by multiple ethical principles is a critical and technically challenging goal. Third, effective evaluation methods must be developed and applied before deployment, with domain-specific benchmarks playing a key role in high-stakes areas such as healthcare and autonomous vehicles. Finally, integrating normative ethics—based on innate moral reasoning—with evolutionary ethics—based on learning and moral development over time—could enable AI systems to continuously acquire and refine moral competencies, offering a promising direction for future research.

CONCLUSION

This article has provided a comprehensive overview of AI ethics, addressing the ethical risks and issues raised by AI, the ethical guidelines and principles published by organisations worldwide, approaches for addressing ethical issues in AI, and methods for evaluating the ethics of AI systems. Based on this review, it is clear that designing AI systems that behave ethically is a complex and multifaceted challenge. Nevertheless, the extent to which AI can play a beneficial role in our future society depends substantially on how successfully the field rises to this challenge. The discipline of AI ethics requires a genuine joint effort from AI scientists, engineers, philosophers, users, and government policymakers. We hope that this article will serve as a useful starting point for those entering the field and a valuable reference for researchers and practitioners already engaged in advancing the practice of ethical AI.

REFERENCES

1. Z. Liu and Y. Zheng, AI Ethics and Governance: Black Mirror and Order. Singapore: Springer Singapore, 2022.
2. L. Floridi and J. Cowls, Eds., Ethics, Governance, and Policies in Artificial Intelligence. Cham: Springer Nature, 2020.
3. F. Coleman, A Human Algorithm: How Artificial Intelligence Is Redefining Who We Are. Berkeley, CA: Counterpoint Press, 2019.
4. A. Batool, D. Zowghi, and M. Bano, “Responsible AI governance: A systematic literature review,” arXiv preprint arXiv:2307.XXXXX, 2023.
5. A. Batool, D. Zowghi, and M. Bano, “AI governance: A systematic literature review,” AI and Ethics, 2025, doi: 10.1007/s43681-025-XXXX-X.
6. X. Chen, “Ethical governance of AI: An integrated approach via human-in-the-loop machine learning,” Computer Science and Mathematics Forum, vol. X, no. X, 2023.
7. J. Zhang and Z.-M. Zhang, “Ethics and governance of trustworthy medical artificial intelligence,” BMC Medical Informatics and Decision Making, vol. 23, no. 1, 2023.
8. S. Singh and H. Mohapatra, “AI ethics governance and privacy in a rapidly advancing digital world,” Systemic Analytics, vol. X, no. X, 2025.