**International Journal of Software Innovation**
Volume 14 • Issue 1 • January-December 2026

# Bridging the AI Ethics Gap:
## A Tripartite Framework for Accountability, Implementation, and Governance

Masafumi Nakano
https://orcid.org/0009-0000-9494-6152
*Toyo University, Japan*
**Received:** September 10th, 2025 | **Accepted:** February 4th, 2026

## ABSTRACT

To address the largely unmitigated societal impact of artificial intelligence (AI) technologies, this study proposes a Tripartite Ethical Policy Framework for AI governance in global industries. The framework comprises three key components: AI ethics, technical implementation, and institutional governance. Drawing on current international standards, it integrates six core principles (i.e., human-centricity, fairness, accountability, transparency, privacy, and safety) and translates them into three actionable ethical domains: data, algorithms, and autonomy. An empirical analysis of corporate disclosures in Japan reveals a gap between stated commitments and actual implementation, highlighting the need for stronger governance and credible accountability. By embedding AI ethics in internal and external audits, the framework enhances transparency, strengthens oversight, and promotes responsible AI use. Its emphasis on adaptability provides a foundation for AI audit and responsible innovation amid rapid technological change.

## KEYWORDS

AI Ethics, AI Audit, Ethical Policies Framework, Ethical Data Governance, Ethical Algorithms, Ethical Autonomy, AI Governance, Technology Disclosure, External Audit, Internal Audit

## INTRODUCTION

The rapid advancement and widespread adoption of artificial intelligence (AI) technologies have delivered substantial benefits to society. Yet these same developments have generated profound ethical risks, including discriminatory outcomes from biased data, opaque and unexplainable decision-making processes, unclear accountability for errors, privacy violations, and the potential for serious harm arising from insufficient attention to safety. Moreover, the possible emergence of artificial general intelligence—AI systems surpassing human intelligence—raises the alarming prospect of losing human control over AI behavior and intentions, posing existential risks to humanity.

In response to these urgent concerns, AI audits have gained prominence as a central mechanism for ensuring the ethical governance of AI systems. For such audits to be effective, however, they must rest on a comprehensive framework of ethical policies that guide the full life cycle of AI and software development. To this end, this study proposes a detailed AI ethical policy framework designed to address ethical risks systematically across the design, development, deployment, and operational phases of AI systems.

DOI: 10.4018/IJSI.401497

A cornerstone of ethical AI governance is the transparent disclosure of organizational ethics practices. Transparency builds public trust, enables external oversight, and ensures that technological innovation remains aligned with societal values. Yet disclosure is credible only when supported by robust internal governance mechanisms. By strengthening internal ethical oversight and establishing transparent structures for external accountability, organizations can reduce societal risks associated with AI while promoting responsible and sustainable innovation.

This study proceeds as follows. First, it reviews prior research and examines international normative trends relevant to AI ethics. Second, it presents empirical findings on corporate disclosure practices in Japan. Based on these insights, the study introduces the three principal components of the proposed framework. It then explores the application of this framework to AI audit before concluding with implications, limitations, and directions for future research.

## PRIOR STUDIES

### Trends in the Normative and Institutional Implementation of AI Ethics

Jobin et al. (2019) identified five foundational principles commonly shared across early AI ethics guidelines: transparency, justice/fairness, non-maleficence, responsibility, and privacy. However, more recent international frameworks have expanded upon this ethical foundation. Documents issued by the Organisation for Economic Co-operation and Development (OECD; Organisation for Economic Co-operation and Development, 2019), the United Nations Educational, Scientific, and Cultural Organization (UNESCO; United Nations Educational, Scientific, and Cultural Organization, 2021), and the European Union (EU; European Union, 2024) increasingly emphasize human-centricity and safety/robustness as distinct and essential ethical imperatives. As a result, despite regional and institutional differences, contemporary international AI ethics-related frameworks have begun to converge around six core principles (i.e., human-centricity, fairness and nondiscrimination, accountability, transparency and explainability, privacy and data protection, and safety and robustness).

First, human-centricity is a central focus of nearly all international normative instruments. The OECD explicitly endorses the protection of dignity, autonomy, and fundamental human rights (Organisation for Economic Co-operation and Development, 2019, Section 1.2). UNESCO (United Nations Educational, Scientific, and Cultural Organization, 2021, Section III.1) identifies human dignity as the foundational premise of ethical AI. The Institute of Electrical and Electronics Engineers (IEEE; Institute of Electrical and Electronics Engineers, 2019) extends this idea by embedding human well-being as a guiding priority throughout all stages of AI design (Principles 1 and 2).

Second, fairness and nondiscrimination are widely acknowledged as foundational to AI governance. These principles aim to prevent the reinforcement or amplification of societal biases and to promote equitable treatment across diverse populations. They are particularly relevant to the protection of economically and socially marginalized groups. The OECD (Organisation for Economic Co-operation and Development, 2019, Section 1.2) includes diversity, fairness, and nondiscrimination as key aspects of human rights protection. UNESCO (United Nations Educational, Scientific, and Cultural Organization, 2021, paras. 28–30) emphasizes equitable access to AI's benefits and the prevention of discriminatory outcomes. Likewise, the European Commission (2019, p. 19) asserts that fairness entails the elimination of unfair bias and the assurance of equal treatment.

Third, accountability in AI and software development involves clearly defining responsibilities and establishing institutional mechanisms that enable stakeholders to be held accountable for AI outcomes. This is essential for building trust and maintaining ethical integrity throughout the system life cycle. The OECD (Organisation for Economic Co-operation and Development, 2019, Section 1.5) highlights accountability as a core principle and calls for procedural safeguards and mechanisms for assigning responsibility. Similarly, UNESCO (United Nations Educational, Scientific, and Cultural Organization, 2021, para. 42) stresses the importance of traceability, auditability, and responsibility in supporting ethical oversight. IEEE (Institute of Electrical and Electronics Engineers, 2019, General