ARTIFICIAL INTELLIGENCE AND ETHICS: A GLOBAL PERSPECTIVE

YAPAY ZEKA VE ETİK: KÜRESEL BİR BAKIŞ

Elif Simge GÜZELERGENE, Deniz Baransel CINAR , Funda NAYIR

78

Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi
Yıl: 2026, Sayfa:1-27
https://dergipark.org.tr/tr/pub/cusosbil

DOI: 10.35379/cusosbil.1685334

Geliş Received: **28.04.2025**
Kabul Accepted **05.01.2026**

Keywords:
Artificial intelligence ethics, ethical implementation gap, governance logics, international policy frameworks, transparency, accountability, trustworthiness.

Anahtar Kelimeler:
Yapay zekâ etiği, etik uygulama boşluğu, yönetişim yaklaşımları, uluslararası politika çerçeveleri, şeffaflık, hesap verebilirlik, güvenilirlik.

ARTIFICIAL INTELLIGENCE AND ETHICS: A GLOBAL PERSPECTIVE [1]

YAPAY ZEKA VE ETİK: KÜRESEL BİR BAKIŞ

Elif Simge GÜZELERGENE [2], Deniz Baransel CINAR [3] , Funda NAYIR [4]

ABSTRACT

The rapid development and widespread adoption of artificial intelligence technologies necessitate their alignment with ethical principles throughout design and deployment processes. This study comparatively examines five major international reports on AI ethics published by UNESCO, the European Commission, the OECD, the European Parliament, and the Council of Europe. Employing document analysis and thematic coding methods, the research identifies transparency, responsibility, and trustworthi-ness as core ethical principles commonly emphasized across these frameworks. How-ever, the findings reveal a fundamental pattern: while organizations demonstrate normative convergence around shared ethical values, they exhibit operational diver-gence in how these principles are framed, justified, and institutionalized within distinct governance contexts. This produces an ethical implementation gap, the distance be-tween widely endorsed norms and their practical realization in policy frameworks. UNESCO pursues culturally adaptive norms through education and capacity building; the European Commission and European Parliament enforce compliance through binding regulations; OECD coordinates flexible implementation across member states; and the Council of Europe anchors AI ethics within human rights law. These divergent governance logics reflect deeper philosophical disagreements about whether ethical AI should be realized through voluntary norms, technical standards, binding legisla-tion, or rights-based frameworks. The study demonstrates that contemporary chal-lenges in AI ethics stem not from a lack of shared principles but from incompatible mechanisms for their operationalization. By analytically unpacking this implementa-tion gap, the research advances AI ethics scholarship beyond principle identification and provides actionable insights for policymakers and practitioners.

ÖZ

Yapay zekâ teknolojilerinin hızla gelişmesi ve yaygınlaşması, bu teknolojilerin etik ilkelerle uyumlu biçimde geliştirilmesini zorunlu kılmaktadır. Bu çalışma, UNESCO, Avrupa Komisyonu, OECD, Avrupa Parlamentosu ve Avrupa Konseyi tarafından yayımlanan beş büyük uluslararası yapay zekâ etiği raporunu karşılaştırmalı olarak incelemektedir. Doküman analizi ve tematik kodlama yöntemiyle yürütülen araştırmada, şeffaflık, sorumluluk ve güvenilirlik ilkelerinin tüm raporlarda ortak biçimde vurgulandığı tespit edilmiştir. Bulgular, kuruluşların ortak etik değerlerde uzlaşırken, bu ilkelerin uygulanmasında önemli farklılıklar gösterdiğini ortaya koymaktadır. UNESCO eğitim ve kapasite geliştirmeyle kültürel uyarlama; Avrupa Komisyonu ve Parlamentosu bağlayıcı yasal düzenlemeler; OECD esnek işbirliği; Avrupa Konseyi ise insan hakları hukuku yoluyla etik uygulamayı hayata geçirmektedir. Bu farklılaşma, çalışmanın temel kavramı olan "etik uygulama boşluğunu", başka bir ifade ile "yaygın kabul gören ilkeler ile gerçek uygulamalar arasındaki açığı" doğurmaktadır. Çalışma, yapay zekâ etiğindeki güncel zorlukların paylaşılan ilkelerin eksikliğinden değil, bu ilkelerin hayata geçirilmesine yönelik uyumsuz mekanizmalardan kaynaklandığını göstermektedir. Bu uygulama boşluğunu analitik olarak açıklayarak, araştırma yapay zekâ etiği literatürünü ilke belirlemenin ötesine taşımakta ve politika yapıcılar ile uygulayıcılar için eylem odaklı öneriler sunmaktadır.

[1] This study was presented as an oral presentation at the EJER Congress in 2024.

[2] Research Assistant, Pamukkale University, Faculty of Education, eguzelergene@pau.edu.tr, ORCID: 0000-0001-6629-6543

[3] Research Assistant Dr., Pamukkale University, Faculty of Education, baransel@pau.edu.tr, ORCID: 0000-0003-1614-1214. Correspondence should be addressed to Deniz Baransel Cinar.

[4] Professor Dr., Ağrı İbrahim Çeçen University, Faculty of Education, fnayir09@gmail.com, ORCID: 0000-0002-9313-4942

Alıntılamak için/Cite as: Güzelgene E. S., Cınar D. B. ve Nayır F. (2026) Artificial Intelligence And Ethics: A Global Perspective, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, s.1-27

INTRODUCTION

Artificial intelligence (AI) is defined as systems capable of mimicking human intelligence to perform specific tasks while independently managing processes such as learning, data processing, and decision-making. The primary objective of AI is to replicate functions unique to human intelligence, thereby enhancing efficiency across various fields and addressing complex problems (Russell & Norvig, 2021). Powered by advances in big data analytics, machine learning, and deep learning, these systems continually evolve and self-improve over time (Goodfellow, Bengio, & Courville, 2016). Positioned at the core of contemporary technological change, AI has been a catalyst for profound transformations in multiple sectors, including healthcare, education, and transportation, particularly in recent years.

In healthcare, AI-powered diagnostic systems assist physicians by enabling early disease detection and accelerating treatment processes, while supporting clinical decision-making (Esteva et al., 2021). In education, AI-driven personalized learning systems analyze students' learning processes to provide customized content tailored to individual needs (Luckin, Holmes, Griffiths, & Forcier, 2016). Similarly, in transportation, autonomous vehicles and intelligent traffic management systems are developed to enhance safety, reduce congestion, and improve energy efficiency (Milakis, Arem, & Van Wee, 2017). The acceleration of digitalization following the COVID-19 pandemic has further amplified the economic and societal significance of AI technologies (Acemoglu & Restrepo, 2020; McKinsey Global Institute, 2023). Estimates suggest that AI could contribute up to $13 trillion to the global economy by 2030, creating new job opportunities while simultaneously reshaping labor markets and occupational structures (PwC, 2023). At the same time, scholars warn that these transformations may intensify risks such as automation-driven unemployment and widening income inequalities (Brynjolfsson & McAfee, 2014). Moreover, global competition in AI development has intensified, with major actors such as China, the United States, and the European Union investing heavily to strengthen their strategic positions (European Parliamentary Research Service [EPRS], 2023, 2024). These developments underscore that AI is not merely a technological innovation but also a powerful driver of social, economic, and political transformation, making its ethical governance an urgent concern.

The rapid rise and widespread applications of AI have simultaneously raised pressing ethical and societal concerns. A growing body of research emphasizes the critical importance of ethical principles such as transparency, fairness, privacy, and accountability in the design and deployment of AI systems (Binns, 2018; Floridi & Cowls, 2022). As algorithmic decision-making increasingly replaces or supplements human judgment, challenges surrounding bias, discrimination, and unequal treatment have become more pronounced, particularly in sensitive domains such as education, healthcare, and employment (Mittelstadt et al., 2022). Empirical studies have demonstrated that certain AI systems may reproduce or amplify existing social inequalities, including biases related to gender or ethnicity (Buolamwini & Gebru, 2018). Similarly, AI's capability to leverage vast datasets has raised profound privacy concerns, particularly in contexts where individual rights to confidentiality may be compromised (Crawford, 2023).

Within this context, the field of AI ethics has emerged as a pivotal discipline aimed at ensuring the ethical development and trustworthy application of artificial intelligence (Allen, Wallach, & Smith, 2006; Anderson & Anderson, 2007). Rooted in ethical theories, regulatory principles, and policy frameworks, AI ethics seeks to align technological innovation with societal values and moral expectations (Siau & Wang, 2020). AI ethics serves as both a framework for guiding the ethical development of AI systems and a repository of moral values and principles that define the boundaries between acceptable and unacceptable practices. Core ethical principles such as fairness, non-discrimination, transparency, accountability, and respect for human rights constitute the normative foundation for trustworthy AI systems (Jobin, Ienca, & Vayena, 2019). Upholding these ethical standards is vital not only for mitigating risks but also for enhancing the predictability and reliability of AI's societal impact, thereby strengthening public trust and ensuring that AI contributes to socially desirable outcomes (Floridi, 2014).

In response to these concerns, international organizations have developed a range of policy documents and ethical guidelines aimed at promoting the responsible use of AI. The European Commission's Ethics Guidelines for Trustworthy AI emphasize human rights, transparency, accountability, and technical robustness as core requirements for trustworthy systems (European Commission, 2019). Similarly, the OECD AI Principles advocate for human-centered AI that serves the public interest and promotes fairness, accountability, and transparency (OECD, 2019). UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021) further stresses the need for AI to contribute to sustainable development without compromising human rights. Despite shared normative commitments, these frameworks differ in their priorities, scope, and underlying governance rationales. For example, while the European Union adopts a rights-based and regulatory approach, China's AI strategy places greater emphasis on economic development and collective societal benefits (Ding, 2018). As Whittlestone et al. (2024) argue, understanding such diversity is crucial for developing ethical AI standards and establishing a coherent global framework. The coexistence of converging ethical values and divergent policy interpretations highlights the need for systematic comparative analysis to identify both common ground and points of tension across international frameworks.

Despite the growing volume of literature on AI ethics, much existing research either focuses on articulating normative principles or summarizes policy documents without sufficiently examining the governance logics underlying their implementation. Previous studies document widespread convergence around core ethical values such as transparency, accountability, and fairness (Jobin, Ienca, & Vayena, 2019), yet they rarely analyze why this normative consensus fails to produce uniform governance practices across different institutional contexts. This phenomenon, whereby widely endorsed principles coexist with fragmented and divergent implementation mechanisms, constitutes what we term the *"ethical implementation gap."* Addressing this gap, the present study adopts a comparative and analytical perspective to examine how major international organizations conceptualize, operationalize, and constrain ethical principles in their AI governance frameworks. Rather than introducing new ethical norms, the originality of this research lies in systematically examining how shared principles such as transparency, responsibility, and trustworthiness are framed, operationalized, and constrained within distinct institutional and political contexts. By comparatively analyzing reports published by UNESCO, the European Commission, the OECD, the European Parliament, and the Council of Europe, the study reveals not only a normative convergence around core ethical values but also a significant divergence in their practical interpretation and implementation. This divergence reflects the ethical implementation gap: while organizations endorse identical principles, they pursue incompatible mechanisms for their realization, shaped by differing institutional mandates, legal traditions, and governance philosophies. This analytical focus allows the study to uncover the ethical governance logics embedded in global AI policy frameworks and to contribute to the literature by illuminating the gap between normative aspiration and regulatory reality. Accordingly, the study addresses the following research questions:

1. What common themes emerge across international reports on AI ethics?
2. How are different approaches and recommendations regarding AI ethics constructed and justified?
3. What unique contributions do these reports offer, and what challenges arise in translating ethical principles into practice?

METHODOLODY

Research Design

The purpose of this study is to conduct an in-depth analysis of reports, guidelines, and recommendations published by major international institutions in order to understand the global relationship between artificial intelligence (AI) and ethics. Within this scope, the study examines emerging common themes, diverse approaches, and unique contributions in the field of AI ethics through a comparative analytical lens. The research is structured using a basic qualitative research design. According to Merriam and

Tisdell (2016), basic qualitative research is a flexible and descriptive approach that enables an in-depth exploration of complex social phenomena by focusing on meaning, interpretation, and context. Given the multidimensional and value-laden nature of AI ethics, this design provides an appropriate framework for systematically examining how ethical principles are framed, interpreted, and operationalized across different institutional and governance contexts. Although the primary data sources consist of institutional reports and policy documents, the study does not aim to function as a systematic literature review or a descriptive compilation of existing materials. Instead, it treats these documents as qualitative data and applies thematic and comparative analysis to produce an interpretive and analytical synthesis. In this regard, the originality of the study does not stem from the novelty of the documents analyzed, but from its analytical examination of convergences and divergences in ethical framing, governance logics, and implementation rationales within global AI ethics frameworks. By leveraging this methodological approach, the study aims to contribute to both the theoretical and practical understanding of AI ethics beyond descriptive policy analysis.

Data Collection

The data utilized in this study were obtained from reports, guidelines, and recommendations published in the field of AI ethics on a global scale. The document analysis method was employed to systematically examine and evaluate these documents (Bowen, 2009). The following documents were analyzed in detail:

1. UNESCO (2021): Recommendation on the Ethics of Artificial Intelligence
2. European Commission (2019): Ethics Guidelines for Trustworthy AI by the AI High-Level Expert Group (AI HLEG)
3. OECD (2023): State of Implementation of the OECD AI Principles: Four Years Later
4. European Parliament (2020): AI Ethics: Issues and Initiatives
5. Council of Europe CAHAI (2020): Ad Hoc Committee on Artificial Intelligence - Legal Framework for AI Systems

The selected documents were published between 2019 and 2023, representing a critical period in the development of international AI ethics frameworks. Most documents (UNESCO 2021, European Commission 2019, European Parliament 2020, Council of Europe 2020) were issued during the initial articulation phase of AI ethics governance, when major organizations first established comprehensive normative frameworks. The inclusion of OECD's 2023 report, which evaluates the implementation of principles originally established in 2019, provides an updated perspective on how these foundational commitments have been interpreted and applied across member states over time. This temporal range enables the study to examine both the initial formulation of ethical principles and early reflections on their operationalization, without extending into the post-2024 regulatory phase marked by binding legal instruments such as the EU AI Act.

The documents included in this study were selected based on their relevance, institutional authority, and comprehensive treatment of artificial intelligence ethics. All analyzed texts were published by internationally recognized organizations that play a significant role in shaping global and regional AI governance agendas. These institutions exert substantial influence on policy formulation, norm-setting processes, and regulatory debates related to artificial intelligence. The selected reports collectively represent major institutional perspectives on AI ethics, each offering distinct, yet complementary frameworks grounded in their respective organizational missions and mandates. In addition, the corpus reflects both global and regional dimensions of AI governance, encompassing worldwide frameworks developed by organizations such as UNESCO and the OECD, as well as regionally oriented initiatives produced by European Union institutions and the Council of Europe. Finally, the selected documents have been widely cited in academic literature and policy discussions, indicating their substantial impact on subsequent research and regulatory developments in the field of AI governance. Details about the documents analyzed in this study are presented in Table 1.

Table 1. Reports on AI Ethics by International Organizations and their Scopes

Report / Organization	Purpose (Organization)	Core Focus and Principle(s)	Scope and Objectives
UNESCO-Recommendation on the Ethics of AI (2021)	Promoting international collaboration in education, science, and culture.	Alignment with human rights and the environment, data privacy, individual autonomy, algorithmic transparency.	Inclusive and sustainable integration, international cooperation, strategies, and capacity building.
European Commission - Ethics Guidelines for Trustworthy AI (2019)	Ensuring the implementation of EU policies.	Seven principles for trustworthy AI: human autonomy, non-maleficence, fairness, transparency, social and environmental well-being, accountability, privacy.	Practical recommendations, technical robustness, and oversight mechanisms.
OECD - State of Implementation of AI Principles (2023)	Supporting economic cooperation and sustainable development.	OECD principles: trust, transparency, societal welfare, and public interest.	Implementation analyses, assessment of successes and challenges, policy directions.
European Parliament - AI Ethics: Issues and Initiatives (2020)	Regulating technology policies and serving as a legislative body.	Ethical, legal, and socio-economic dimensions; focus on human rights and democratic values.	Policy recommendations, development of regulatory frameworks, evaluation of social and economic dimensions.
Council of Europe - Legal Framework for AI Systems (2020)	Establishing standards in human rights, democracy, and the rule of law.	Compliance with human rights, democracy, and the rule of law; balancing risks and opportunities.	Developing a global legal framework, promoting ethical use, and harmonizing policies.

Table 1 summarizes the core principles, scopes, and objectives presented by each organization regarding AI ethics. As shown in the table, each institution approaches AI ethics through the lens of its organizational mandate: UNESCO emphasizes global collaboration and cultural sensitivity; the European Commission and Parliament prioritize rights-based regulatory frameworks; OECD focuses on policy coordination and implementation assessment; and the Council of Europe grounds its approach in human rights law. Collectively, these reports demonstrate that while organizations share commitment to ethical AI development, their governance priorities and implementation strategies vary considerably.

Data Analysis

The collected data were analyzed using thematic analysis, a systematic qualitative method designed to identify and analyze patterns within textual data (Braun & Clarke, 2006). This method is particularly suited to examining how ethical principles are articulated across diverse institutional contexts. The analysis process followed the structured approach illustrated in Figure 1.

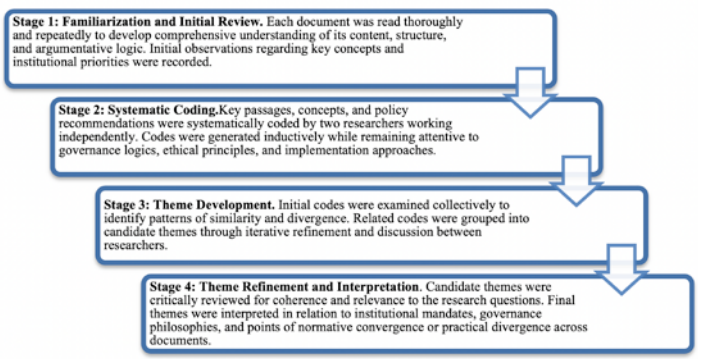

Figure 1. Data Analysis Process

Trustworthiness and Rigor

To enhance the reliability of the analysis, coding was conducted independently by two researchers. Inter-coder agreement was calculated using Cohen's Kappa, resulting in a coefficient of 0.85, which indicates a high level of reliability. Discrepancies were discussed until consensus was reached. Validity was strengthened through contextual interpretation of each document and expert consultation during the analytical process. Expert feedback supported the consistency and conceptual relevance of the identified

themes. This triangulated approach helped minimize interpretive bias and ensured that findings accurately reflected the intent and scope of the analyzed documents.

Ethical Considerations

All data used in this study were obtained from publicly accessible documents; therefore, formal ethical approval was not required. Nevertheless, ethical research principles were strictly observed. All sources were accurately cited, and the original context of each document was preserved to avoid misrepresentation. No conflicts of interest were identified during the research process.

Limitations

This study is limited to a selected set of international AI ethics reports published between 2019 and 2023 by major intergovernmental and regional organizations. National-level policies, sector-specific regulations, and corporate AI ethics frameworks fall outside the scope of the analysis. For instance, national frameworks such as Singapore's Model AI Governance Framework, China's New Generation AI Development Plan, or Brazil's AI Strategy reflect region-specific cultural values, economic priorities, and regulatory traditions that may differ substantially from the international frameworks analyzed here. Similarly, sector-specific guidelines developed by professional associations (e.g., medical AI ethics by healthcare regulatory bodies) or corporate AI principles (e.g., Microsoft's Responsible AI Standards, Google's AI Principles) embody organizational and industry-specific interpretations of ethical commitments that are not captured in this study. As a result, certain contextual variations in how AI governance is interpreted and implemented at national, sectoral, or organizational levels may not be fully captured. Moreover, these documents precede several significant regulatory developments, most notably the adoption of the European Union Artificial Intelligence Act in 2024, which marks a transition from predominantly voluntary ethical guidance toward legally binding regulatory obligations. Nevertheless, the analytical contribution of the present study does not lie in assessing the effectiveness of these newer regulatory instruments, but in systematically examining how foundational ethical principles were initially articulated and embedded within distinct institutional and governance logics. Understanding these foundational frameworks remains essential for interpreting subsequent regulatory evolution, as contemporary binding regulations continue to draw upon and institutionalize the normative commitments established in earlier international policy documents. In addition, this study adopts a document-centered qualitative approach and does not incorporate empirical data on implementation practices or the perspectives of key stakeholders such as policymakers, developers, or end users.

Future research could address these limitations by expanding the corpus to include national AI frameworks from regions underrepresented in this study, such as Singapore's Model AI Governance Framework, China's New Generation AI Development Plan, or Brazil's AI Strategy, which reflect distinct cultural, economic, and regulatory contexts. Additionally, incorporating sector-specific guidelines (e.g., healthcare regulatory standards, financial sector AI frameworks) and corporate AI principles would reveal how ethical commitments are interpreted within specific industry and organizational contexts. Empirical research examining stakeholder perspectives would complement this document-centered analysis. Qualitative studies with AI developers could explore how transparency requirements are operationalized in practice; interviews with compliance officers could reveal organizational barriers to accountability; and surveys with end users could assess whether deployed systems meet ethical expectations. Finally, longitudinal research tracking how principles evolve from voluntary guidelines (2019-2023) to binding regulations (post-2024) would clarify the conditions under which ethical implementation gaps narrow or persist.

FINDINGS

The comparative analysis of the five international reports on AI ethics identified three main themes and their associated sub-themes: *Common Themes, Divergent Approaches and Recommendations, and Unique Contributions and Potential Challenges. These* themes provide a framework for examining the reports' core approaches and areas of focus. Figure 2 illustrates the hierarchical relationships between the main themes and sub-themes identified in the analysis.

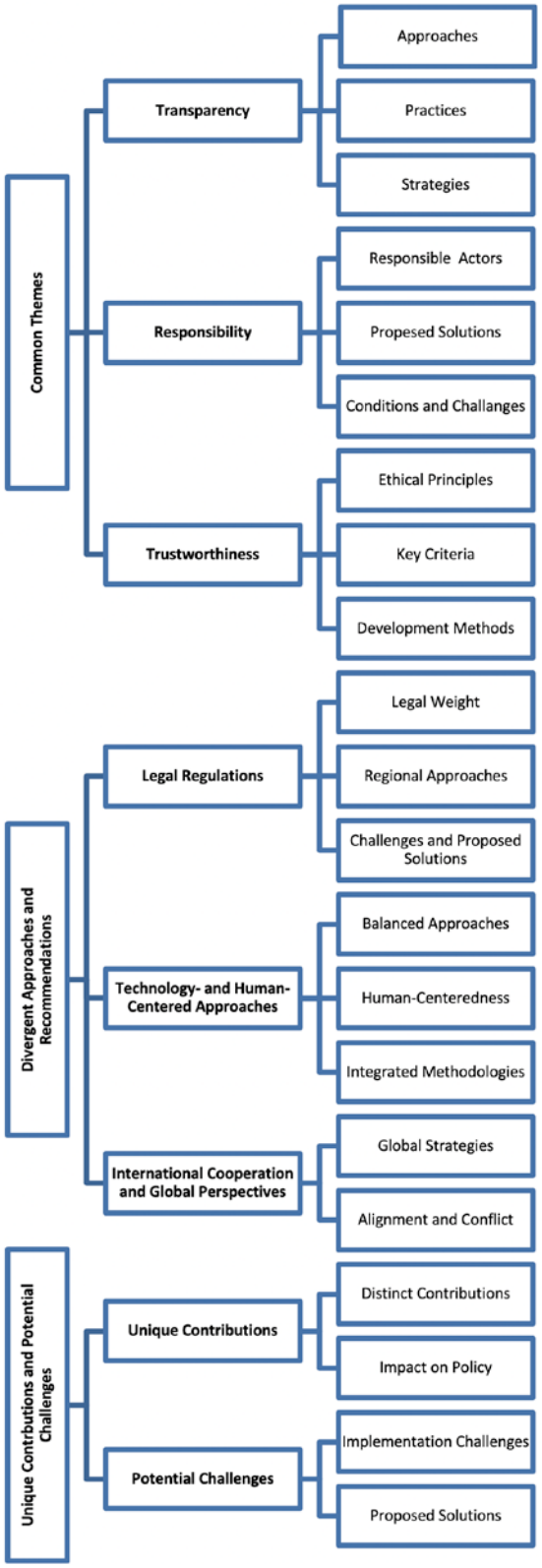

Figure 2. Thematic Framework of International AI Ethics Reports

As shown in Figure 2, the thematic structure reveals both areas of normative convergence, particularly around transparency, responsibility, and trustworthiness, and points of divergence in how these principles are operationalized across different institutional contexts. The hierarchical organization demonstrates that while all reports share commitment to core ethical values, their governance approaches and implementation strategies vary considerably. Table 2 provides a detailed description of each theme and its constituent sub-themes, outlining their specific focus areas and analytical significance.

Table 2. Scope of Themes and Sub-Themes Identified in the Analysis

Theme/Sub-Theme		Description
1. Common Themes	**1.a Transparency**	Emphasizes the need for AI systems to be transparent and understandable. Includes approaches, practical examples, and strategies.
	1.b Responsibility	Focuses on identifying responsible actors, proposing solutions, and addressing potential challenges in implementation.
	1.c Trustworthiness	Aims to develop a reliable infrastructure through ethical principles, key criteria, and development methods.
2. Divergent Approaches and Recommendations	**2.a Legal Regulations**	Proposes regional and international legal frameworks for AI systems, emphasizing the applicability of ethical principles.
	2.b Technology- and Human-Centered Approaches	Highlights some reports' technology-focused approaches and others' human-centered perspectives, advocating for a balanced methodology.
	2.c International Cooperation and Global Perspectives	Explores global collaborations, strategies, and alignment efforts. Evaluates cooperation processes and diverse recommendations.

3. Unique Contributions and Potential Challenges	**3.a Unique Contributions**	Assesses original contributions to AI ethics, including policy impacts and guidance for implementation.
	3.b Potential Challenges	Identifies obstacles to the ethical implementation of AI systems and proposes strategies to overcome these challenges.

The distribution of sub-themes across these three main categories reflects the dual nature of international AI ethics discourse: while organizations converge around fundamental ethical principles, they diverge significantly in their regulatory approaches, institutional priorities, and implementation mechanisms. This pattern suggests that global AI ethics governance is characterized by normative consensus alongside operational pluralism. The following sections examine each theme and its sub-themes in detail.

Findings on Common Themes Identified in International Reports on AI Ethics

The analysis indicates that three ethical principles recur across all five reports and form the shared normative core of global AI ethics frameworks: Transparency, Responsibility, and Trustworthiness. Across the documents, transparency is primarily linked to explainability and disclosure practices; responsibility is framed through the distribution of accountability among actors and oversight mechanisms; and trustworthiness is associated with the reliability of AI systems and safeguards against societal harm. Tables 3-5 present the codes and categories underpinning these common themes.

Table 3. Findings on the Common Theme of Transparency

Reports	Common Theme- Transparency		
	Approach to the Principle of Transparency	**Examples of Practices and Recommendations**	**Strategies**
UNESCO (2021)	Transparency in development and implementation phases.	Examples: Data provenance (documenting data sources, collection methods, and intended uses), Ethical labeling (indicating the ethical principles guiding algorithm design and testing). Recommendations: Transparency reports, international standards.	Development of global frameworks. Education and awareness programs.
European Commission (2019)	Comprehensibility and traceability. Technical and communicative transparency.	Examples: Use of explainability tools in decision-making processes, Simulation tests. Recommendations: Development of ethical codes, independent audits.	Use of technical explanation tools. Regular oversight and audits.
OECD (2023)	Comprehensibility and auditability. Openness and accountability.	Examples: Source identification systems, User experience reports. Recommendations: International protocols, Ethical auditing mechanisms.	Guides for policymakers. International cooperation and sharing.
European Parliament (2020)	Transparency in monitoring and management. Democratic oversight. Openness.	Examples: Real-time monitoring tools, Dialogue and participation platforms. Recommendations: Transparency training.	Legal frameworks and regulations. Public participation and oversight.
Council of Europe (2020)	Technical transparency. Compliance with human rights, democracy, and the rule of law.	Examples: Multi-stakeholder review committees, Open-source tools. Recommendations: International transparency agreements, Approaches respectful of human rights	Multi-stakeholder platforms. Human rights-based approaches.

As shown in Table 3, transparency is conceptualized differently across institutional contexts. UNESCO and OECD frame transparency through global normative frameworks, emphasizing comprehensibility and accountability across diverse cultural contexts. The European Commission distinguishes between technical transparency (algorithm explainability) and communicative transparency (user-facing disclosure), reflecting the EU's dual emphasis on technical robustness and consumer protection. The European Parliament uniquely positions transparency as essential for democratic oversight, linking algorithmic disclosure to public participation mechanisms. The Council of Europe integrates transparency within its human rights framework, treating explainability not merely as a technical requirement but as a precondition for rights-based governance. Despite these variations, all reports converge on the necessity of documentation practices (data provenance, algorithmic audits) and stakeholder engagement (transparency reports, public dialogue). The proposed strategies reveal a common trajectory from voluntary guidelines toward institutionalized oversight mechanisms, though implementation pathways differ ranging from UNESCO's education-focused approach to the European Parliament's legally mandated disclosure requirements.

Tablo 4. Findings on the Common Theme of Responsibility

Reports	Common Theme- Responsibility		
	Responsible Actors	**Responsibility Mechanisms: Solutions**	**Conditions and Challenges**
UNESCO (2021)	Developers, implementers, and government bodies.	International standards, global oversight.	**Conditions:** International cooperation and inclusive policies. **Challenges:** Incompatibility between legal and cultural frameworks of different countries.
European Commission (2019)	Designers, implementers, and sectoral regulators.	Regulatory frameworks, technical oversight tools.	**Conditions:** Strong regulatory frameworks and advanced oversight tools. **Challenges:** Adapting to technology and ensuring continuous updates.
OECD (2023)	Developers and national and international regulators.	Policy guidelines, international collaboration, sharing of best practices.	**Conditions:** Cross-country information sharing and clear policy guidelines. **Challenges:** Ensuring stability in the implementation of standards.
European Parliament (2020)	Developers, public, and private sectors.	Legal regulations, public engagement.	**Conditions:** Public participation, transparency, and legal frameworks. **Challenges:** Balancing public and private sector interests, legal enforcement.
Council of Europe (2020)	Multi-stakeholder committees, developers.	Human rights-based and multi-stakeholder approach.	**Conditions:** Ethical frameworks based on human rights and multi-stakeholder review mechanisms. **Challenges**: Multi-stakeholder coordination and conflicts of interest.

Table 4 reveals significant variation in how responsibility is distributed and enforced. UNESCO adopts a diffuse accountability model, distributing responsibility across developers, implementers, and government bodies through international cooperation. In contrast, the European Commission and European Parliament favor a regulatory accountability model, clearly delineating roles for sectoral regulators and establishing enforceable legal frameworks. OECD emphasizes policy-based coordination, relying on voluntary guidelines and best practice sharing rather than binding obligations. A notable divergence emerges in accountability mechanisms: while UNESCO and OECD propose global standards and international protocols, the European Parliament and European Commission pursue legally enforceable regulations backed by technical oversight tools. The Council of Europe uniquely integrates multi-stakeholder governance, distributing responsibility across diverse actors while grounding accountability in human rights obligations. All reports acknowledge similar challenges, legal and cultural incompatibilities (UNESCO), rapid technological change (European Commission), and coordination difficulties (Council of Europe), yet propose divergent solutions reflecting their institutional mandates. This pattern suggests that while responsibility is universally recognized as essential, its operationalization remains contested and context dependent.

As illustrated in Table 5, trustworthiness is anchored in divergent ethical foundations. UNESCO emphasizes procedural fairness and cultural sensitivity, reflecting its global mandate to accommodate diverse value systems. The European Commission prioritizes technical reliability (algorithm accuracy, data management), consistent with the EU's risk-based regulatory approach. OECD uniquely foregrounds individual autonomy and inclusivity, aligning with liberal democratic principles central to its member states. The Council of Europe distinguishes itself by explicitly subordinating AI trustworthiness to human rights compliance, treating technical criteria as secondary to normative alignment with democratic values. This contrasts sharply with the European Commission's technocratic focus on measurable performance indicators. Development methods reveal a spectrum from soft governance (UNESCO's education programs, OECD's voluntary harmonization) to hard governance (European Parliament's independent regulatory bodies, European Commission's binding sectoral standards). Despite these differences, all reports converge on the necessity of continuous auditing and multi-stakeholder participation, suggesting emergent consensus on process-based trust rather than purely technical assurance. Collectively, these three themes, *transparency, responsibility, and trustworthiness*, constitute the normative foundation of international AI ethics discourse. However, their translation into governance practice varies substantially across institutional contexts, reflecting divergent priorities between global inclusivity, technical robustness, democratic accountability, and human rights protection.

Table 5. Findings on the Common Theme of Trustworthiness

Reports	Common Theme- Trustworthiness		
	Ethical Principles	**Key Criteria**	**Development Methods**
UNESCO (2021)	Fairness, justice, transparency, accountability.	Transparency in decision-making processes, culturally sensitive algorithms.	International cooperation and information sharing, education, and awareness programs.
European Commission (2019)	Trustworthiness, protection of privacy.	Algorithm accuracy and reliability, data management.	Sectoral regulations and standards, technical audits, and continuous development.
OECD (2023)	Autonomy, freedom, collaboration, inclusivity.	Multi-stakeholder collaboration, establishment of standards, continuous auditing.	Harmonization of international standards, cross-sectoral audits, and transparency reports.
European Parliament (2020)	Fair access, justice, cultural diversity.	Transparency, fair oversight, respect for consumer rights.	Public participation, independent regulatory bodies.
Council of Europe (2020)	Respect for human rights, democracy, and the rule of law.	Compliance with human rights, development of policies aligned with democratic values.	Multi-stakeholder review committees, ethical frameworks.

Table 6. Findings on Divergent Approaches and Recommendations Regarding Legal Regulations

Reports	Divergent Approaches and Recommendations- Legal Regulations		
	Legal Emphasis	**Regional Recommendations**	**Legal Challenges and Solutions**
UNESCO (2021)	Establishing global principles and ethical norms.	Developing legal frameworks reflecting global diversity and cultural norms.	**Challenges:** Preserving cultural diversity, establishing and implementing universal legal standards and ethical norms. **Solutions:** Promoting cultural diversity and ethical norms through international cooperation.
European Commission (2019)	Strengthening legal frameworks within Europe.	A Europe-centered approach, alignment with other regions.	**Challenges:** Ensuring reliability and transparency in AI applications. **Solutions:** Developing ethical standards and regulatory frameworks.
OECD (2023)	Legal harmonization and standardization, providing a general framework.	Aligning legal regulations among member countries, promoting global cooperation.	**Challenges:** Data protection and privacy. **Solutions:** Developing international standards and policies, fostering cooperation among member countries.
European Parliament (2020)	Legal regulations and policy recommendations within Europe.	Europe-focused legal frameworks.	**Challenges:** Addressing the impact of AI on the workforce. **Solutions:** Updating and enhancing legal frameworks, focusing on consumer rights.
Council of Europe (2020)	Detailed legal framework based on human rights.	Global legal frameworks rooted in human rights.	**Challenges:** Adapting legal frameworks to the rapid development of AI. **Solutions:** Continuously updated legal frameworks, ethical solutions based on human rights.

Findings on Divergent Approaches and Recommendations in International Reports on AI Ethics

While the reports converge on core ethical principles, they diverge substantially in how these principles are translated into governance strategies. These differences cluster around three sub-themes: *Legal Regulations, Technology- and Human-Centered Approaches, and International Cooperation and Global Perspectives* (Tables 6-8).

Table 6 reveals a fundamental tension between universalist and regionalist approaches to AI regulation. UNESCO adopts a universalist stance, emphasizing culturally sensitive global norms that accommodate diverse legal traditions. This contrasts sharply with the European Commission and European Parliament, which prioritize Europe-centered regulatory frameworks designed to protect EU citizens while maintaining external alignment capacity. OECD occupies a middle position, pursuing legal harmonization among member states through voluntary coordination rather than binding regulation. The Council of Europe uniquely grounds its legal framework in human rights obligations, treating legal compliance not as technical standardization but as normative alignment with democratic principles. This human rights-centric approach diverges from the European Commission's risk-based regulatory model and OECD's economic coordination framework. Regarding implementation challenges, the reports identify distinct obstacles reflecting their institutional priorities: UNESCO emphasizes cultural diversity preservation, the European Commission focuses on technical reliability assurance, OECD highlights data protection harmonization, and the Council of Europe addresses adaptive legal frameworks responsive to technological change. These divergent challenge framings reveal underlying differences in governance philosophies, between cultural pluralism, consumer protection, economic coordination, and rights-based regulation.

Table 7. Findings on Divergent Approaches and Recommendations Regarding Technology-and-Human-Centered Approaches

Reports	Divergent Approaches and Recommendations- *Technology- and Human-Centered Approaches*		
	Balanced Approach	***Human-Centeredness***	***Integrated Models***
UNESCO (2021)	AI systems supporting cultural diversity and aligning with human rights.	Increasing human participation in education and capacity development.	International standardization of ethical principles and their integration into technology design.
European Commission (2019)	Reliable AI systems sensitive to user needs.		Policy recommendations and legal regulations for integrating ethical principles into technological processes.
OECD (2023)	Joint evaluation of technology and human resources to enhance human capacity.	Active human involvement in the design and implementation of AI systems, sensitivity to human needs.	Guidelines and frameworks for the tangible integration of ethical principles into AI design.
European Parliament (2020)	Development of AI systems harmonious with humans, emphasizing ethical and human-rights-respecting solutions.	Human involvement in processes of compliance with legal and ethical norms.	Continuous and dynamic integration of ethical principles into AI technologies.
Council of Europe (2020)		Policy recommendations and legal regulations for integrating ethical principles into technological processes.	Policy recommendations and legal regulations for integrating ethical principles into technological processes.

As shown in Table 7, the reports differ in whether they foreground technological capability or human agency as the primary driver of ethical AI. UNESCO and OECD adopt explicitly human-centered frameworks, emphasizing capacity building, education, and active human involvement in AI design and governance. The European Commission, while acknowledging user needs, focuses more on technical reliability standards, treating human-centeredness as an output of robust engineering rather than a participatory governance principle. The European Parliament and Council of Europe position human-centeredness within legal and rights-based frameworks, emphasizing compliance mechanisms rather than participatory design processes. This reflects a governance logic that prioritizes regulatory protection over participatory empowerment. Integration models also diverge: UNESCO pursues international standardization, the European Commission and Council of Europe rely on binding legal regulations, OECD provides non-binding guidelines, and the European Parliament advocates for continuous adaptive integration. This spectrum, *from soft norms to hard law*, reflects institutional variation in regulatory capacity and political mandate.

Table 8. Findings on Divergent Approaches and Recommendations Regarding International Cooperation and Global Perspectives

Reports	Divergent Approaches and Recommendations- International Cooperation and Global Perspectives		
	Importance of International Cooperation	**Global Strategies**	**Points of Alignment and Conflict**
UNESCO (2021)	Establishing global standards and harmonization.	Promoting participation and reflecting universal values.	Misalignment due to cultural and legal differences.
European Commission (2019)	Creating ethical standards at the international level.	Promoting international cooperation and multi-stakeholder approaches.	Misalignment between regional values and international norms => Addressing diversity as common ground.
OECD (2023)	Adopting, implementing, and standardizing AI principles globally.	Supporting policy and information exchange, providing coordination guidelines.	Misalignment due to differences in economic and technological development => Flexible policies.
European Parliament (2020)	Establishing a framework to support ethical and secure AI usage.	Aligning AI ethical principles with European norms.	Misalignment between European and other international norms => Broader and inclusive policies.
Council of Europe (2020)	Creating a global legal framework rooted in human rights and democratic values with broad impact.	Establishing a comprehensive legal framework and extensive multi-stakeholder collaborations.	Misalignment between global and regional ethical standards => Detailed legal frameworks.

Table 8 reveals contrasting visions of global AI governance. UNESCO advocates for bottom-up harmonization that accommodates cultural diversity, while the Council of Europe pursues top-down human rights frameworks intended as universal standards. The European Commission and European Parliament exhibit regional assertiveness, seeking to align international standards with European norms rather than adapting European regulation to global consensus. OECD uniquely emphasizes economic and technological development disparities as barriers to cooperation, proposing flexible policies that accommodate varying national capacities. This contrasts with UNESCO's cultural pluralism framing and the Council of Europe's rights-uniformity approach. Regarding conflict resolution strategies, the reports propose divergent pathways: UNESCO emphasizes cultural accommodation, the European Commission seeks diversity as common ground, OECD advocates flexible implementation, the European Parliament pursues broader inclusivity, and the Council of Europe relies on detailed legal frameworks. These strategies reflect fundamentally different assumptions about whether global AI governance should prioritize procedural flexibility (UNESCO, OECD) or substantive universalism (Council of Europe, European Parliament). Collectively, these divergences suggest that while international organizations share rhetorical commitment to cooperation, they envision fundamentally different governance architectures, ranging from soft norm harmonization to binding legal universalism, from participatory multi-stakeholder processes to regulatory enforcement mechanisms, and from cultural pluralism to rights-based uniformity.

Findings on Unique Contributions and Potential Challenges in AI Ethics

Beyond shared principles and divergent strategies, the analysis identifies distinct institutional contributions and implementation challenges. These are organized under two sub-themes: Unique Contributions and Potential Challenges and Solutions (Tables 9-10).

As shown in Table 9, each organization contributes a distinctive conceptual or institutional innovation to global AI ethics discourse. UNESCO's primary contribution lies in establishing normative universalism, articulating ethical principles intended to transcend regional variations while accommodating cultural diversity. The European Commission introduces the operationalized concept of "Trustworthy AI", which has gained significant traction beyond the EU and now serves as a reference point in international policy discussions. OECD's unique contribution is implementation-focused pragmatism: rather than articulating new ethical principles, it systematically evaluates how existing principles are translated into practice across diverse national contexts, providing evidence-based policy guidance. The European Parliament contributes legislative integration mechanisms, demonstrating pathways for embedding ethical principles within binding legal frameworks. The Council of Europe uniquely grounds AI ethics within international human rights law, treating AI governance not as a new regulatory domain but as an extension of existing rights obligations. These contributions reflect divergent institutional mandates: UNESCO operates through norm diffusion, the European Commission through regulatory standardization, OECD through policy coordination, the European Parliament through legislative action, and the Council of Europe through rights-based juridification. Collectively, these approaches create a multi-layered global governance ecosystem in which voluntary guidelines, binding regulations, and legal obligations coexist and interact.

Table 9. Findings on the Unique Contributions of Reports

Reports	Unique Contributions and Potential Challenges- Unique Contributions	
	Unique Contributions	**Impact on Policy**
UNESCO (2021)	Global standards, legal frameworks.	Shaping global policies, broad acceptance.
European Commission (2019)	The concept of 'Trustworthy AI,' ethical guidelines.	Impact on European and global policy frameworks.
OECD (2023)	Evaluation of global practices, policy recommendations.	Coordination among member states, promoting ethical and human-centered AI usage.
European Parliament (2020)	Review of legal frameworks, integration of ethics.	Shaping European AI policies and legislation.
Council of Europe (2020)	Human rights-based legal frameworks.	Enhancing international cooperation, promoting global alignment.

Table 10. Findings on the Potential Challenges Presented by the Reports

Reports	Unique Contributions and Potential Challenges- Potential Challenges	
	Potential Challenges	**Proposed Solutions**
UNESCO (2021)	Cultural and legal alignment challenges.	Increasing international cooperation.
European Commission (2019)	Applicability of abstract principles.	Providing concrete examples and updated guidelines.
OECD (2023)	Capacity differences among countries.	Encouraging information sharing.
European Parliament (2020)	Adaptation to technological changes.	Developing flexible legal frameworks.
Council of Europe (2020)	Challenges in international cooperation.	Developing international standards and agreements.

Table 10 reveals that the reports identify challenges corresponding to their respective governance approaches. UNESCO confronts the inherent tension in global norm-setting: establishing universal principles while respecting legal and cultural pluralism. The European Commission acknowledges the abstraction-application gap, the difficulty of translating high-level ethical principles into concrete technical requirements and organizational practices. OECD highlights asymmetric capacity as a barrier to harmonization: countries with varying technological capabilities, regulatory infrastructures, and economic resources face different implementation challenges, making uniform standards difficult to achieve. The European Parliament identifies regulatory lag: the challenge of maintaining legal frameworks that remain relevant amid rapid technological change. The Council of Europe emphasizes coordination complexity: the difficulty of aligning diverse legal systems and institutional actors within multilateral governance mechanisms. Proposed solutions reflect each organization's governance logic. UNESCO relies on voluntary cooperation, the European Commission on technical guidance and exemplars, OECD on information sharing and peer learning, the European Parliament on adaptive legal design, and the Council of Europe on binding international agreements. This solution diversity underscores a fundamental governance dilemma: whether AI ethics implementation should prioritize flexibility (accommodating diverse contexts) or uniformity (ensuring consistent protection).

The comparative analysis of five international AI ethics reports identified three main thematic clusters. First, all reports converge around transparency, responsibility, and trustworthiness as core ethical principles (Tables 3-5). Second, they diverge substantially in how these principles are translated into governance strategies, particularly regarding legal regulations, technology-human balance, and international cooperation (Tables 6-8). Third, each organization contributes distinct institutional innovations while confronting different implementation challenges (Tables 9-10). These patterns reveal both normative alignment and operational variation across international AI ethics frameworks. While organizations share commitment to fundamental ethical values, they differ in how these values are framed, justified, and embedded within institutional governance mechanisms. This divergence between widely endorsed principles and diverse operationalization approaches constitutes what we term the "ethical implementation gap", a pattern that will be examined analytically in the Discussion section

DISCUSSION

Introduction: From Normative Convergence to Operational Divergence

The rapid advancement of AI technologies and their deepening societal impacts have prompted international organizations to articulate ethical frameworks for trustworthy, human-centered, and socially beneficial AI development. This study analyzed five major international AI ethics reports to identify common themes, divergent approaches, unique contributions, and implementation challenges. The findings reveal substantial normative convergence around core ethical principles, particularly transparency, responsibility, and trustworthiness, yet significant operational divergence in how these principles are interpreted, justified, and embedded within distinct governance architectures.

Recent scholarship on AI ethics and governance similarly highlights both the proliferation of high-level principles and the persistent gap between ethical aspirations and implementation practices. Large-scale reviews of ethics guidelines demonstrate growing convergence around values such as fairness, accountability, and transparency, while substantial divergence remains in how these principles are interpreted, which actors they address, and how they are enforced in practice (Jobin, Ienca, & Vayena, 2019; Corrêa et al., 2023). More recent work emphasizes the need to move beyond abstract ethical principles toward concrete governance mechanisms, organizational routines, and accountability structures that support implementation across the AI lifecycle (Batool et al., 2025; Papagiannidis et al., 2025).

The adoption of the European Union Artificial Intelligence Act in 2024 exemplifies how ethical principles articulated in earlier international reports have begun to crystallize into binding, risk-based regulatory mechanisms (Cancela-Outeda, 2024). Rather than rendering prior ethical

frameworks obsolete, these developments underscore their formative role in shaping contemporary AI regulation. From this perspective, the present study provides an analytical foundation for understanding the ethical genealogy of current and emerging AI regulations by clarifying how widely endorsed normative principles are transformed into governance rationales.

It is important to clarify the analytical scope and contribution of this study in relation to more recent regulatory developments. The documents analyzed (2019-2023) represent the formative period during which international organizations first articulated comprehensive AI ethics frameworks—before these principles were codified into binding law. The adoption of the EU Artificial Intelligence Act in 2024 marks a critical transition from voluntary guidelines to legally enforceable obligations (Cancela-Outeda, 2024). However, rather than rendering earlier frameworks obsolete, this regulatory evolution validates the importance of understanding their foundational logic: contemporary binding regulations draw directly upon the normative commitments and governance rationales established in these earlier documents.

The originality of this study lies not in surveying the most recent legal instruments, but in analytically unpacking the governance logics through which widely endorsed ethical principles are interpreted and institutionalized differently across international contexts. By examining this formative period, the study clarifies why normative consensus has not produced uniform implementation, a pattern that persists even as voluntary guidelines evolve into binding regulations. Understanding this ethical implementation gap is essential for interpreting current and future regulatory developments, as the challenges identified in 2019-2023 frameworks continue to shape post-2024 governance debates. Recent scholarship on responsible AI governance reinforces this point, demonstrating that effective implementation requires not merely updating principles but transforming the organizational structures and enforcement mechanisms through which they are realized (Batool et al., 2025; Xiong et al., 2025).

This discussion contextualizes the findings within existing literature, interprets their theoretical and practical implications, and advances the central argument of the study: that contemporary debates on AI ethics are not primarily hindered by a lack of shared ethical principles, but rather by the institutional, legal, and political mechanisms through which these principles are translated into practice.

The Ethical Implementation Gap: Shared Principles, Divergent Governance Logics

The comparative analysis reveals a fundamental pattern in global AI ethics governance: normative convergence at the level of principles coexists with operational divergence at the level of implementation. While all five reports endorse transparency, responsibility, and trustworthiness as core ethical commitments, they frame and operationalize these principles differently depending on their institutional mandates, legal traditions, and governance philosophies. This divergence produces what we term an ethical implementation gap, the distance between widely endorsed ethical norms and their practical realization in policy frameworks.

As demonstrated in the findings, transparency, responsibility, and trustworthiness function not as universal, self-executing principles but as conceptual frameworks that are interpreted and institutionalized through distinct governance logics:

In UNESCO's framework, ethical principles are primarily articulated as value-based and normatively aspirational, emphasizing education, cultural sensitivity, and global moral responsibility. UNESCO treats transparency as a matter of cultural accommodation and capacity building, responsibility as distributed across developers and governments through voluntary cooperation, and trustworthiness as grounded in respect for diverse value systems (UNESCO, 2021). This approach reflects UNESCO's institutional mandate to promote international collaboration while preserving cultural diversity, a governance logic prioritizing normative pluralism over regulatory uniformity.

The European Commission and European Parliament, in contrast, translate similar principles into legally oriented and operational mechanisms, linking ethics to compliance,

accountability, and regulatory enforceability. Transparency is framed through technical explainability requirements and mandatory disclosure obligations; responsibility is assigned to specific actors (developers, deployers, regulators) with clear legal consequences for non-compliance; and trustworthiness is operationalized through conformity assessment procedures and certification schemes (European Commission, 2019; European Parliament, 2020). This governance logic reflects the EU's regulatory tradition of rights-based enforcement and consumer protection.

The OECD adopts a policy-coordination perspective, framing ethical principles as instruments for international alignment, risk management, and voluntary implementation guidance rather than binding regulation. The OECD treats transparency as a matter of best practice sharing, responsibility as multi-stakeholder coordination, and trustworthiness as continuous evaluation and adaptive policymaking (OECD, 2023). This reflects the OECD's institutional role as a forum for soft governance among economically diverse member states, prioritizing flexibility and mutual learning over legal harmonization.

The Council of Europe embeds ethical principles within human rights jurisprudence, treating AI ethics not as a new regulatory domain but as an extension of existing legal obligations under international human rights law. Transparency is subordinated to due process and the right to explanation; responsibility is anchored in state obligations to protect fundamental rights; and trustworthiness is measured against compatibility with democratic values and the rule of law (Council of Europe, 2020). This governance logic reflects the Council's mandate to safeguard human rights through binding legal frameworks.

These divergent operationalization logics reveal that ethical consensus at the normative level does not automatically translate into uniform implementation practices. Instead, shared principles are reshaped by institutional contexts, legal traditions, and governance rationales. As Whittlestone et al. (2024) argue, understanding such diversity is essential for developing coherent global approaches to AI governance. The present study advances this argument by demonstrating that divergence occurs not only at the level of policy priorities but also in the underlying governance logics that determine how ethical principles are translated into practice.

This ethical implementation gap has significant implications. It suggests that efforts to harmonize AI ethics internationally cannot rely solely on articulating shared principles; they must also address the governance mechanisms, enforcement capacities, and institutional pathways through which ethical commitments are realized. The gap between normative aspiration and regulatory reality is not merely a technical or administrative challenge but reflects deeper philosophical divergences about the nature of ethical governance itself, whether ethics should function primarily through voluntary norms (UNESCO), technical standards (European Commission), policy coordination (OECD), binding legislation (European Parliament), or legal rights frameworks (Council of Europe).

The following sections examine how this implementation gap manifests across specific ethical principles and governance domains, providing a more granular analysis of the tensions between shared values and divergent practices in global AI ethics discourse.

Transparency: From Shared Value to Divergent Operationalization

Transparency, identified as a foundational ethical principle across all analyzed reports, is directly linked to the traceability and explainability of algorithmic decisions (Floridi & Cowls, 2022; Lipton, 2018). However, the findings reveal that transparency is operationalized differently depending on institutional governance logics. The European Commission (2019) frames transparency primarily as a technical and regulatory requirement, emphasizing tools that facilitate user understanding of AI systems. This aligns with Lipton (2018) and Doshi-Velez & Kim's (2017) models for explainability, which aim to simplify algorithmic processes for comprehension. The EU approach treats transparency as enforceable through mandatory disclosure obligations and conformity assessments reflecting a governance logic that prioritizes consumer protection through binding standards. UNESCO (2021), in contrast, conceptualizes transparency as

culturally adaptive capacity building. Its proposal for data provenance and ethical labeling systems broadens transparency beyond technical explainability to encompass ethical provenance and cultural sensitivity. O'Neil (2016) supports this view, asserting that transparent data sourcing can reduce societal biases. UNESCO's approach reflects a governance logic emphasizing education and voluntary adoption rather than regulatory enforcement. OECD (2023) treats transparency as an instrument for international policy coordination, advocating for guidelines and best practice sharing rather than binding standards. This positions transparency as a tool for alignment across diverse national contexts, acknowledging that what constitutes "transparent" AI may vary depending on technological capacity and regulatory infrastructure. This reflects the pragmatic challenge of bridging technology-focused approaches (as in China; Ding, 2018) with rights-based transparency principles (as in Europe). However, Crawford (2021) and Zarsky (2016) highlight a fundamental tension: transparency and privacy may conflict, complicating the development of globally harmonized standards. This tension manifests differently across governance contexts. The EU addresses it through legal balancing tests embedded in data protection law; UNESCO emphasizes cultural negotiation; and OECD advocates flexible implementation allowing national variation. These divergent operationalizations reveal that transparency functions not as a universal standard but as a governance-dependent principle shaped by institutional mandates. The ethical implementation gap emerges precisely because organizations endorse transparency while pursuing incompatible mechanisms for its realization, ranging from binding legal obligations to voluntary guidelines to culturally adaptive frameworks.

Responsibility: Multi-Stakeholder Rhetoric vs. Institutional Reality

Responsibility, as Whittlestone et al. (2019) argue, requires a multi-stakeholder approach for effective AI governance. Yet the findings demonstrate substantial variation in how responsibility is distributed, enforced, and justified across institutional contexts. The Council of Europe (2020) and OECD (2023) advocate multi-stakeholder governance, distributing responsibility across developers, regulators, users, and civil society. This aligns with Sharma's (2024) argument that responsibility must extend beyond developers to include all actors in the AI ecosystem. The proposed multi-stakeholder review committees exemplify mechanisms supporting participatory accountability (Diakopoulos, 2016; Wieringa, 2020). However, the European Commission (2019) and European Parliament (2020) adopt a more legally centralized approach, assigning clear responsibility to specific actors with enforceable consequences. This reflects a governance logic prioritizing regulatory clarity over participatory flexibility. While multi-stakeholder consultation is encouraged, ultimate responsibility rests with identifiable legal entities subject to oversight and sanction. UNESCO (2021) distributes responsibility through international cooperation and voluntary norms, reflecting its mandate to accommodate diverse legal and cultural frameworks. This approach acknowledges that responsibility mechanisms must be adapted to local contexts, a view supported by Crawford (2021) and Brynjolfsson & McAfee (2014), who note that economic and cultural differences pose significant implementation challenges. The tension between these approaches is evident in practice. As Ding (2018) observes, centralized governance models (such as China's) distribute responsibility differently than decentralized multi-stakeholder systems (such as the EU's). Siau & Wang (2020) argue for shared global policies, yet the findings reveal that even among democratic governance systems, responsibility is institutionalized through incompatible mechanisms, binding legal liability (EU), voluntary coordination (OECD), or normative aspiration (UNESCO). This divergence reflects a deeper challenge: responsibility without enforcement capacity risks becoming rhetorical, while enforcement without multi-stakeholder legitimacy risks becoming authoritarian. The ethical implementation gap manifests as a tension between participatory ideals and regulatory pragmatism, with no consensus on how to balance accountability with inclusivity.

Trustworthiness: Technical Standards vs. Rights-Based Governance

Trustworthiness, as conceptualized across the reports, encompasses transparency, accountability, security, and societal acceptance (OECD, 2019). Yet the findings reveal

divergent approaches to what constitutes a "trustworthy" AI system. The European Commission (2019) operationalizes trustworthiness through technical robustness and regulatory compliance, associating it with ethical conformity, algorithmic accuracy, and legal adherence. Jobin, Ienca, & Vayena (2019) support this emphasis on measurable criteria such as algorithmic accuracy and data security. This approach reflects a governance logic treating trustworthiness as verifiable technical performance subject to conformity assessment. The Council of Europe (2020), in contrast, grounds trustworthiness in human rights compliance and democratic values. From this perspective, technical reliability is necessary but insufficient; trustworthiness requires alignment with fundamental rights and the rule of law. This reflects a governance logic subordinating technical criteria to normative legitimacy, a system may be technically robust yet untrustworthy if it violates rights or undermines democratic processes. OECD (2023) frames trustworthiness through continuous evaluation and adaptive governance, emphasizing harmonization of international standards and cross-sectoral audits. This reflects a view of trustworthiness as dynamic and context-dependent rather than static compliance with fixed criteria. Whittlestone et al. (2019) support this perspective, emphasizing the necessity of cross-sectoral oversight mechanisms. UNESCO (2021) anchors trustworthiness in cultural sensitivity and inclusive development, treating it as inseparable from respect for diverse value systems. This approach recognizes that what constitutes "trustworthy" AI varies across cultural contexts, a principle supported by international cooperation and capacity building rather than uniform standards. However, Crawford (2021) identifies a persistent challenge: infrastructure deficiencies in developing countries limit the implementation of trustworthiness standards. This highlights the risk that technically sophisticated frameworks developed by organizations like the European Commission may be inapplicable in contexts lacking regulatory capacity or technological infrastructure. The divergence in trustworthiness operationalization reveals a fundamental tension between universal standards (which risk being culturally insensitive or economically inaccessible) and context-adaptive frameworks (which risk fragmenting global AI governance). The ethical implementation gap emerges because organizations pursue incompatible pathways, binding technical standards (EU), rights-based legal frameworks (Council of Europe), voluntary harmonization (OECD), or culturally adaptive norms (UNESCO), while claiming to pursue the same goal of trustworthy AI.

Divergent Governance Strategies: Legal, Technological, and Collaborative Dimensions

Beyond the core ethical principles, the analysis identified three domains where governance strategies diverge substantially: legal regulations, technology-human balance, and international cooperation. Legal regulations reveal a spectrum from universalist to regionalist approaches. UNESCO (2021) advocates global principles accommodating cultural diversity, while the European Commission (2019) and European Parliament (2020) pursue Europe-centered frameworks designed for regional implementation with external alignment capacity. This tension reflects broader debates in AI governance scholarship: as Floridi & Cowls (2019) and Crawford (2021) argue, legal frameworks must adapt swiftly to technological change while respecting regional variation. The challenge, as Ding (2018) observes, is that China's focus on collective benefit conflicts with the EU's rights-based approach, complicating global legal harmonization. OECD's (2023) proposal for multi-stakeholder platforms represents an intermediary strategy, yet Whittlestone et al. (2019) note that translating abstract principles into culturally adaptable legal practices remains contested. The technology-human balance similarly reveals divergent governance philosophies. The European Commission (2019) emphasizes technical reliability and user protection, prioritizing measurable criteria like algorithmic accuracy (Lipton, 2018; Doshi-Velez & Kim, 2017). UNESCO (2021) and the Council of Europe (2020), in contrast, foreground human-centered participatory design, emphasizing human involvement in development processes (Van Otterlo, 2017). OECD (2023) advocates multi-stakeholder collaboration to balance technical and ethical dimensions. Yet as Crawford (2021) observes, technological advancements evolve faster than ethical alignment, creating persistent tensions.

Brynjolfsson and McAfee (2014) argue that human-centered approaches can reduce inequalities, but their effectiveness depends on comprehensive regulation and education, which remain unevenly distributed across the globe. International cooperation exposes competing visions of global governance. UNESCO (2021) pursues bottom-up harmonization respecting cultural diversity, while the Council of Europe (2020) advocates top-down human rights frameworks as universal standards. The European Commission exhibits regional assertiveness, seeking to align international standards with European norms (Floridi & Cowls, 2022). OECD (2023) emphasizes flexible implementation accommodating economic and technological disparities, a pragmatic response to the reality that countries possess vastly different regulatory capacities (Crawford, 2021). Yet as Ding (2018) notes, divergent priorities among China, the United States, and the European Union complicate consensus-building. Whittlestone et al. (2019) argue that effective cooperation requires integration of technical, ethical, and political dimensions, yet the findings reveal little agreement on whether this should occur through binding treaties, voluntary coordination, or cultural negotiation. Collectively, these divergences underscore that the ethical implementation gap extends beyond individual principles to encompass fundamental disagreements about governance architecture, whether AI ethics should be realized through binding law, voluntary norms, technical standards, or rights-based frameworks.

Bridging the Implementation Gap: Recommendations for Policy and Practice

Based on the findings and their interpretation within existing scholarship, this study offers targeted recommendations for developers, practitioners, and policymakers to narrow the ethical implementation gap.

For AI developers and practitioners, the priority must be embedding transparency, accountability, and ethical compliance within organizational routines rather than treating them as external constraints. Recent research emphasizes that ethical assessments become effective when institutionalized through formal governance structures, such as ethics committees, algorithmic impact assessments, cross-functional review boards, and lifecycle-based monitoring tools, rather than remaining loosely defined practices (Corrêa et al., 2023; Batool et al., 2025). Developers should adopt explainability tools that make algorithmic decision-making comprehensible to users, conduct regular ethical assessments to identify and mitigate biases, and establish feedback mechanisms enabling user participation. A human-centered approach requires continuous monitoring of societal impacts, particularly in high-stakes domains like healthcare, education, and employment where AI decisions directly affect individual rights. Multi-stakeholder collaboration involving academics, technology users, and civil society should be integrated into development workflows. Capacity-building programs are essential to equip practitioners with both technical skills and ethical literacy, enabling them to navigate the tension between innovation and responsibility.

For policymakers, the challenge is designing regulatory frameworks that balance enforceability with flexibility. The European Union's AI Act (2024) illustrates the shift from voluntary guidelines to binding, risk-based governance with concrete responsibilities and sanctioning mechanisms (Cancela-Outeda, 2024). Yet as recent scholarship emphasizes, effective regulation requires embedding ethical principles within institutional structures capable of sustaining implementation across the AI lifecycle (Batool et al., 2025; Batool et al., 2026). Policymakers should establish comprehensive data protection laws, create independent oversight mechanisms such as ethical review committees, and invest in technological infrastructure, particularly in developing countries where capacity gaps hinder implementation. International cooperation mechanisms must be strengthened, recognizing that globally harmonized AI ethics requires accommodating cultural, economic, and political diversity. This demands flexible norms adaptable to local contexts while maintaining core protections. Public awareness campaigns and educational initiatives are crucial for building societal trust and ensuring that citizens understand both the benefits and risks of AI technologies. Ultimately, policymakers must recognize that ethical AI governance is not a one-time regulatory achievement but an ongoing process requiring adaptive legal frameworks responsive to technological evolution.

For organizations and institutions, translating ethical principles into practice requires concrete operational mechanisms that bridge the gap between normative commitments and day-to-day operations. Internal AI governance structures, such as ethics committees, algorithmic impact assessment protocols, and cross-functional review boards, enable organizations to embed ethical considerations into routine decision-making processes (Corrêa et al., 2023; Batool et al., 2025). For instance, establishing regular ethics audits can ensure that deployed AI systems continue to meet fairness and transparency standards as they evolve, while incident reporting mechanisms create feedback loops that enable organizations to identify and address ethical failures proactively. Sector-specific guidelines tailored to the unique ethical challenges of domains such as healthcare, finance, or education can operationalize abstract principles within contextual constraints. Professional associations and regulatory bodies play a critical role in developing these tailored frameworks, which translate universal commitments into actionable practices suited to specific risk profiles, technical capabilities, and stakeholder needs. For example, healthcare AI ethics guidelines must account for patient safety, clinical accountability, and informed consent in ways that differ substantially from requirements in automated hiring or credit scoring systems. Furthermore, continuous monitoring and feedback mechanisms are essential for adaptive governance. Organizations should implement systems that track not only technical performance metrics (e.g., accuracy, efficiency) but also ethical outcomes, such as fairness indicators across demographic groups, user satisfaction with transparency measures, and rates of successful appeals when algorithmic decisions cause harm. These mechanisms enable iterative improvement and ensure that ethical governance remains responsive to technological change, emerging risks, and evolving societal expectations. Without such operational infrastructure, even well-intentioned ethical commitments risk remaining aspirational rather than enforceable.

These recommendations reflect the central insight of this study: closing the ethical implementation gap requires aligning normative principles with governance mechanisms, enforcement capacities, and institutional pathways capable of translating ethical aspirations into enforceable practices.

Contributions, Limitations, and Future Research Directions

This study makes both theoretical and practical contributions while acknowledging certain limitations that open pathways for future research.

Theoretically, the study advances AI ethics scholarship by moving beyond principle identification to examine the governance logics through which ethical commitments are institutionalized. While existing literature documents normative convergence around transparency, accountability, and fairness (Jobin, Ienca, & Vayena, 2019; Corrêa et al., 2023), this study demonstrates that shared principles are reshaped by institutional mandates, legal traditions, and governance philosophies, producing an ethical implementation gap between normative aspiration and regulatory reality. By comparatively analyzing how UNESCO, the European Commission, OECD, the European Parliament, and the Council of Europe operationalize identical principles through divergent mechanisms, the study clarifies the conceptual links between ethical guidance documents and their subsequent regulatory enactment. This contributes to ongoing debates about whether global AI governance should prioritize uniformity or flexibility, binding law or voluntary norms, technical standards or rights-based frameworks.

Practically, the study provides actionable insights for policymakers and practitioners. The identification of divergent governance strategies, ranging from soft norms to binding regulations, from technical assurance to participatory design, from cultural pluralism to rights-based universalism, enables stakeholders to make informed choices about which governance models best suit their institutional contexts. The recommendations emphasize that effective AI ethics requires not merely endorsing principles but establishing enforcement mechanisms, accountability structures, and organizational routines capable of sustaining implementation.

However, the study has *limitations* that must be acknowledged. First, the analysis was restricted to five international reports, shaped primarily by European

and global intergovernmental perspectives. National-level policies, regional frameworks outside Europe, and sector-specific guidelines fall outside the scope, limiting generalizability to diverse contexts. Second, the focus on institutional documents excludes the perspectives of individual actors, developers, users, and civil society organizations, whose experiences with AI ethics implementation may reveal challenges not evident in policy texts. Third, the rapidly evolving nature of AI technologies and regulatory frameworks means that findings represent a snapshot of a formative period (2019-2023) rather than a comprehensive account of contemporary AI governance, which now includes binding regulations such as the EU AI Act (2024). Fourth, the comparative document analysis method, while effective for identifying patterns across reports, does not capture implementation effectiveness or compliance outcomes, dimensions that require empirical case studies or longitudinal research.

Future research should address these limitations through several pathways. First, expanding the analysis to include national AI strategies, particularly from countries in Asia, Africa, and Latin America, would provide a more globally representative understanding of AI ethics governance. Second, empirical studies examining how individual actors interpret and apply ethical principles in practice would complement this document-centered analysis. For instance, qualitative research with AI developers could explore how transparency requirements are operationalized during algorithm design and testing, investigating questions such as: Are explainability tools used in development workflows? How are algorithmic decisions documented? What trade-offs do developers perceive between transparency and performance? Interviews with compliance officers and auditors could reveal organizational barriers to implementing accountability mechanisms, examining challenges such as resource constraints, conflicting regulatory requirements across jurisdictions, or insufficient technical expertise within oversight teams. Surveys with end users could assess whether deployed AI systems meet ethical expectations regarding fairness, explainability, and privacy protection, measuring user perceptions of algorithmic transparency, trust in AI-driven decisions, and satisfaction with recourse mechanisms when harms occur. Such research would illuminate not only what principles organizations endorse but how these principles are, or are not, realized in practice. Third, longitudinal research tracking how ethical principles evolve from voluntary guidelines to binding regulations, and assessing compliance and enforcement outcomes, would clarify the conditions under which ethical implementation gaps narrow or persist. Fourth, empirical case studies examining specific AI applications (e.g., facial recognition, predictive policing, automated hiring) would reveal how abstract principles are translated into concrete practices across diverse sectors. Finally, research analyzing the effectiveness of international cooperation mechanisms, such as multi-stakeholder platforms, bilateral agreements, and treaty-based governance, could identify pathways toward more coherent global AI governance.

In conclusion, this study demonstrates that the challenge facing global AI ethics is not a lack of shared principles but the absence of consensus on how these principles should be institutionalized. By analytically unpacking the ethical implementation gap, the study provides a foundation for future research and policy efforts aimed at aligning normative commitments with governance realities.

CONCLUSION

This study has examined the common themes, divergent approaches, and unique contributions of five major international AI ethics reports, revealing both normative convergence and operational divergence in global AI governance. While organizations converge around transparency, responsibility, and trustworthiness as foundational ethical principles, they operationalize these principles through incompatible governance mechanisms, ranging from voluntary norms and policy coordination to binding legal frameworks and rights-based jurisprudence.

The central finding is *the existence of an ethical implementation gap*: widely endorsed principles do not automatically translate into uniform practices because they are reshaped by institutional mandates, legal traditions, and governance philosophies. UNESCO pursues culturally adaptive norms through education and capacity building; the European Commission and European Parliament enforce compliance through binding regulation; OECD

coordinates flexible implementation across economically diverse states; and the Council of Europe anchors AI ethics within human rights law. These divergent governance logics reflect deeper philosophical disagreements about whether ethical AI should be realized through soft norms or hard law, technical standards or participatory design, universal frameworks or context-sensitive adaptation.

The adoption of the EU AI Act (2024) and the evolution of AI ethics from aspirational guidelines to enforceable regulations underscore the formative role of the documents analyzed in this study. Rather than being rendered obsolete by subsequent developments, these frameworks established the normative foundation upon which contemporary binding regulations are built. Understanding their governance logics clarifies the pathways and obstacles to translating ethical principles into practice.

The study's recommendations emphasize that closing the implementation gap requires aligning ethical principles with institutional capacity, enforcement mechanisms, and governance structures capable of sustaining implementation. For developers, this means embedding ethics within organizational routines; for policymakers, it requires designing adaptive legal frameworks responsive to technological change while respecting cultural diversity. Ultimately, the future of AI ethics depends not on articulating new principles but on building governance architectures capable of realizing existing commitments. Collaboration among international organizations, national governments, developers, and civil society will be essential to ensure that AI technologies develop in ways that are transparent, accountable, and aligned with human rights and democratic values. By illuminating the gap between ethical aspiration and regulatory reality, this study contributes to ongoing efforts to ensure that AI serves humanity equitably and responsibly.

REFERENCES

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy, 128*(6), 2188-2244. https://doi.org/10.1086/705716

Allen, C., Wallach, W., & Smit, I. (2006). Why machine ethics? *IEEE Intelligent Systems*, *21*(4), 12-17. https://doi.org/10.1109/MIS.2006.83

Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine, 28*(4), 15-26. https://doi.org/10.1609/aimag.v28i4.2065

Batool, A., Lee, S., Liu, Y., & Dong, L. (2026). The anatomy of AI policies: a systematic comparative analysis of AI policies across the globe. *AI and Ethics 6*, 55. https://doi.org/10.1007/s43681-025-00886-3

Batool, A., Zowghi, D. & Bano, M. (2025). AI governance: a systematic literature review. *AI and Ethics 5*, 3265-3279. https://doi.org/10.1007/s43681-024-00653-w

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Proceedings of the Conference on Fairness, Accountability and Transparency (FAT*)* (pp. 149–159). PMLR.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27-40. https://doi.org/10.3316/QRJ0902027

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the Conference on Fairness, Accountability and Transparency (FAT*)* (pp. 77–91). PMLR.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, *4*(10), 100857. https://doi.org/10.1016/j.patter.2023.100857

Corrêa, N. K., Santos, J. W., Galvão, C., Pasetti, M., Schiavon, D., Naqvi, F., Hossain, R., & de Oliveira, N. (2025). Crossing the principle-practice gap in AI ethics with ethical problem-solving. *AI and Ethics 5,* 1271-1288. https://doi.org/10.1007/s43681-024-00469-8

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press. https://doi.org/10.12987/9780300252392

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM, 59*(2), 56-62. https://doi.org/10.1145/2844110

Ding, J. (2018). *Deciphering China's AI dream*. Future of Humanity Institute, University of Oxford. https://cdn.governance.ai/Deciphering_Chinas_AI-Dream.pdf

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. https://doi.org/10.48550/arXiv.1702.08608

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, *4*(1), 5. https://doi.org/10.1038/s41746-020-00376-2

European Commission (2019). *Ethics Guidelines for Trustworthy AI.* https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Text with EEA relevance). *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Fox-Skelly, J., Bird, E., & Jenner, N. (2020). *The ethics of artificial intelligence: Issues and initiatives*. European Parliament, Directorate-General for Parliamentary Research Services. https://data.europa.eu/doi/10.2861/6644

European Parliamentary Research Service. (2024). *AI investment: EU and global indicators*. European Parliament. https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/760392/EPRS_ATA(2024)760392_EN.pdf

European Parliamentary Research Service. (2023). China-US global rivalry and the EU (Briefing 749803). European Parliament. https://www.

europarl.europa.eu/RegData/etudes/BRIE/2023/749803/EPRS_BRI(2023)749803_EN.pdf

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In S. Carta (Ed.), *Machine Learning and the City: Applications in Architecture and Urban Design* (pp. 535-545). Wiley. https://doi.org/10.1002/9781119815075.ch45

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2

Israel, I. B., Cerdio, J., Ema, A., Friedman, L., Ienca, M., Mantelero, A., & Matania, E. (2020). *Towards regulation of AI systems: Global perspectives on the development of a legal framework on artificial intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law*. Council of Europe.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *16*(3), 31-57. https://doi.org/10.1145/3236386.3241340

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016*). Intelligence Unleashed: An Argument for AI in Education*. Pearson.

McKinsey Global Institute. (2023). *The economic impact of artificial intelligence*. McKinsey & Company. https://www.mckinsey.com

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Milakis, D., Van Arem, B., & Van Wee, B. (2017). Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of Intelligent Transportation Systems, 21*(4), 324-348. https://doi.org/10.1080/15472450.2017.1291351

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 2053951716679679. https://doi.org/10.1177/2053951716679679

OECD. (2019). *Recommendation of the council on artificial intelligence*. OECD.

OECD. (2023). *The state of implementation of the OECD AI Principles four years on.* https://www.oecd.org/en/publications/the-state-of-implementation-of-the-oecd-ai-principles-four-years-on_835641c9-en.html

O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2023). Toward AI governance: Identifying best practices and potential barriers and outcomes. *Information Systems Frontiers, 25*(1), 123–141. https://doi.org/10.1007/s10796-022-10251-y

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems, 34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

PricewaterhouseCoopers (PwC). (2017). *Sizing the prize: What's the real value of AI for your business and how can you capitalise?* https://www.pwc.com.au/government/pwc-ai-analysis-sizing-the-prize-report.pdf

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Sharma, S. (2024). Benefits or concerns of AI: A multistakeholder responsibility. *Futures*, 157, 103328. https://doi.org/10.1016/j.futures.2024.103328

Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management, 31*(2), 74-87. https://doi.org/10.4018/JDM.2020040105

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science, 361*(6404), 751-752. https://doi.org/10.1126/science.aat5991

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence.* https://unesdoc.unesco.org/ark:/48223/pf0000381137

Van Otterlo, M. (2017). From algorithmic black boxes to adaptive white boxes: Declarative decision-theoretic ethical programs as codes of ethics. *arXiv*. https://doi.org/10.48550/arXiv.1711.06035

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019, January). The role and limits of principles in AI ethics: Towards a focus on tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 195-200).

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1-18. https://

doi.org/10.1145/3351095.3372833

Xiong, H., Ledwidge, M. T., Fadahunsi, K. P., Lee, H. Y., Wu, J., Morrow, S., Nisar, Y. B., Mbakaya, B., O'Donoghue, J., & Gallagher, J. (2025). Global Artificial Intelligence (AI) Governance, Trust, and Ethics for Sustainable Health (GATES): A Protocol for Methodological Framework. *VeriXiv, 2,* 187. https://doi.org/10.12688/verixiv.1380.1

Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, *41*(1), 118-132. https://doi.org/10.1177/01622439156057

Author Contributions

All authors have contributed equally to this work.