Int. J. Soc. & Nat. Sci. Vol. II (I) 30-01, 2026
Journal home page: http://ijsns.com.pk/
ISSN 3079-286X (PRINT)
ISSN 3104-4123 (ONLINE)

An Exploratory Document Analysis of Ethical Considerations in AI Policies and Guidelines: A Critical Review

Tehseen Nuzhat[1*], Maryam Behram[2]
Northern University, Nowshera
*Corresponding author: majtn2006@gmail .com

Accepted: *15 August, 2025* ***Published:*** *30 January,2026*

ABSTRACT

The rapid adoption of artificial intelligence (AI) across sectors such as healthcare, finance, law enforcement and education have increased the need for effective and ethically grounded governance frameworks. This study employs an exploratory qualitative document analysis to examine ethical principles embedded in 25 national and international AI governance documents published between 2020 and 2025. Guided by principle-based ethics, socio- technical systems theory and critical discourse analysis, this study investigates how ethical principles are articulated and operationalized across governance frameworks.

The findings reveal broad convergence around core ethical principles, particularly transparency, fairness, accountability, privacy and human oversight. However, significant differences exist in their interpretation and implementation. International frameworks predominantly rely on normative and aspirational language, whereas regulatory instruments increasingly embed ethical principles within enforceable compliance mechanisms. The analysis also identifies persistent challenges including ambiguous accountability structures, limited operational guidance, insufficient attention to marginalized populations and uneven consideration of emerging AI technologies. These findings highlight a gap between ethical commitments and their practical implementation across governance context.

The study contributes to responsible AI scholarship by highlighting the need for more accountable, inclusive and context sensitive governance approaches capable of translating ethical principles into practice.

Keywords: Artificial Intelligence, Ethics, Governance, Transparency, Fairness, Accountability, Inclusivity.

INTRODUCTION

Artificial intelligence (AI) has become one of the most influential and rapidly advancing technologies of the twenty-first century, reshaping decision making, organizational processes and public services across diverse sectors including healthcare, education, finance, security and governance. Its capacity to analyze large datasets automate complex tasks and generate predictive insights. It has led governments and institutions worldwide to adopt AI strategies and policies at an unprecedented pace. This rapid diffusion underscores AI's potential to enhance efficiency, productivity and economic growth but it also raises pressing ethical questions about how AI systems align with human values and democratic norms.

The importance of ethical scrutiny in AI is widely recognized among scholars and policymakers. Researchers have repeatedly documented how algorithmic models can reproduce social biases embedded in training data, leading to discriminatory outcomes in critical applications such as hiring, credit allocation and criminal justice thereby magnifying existing inequalities (Jobin et al., 2019; Tabbassum & Chintale, 2024). Likewise, the opaque nature of many machine learning models often described as “black boxes” complicates efforts to understand or contest decisions made by AI systems creating barriers to accountability and trust (Organization for Economic Co-operation and Development, (OECD), n.d.; UNESCO, 2022). Privacy concerns have also come to the forefront as AI systems rely on extensive data

collection and processing which can infringe on personal autonomy and data protection rights when regulatory safeguards are inadequate (Barthwal et al., 2025). Collectively these ethical concerns demonstrate that the development and application of AI cannot be disentangled from broader questions about fairness, rights and societal wellbeing.

In response to these challenges national and international bodies have proposed a range of governance frameworks and ethical guidelines to promote responsible AI development. UNESCO's recommendation on the Ethics of Artificial Intelligence provides a comprehensive value-based discursive silences and ambiguities in governance documents that may hinder the equitable context-sensitive development and deployment of AI systems.

Given these limitations there is a clear need for critical systematic examination of how ethical principles are articulated, operationalized and prioritized within AI policies and guidelines. By focusing on national and international governance documents this study aims to map the ethical principles foregrounded in these texts' framework emphasizing human rights, inclusiveness, transparency and accountability as core principles for ethical AI governance (UNESCO, 2022). Similar efforts by the European Union and Organization for Economic Co-operation and Development (OECD) articulate principles aimed at guiding policymakers and developers towards responsible practices including provisions for transparency, fairness and human oversight (OECD,n.d.). However, despite the proliferation of ethical guidelines recent research reveals that such frameworks often remain high level and normative, lacking clear mechanisms for implementation measurable enforcement strategies or contextual adaptability across sectors and jurisdictions (Batool et al., 2025; Corrêa et al., 2022). A systematic literature review of AI governance studies found that while multiple governance solutions and ethical frameworks have been proposed many prioritize compliance and technical risk mitigation with ethical concerns like fairness, transparency and inclusiveness receiving inconsistent attention and practical operationalization (Batool et al., 2025).

Furthermore, emerging AI technologies such as generative models, autonomous systems and large-scale machine learning introduce ethical risks that existing guidelines have yet to address comprehensively. Studies examining practitioner perspectives report that issues such as accountability, privacy protection and transparency are often difficult to implement in real-world settings underscoring gaps between high level ethical principles and practice (Sanderson et al., 2024). Cross national analyses also highlight disparities in regulatory readiness and institutional capacity with developed countries adopting hybrid governance approaches that blend hard and soft laws while many developing regions struggle with infrastructural and policy constraints (Rumlus, 2025). These inconsistencies point to analyze how ethical considerations are framed and implemented and identify gaps inconsistencies and discursive silences that may constrain responsible AI development. Drawing on principle-based ethics, socio-technical perspectives and critical discourse analysis this exploratory document analysis seeks to contribute to both academic scholarship and policy practice by offering actionable insights that can inform future iterations of AI governance frameworks strengthen accountability mechanisms, reduce bias, enhance transparency and promote human centered AI systems aligned with societal values and fundamental rights.

Although previous studies have reviewed AI ethics principles and governance frameworks much of the existing literature focuses on identifying common ethical themes rather than examining how these principles are operationalized across policy contexts. Limited attention has been given to the gap between ethical commitments and their practical implementation within governance documents. This study addresses that gap through a comparative and critical analysis of national and international AI policies, focusing on how

ethical principles are framed, operationalized and translated into governance practice.

OBJECTIVES OF THE STUDY

1. To identify and map the ethical principles foregrounded in national and international AI governance documents.
2. To analyze how ethical considerations are framed, articulated and operationalized across the selected documents.
3. To examine inconsistencies, ambiguities and discursive silences regarding ethical commitments within these documents.
4. To identify gaps and limitations in existing ethical frameworks particularly concerning emerging AI risks and socio-technical challenges.

RESEARCH QUESTIONS

1. Which ethical principles are emphasized in contemporary AI policies and governance guidelines and how are they conceptually framed?
2. How do different AI governance documents construct, prioritize and operationalize ethical considerations and to what extent are there consistency across them?
3. What gaps ambiguities or discursive silences exist within these documents that may constrain responsible equitable and context-sensitive AI development and deployment?

Original Contribution

This study advances AI governance scholarship by moving beyond descriptive summaries of ethical principles to critically interrogate why AI policies are framed in particular ways and whose interests those framings serve. While existing literature frequently reiterates conclusions such as the dominance of high-level ethical principles weak operationalization and the prevalence of ethics washing it rarely examines the discursive and institutional forces that produce these patterns. Using a critical discourse analysis (CDA) approach this research reveals how policy language is strategically constructed through framing choices modal verbs, silences and normative assumptions. A key contribution lies in the comparative analytical contrast between major global governance actors. For example, the European Union predominantly frames AI ethics through legally enforceable and compliance-oriented discourse whereas UNESCO emphasizes moral universalism and voluntary ethical guidance. These contrasting framings reflect underlying power structures, institutional mandates and political priorities rather than neutral ethical reasoning.

The study makes four original contributions. First, it develops a policy to practice translation gap model that explains why ethical commitments articulated in governance documents often fail to translate into concrete implementation mechanisms. Second it introduces a discursive lens on silences demonstrating how omissions related to enforcement, accountability and Global South contexts limit the practical reach of AI ethics frameworks. Third it offers a comparative governance perspective systematically contrasting legalistic and moralized approaches to AI ethics. Fourth it advances a Global South informed critique highlighting how dominant governance narratives marginalize contextual realities of resource constrained and postcolonial settings.

Together these contributions foreground the political and institutional dimensions of AI ethics governance and provide a theoretically grounded and empirically informed framework for bridging the persistent gap between ethical aspiration and operational practice

Review of Related Literature

Artificial Intelligence (AI) has become a transformative technology across multiple sectors including healthcare, education, finance, security and public governance. The rapid adoption of AI has introduced profound ethical concerns as AI systems increasingly influence high-stakes decision making processes that directly affect individuals and societies. Foundational research

emphasizes that AI systems must adhere to ethical principles such as transparency, accountability, fairness, privacy, human oversight and inclusiveness to ensure responsible deployment (Floridi et al., 2020; Jobin et al., 2019). Global governance initiatives have sought to codify these principles into frameworks to guide developers, policymakers and stakeholders. UNESCO's Recommendation on the Ethics of Artificial Intelligence emphasizes a human-centered approach focusing on inclusivity, equity and sustainable development (UNESCO, 2022) while the OECD AI Principles advocate for transparency, robustness, human oversight and responsible data use (OECD, n.d.). The European Union's AI Act further institutionalizes these principles by introducing enforceable obligations for high-risk AI systems (European Commission, 2021). National strategies in countries such as the United States, China, Singapore and Canada reflect a combination of local policy priorities, cultural contexts and institutional capacities (Fjeld et al., 2020; Corrêa et al., 2022). Comparative analyses indicate convergence in the prioritization of fairness, transparency and accountability but reveal significant divergences in operationalization, monitoring and enforcement mechanisms (Raji et al., 2021; Batool et al., 2025)

While these studies demonstrate broad agreement making where decisions carry significant regarding the importance of ethical principles, they differ considerably in their approaches to implementation and governance. International frameworks such as UNESCO and OECD emphasize voluntary guidance and ethical aspirations whereas regulatory initiatives such as the European Union AI Act adopt legally enforceable requirements. These differences suggest that consensus at the conceptual level does not necessarily translate into consistent operational practices across jurisdictions.

Despite the proliferation of AI ethics guidelines operational challenges remain pervasive. While high level principles are widely articulated practical mechanisms for implementing these principles are frequently missing (Sanderson et al., 2024). Practitioners often struggle to translate ethical concepts into organizational workflows particularly in domains such as healthcare diagnostics predictive policing and financial decision societal impact (Obermeyer et al., 2019; Jiao et al., 2025). Socio technical factors including organizational culture, technological infrastructure, regulatory maturity and societal norms significantly influence ethical AI implementation but are often underexplored in governance documents (McGregor, Murray, & Ng, 2020). The phenomenon of "ethics washing" where organizations rhetorically commit to principles without concrete operationalization demonstrates a persistent gap between normative frameworks and practical application (McGregor et al., 2020).

Emerging AI technologies present additional ethical complexities. Generative AI systems can propagate misinformation infringe intellectual property and affect societal trust (McCormack et al., 2025), while autonomous systems raise critical questions about accountability, liability and human oversight (Raji et al., 2021). Large scale machine learning applications often involving sensitive personal data amplify risks of privacy violations, algorithmic bias and inequitable outcomes (Barthwal et al., 2025: Mbabazi et al., 2023). Empirical evidence shows that marginalized populations are particularly vulnerable to algorithmic biases as seen in healthcare predictive tools (Obermeyer et al., 2019) highlighting the need for inclusive and socially aware AI governance. Despite extensive attention to principles like fairness and transparency, discursive silences persist in policy documents regarding distributive justice, environmental impact and sector specific operationalization (Tabbassum & Chintale, 2024; Raji et al., 2021). These gaps underscore the importance of systematic critical analyses to examine both the articulation of principles and their practical enforceability across diverse contexts.

Although existing research has extensively examined AI ethics principles and governance frameworks, several

gaps remain. Much of the literature focuses on identifying ethical values such as fairness, transparency and accountability or evaluating individual policy initiatives within specific jurisdictions. Comparatively fewer studies systematically examine how these principles are operationalized across multiple national and international governance documents limited attention has been given to identifying inconsistencies, implementation challenges and discursive silences within AI governance frameworks through a unified analytical approach. This gap restricts understanding of whether current governance documents provide sufficient practical guidance for responsible AI deployment across diverse contexts.

To address these identified gaps the present study conducts an exploratory document analysis of national and international AI governance frameworks. By mapping ethical principles identifying operational gaps and inconsistencies and examining discursive silences this research bridges normative AI ethics with operational realities. It synthesizes existing literature to highlight both achievements and limitations of current governance frameworks and provides actionable insights for policymakers, developers and organizations seeking to implement responsible, equitable and context-sensitive AI systems. Through this approach the study contributes academically by extending the knowledge base on AI ethics and practically by informing strategies for more robust inclusive and accountable AI deployment.

THEORETICAL FRAMEWORK

The present study is guided by a combination of principle-based ethics, socio-technical systems theory and critical discourse analysis, which together provide a robust lens for analyze AI governance documents. Principle based ethics, rooted in bioethics and philosophical traditions, emphasizes core ethical principles such as autonomy, justice, beneficence, non-maleficence and fairness which are directly relevant to AI systems (Floridi et al., 2020; Jobin et al., 2019). In the context of AI this framework allows the systematic identification of ethical principles articulated in national and international guidelines and provides a normative benchmark against which operationalization and implementation can be evaluated. By employing principle-based ethics this study can assess not only whether documents mention core values but also how consistently these principles are prioritized, interpreted and applied across diverse governance frameworks.

Complementing this socio-technical systems theory provides an analytical perspective for understanding AI deployment within organizational, societal, and technological contexts (Baxter & Sommerville, 2011). This theory posits that technology and social systems are deeply intertwined and ethical compliance in AI cannot be fully understood without considering organizational structures, human interactions regulatory environments and cultural factors (Jiao et al., 2025). By integrating socio-technical systems theory the study recognizes that governance documents are not implemented in isolation their effectiveness depends on infrastructural readiness sector specific workflows and societal norms. This theoretical lens thus enables the identification of practical constraints, challenges and enablers in translating high-level ethical principles into operational practices. Finally critical discourse analysis (CDA) is employed to examine the ways in which language AI governance documents constructs frames and sometimes silences ethical considerations (Fairclough, 2013; Tabbassum & Chintale 2024). CDA allows the study to uncover inconsistencies ambiguities and normative biases embedded in policy texts revealing how ethical commitments are communicated, prioritized or neglected. This approach is particularly useful for highlighting discursive silences regarding emerging AI risks underrepresented populations and socio-technical challenges which may not be explicitly addressed in formal guidelines. By integrating CDA with principle-based ethics and socio-technical systems theory the study establishes a triangulated framework that captures both normative ideals and practical realities providing a comprehensive lens for evaluating AI governance documents.

Together, these theories guide the research by allowing a holistic analysis that is both normative and contextual. Principle-based ethics identifies "what should be done" socio-technical systems theory assesses "how it can be operationalized in context" and critical discourse analysis reveals "how the principles are communicated, framed or omitted" in AI governance documents. This combined theoretical framework ensures that the study systematically maps ethical principles identifies operational gaps and interprets discursive silences thereby bridging the gap between normative ethical commitments and practical implementation challenges in AI governance.

METHODOLOGY

Study Design

This study adopted an exploratory qualitative design specifically a document analysis approach to critically examine ethical considerations in national and international AI governance documents. This analysis is recognized qualitative method that permits systematic assessment of text to uncover meanings, themes and contextual patterns embedded in policy artifacts (Bowen, 2009; Prior, 2008). Research analysis allows the researcher to trace how ethical principles are articulated interpreted, prioritized and operationalized across different governance frameworks offering insight into both normative commitments and implementation strategies (Bryman, 2016). Because AI governance frameworks vary significantly in structure and content an exploratory design is appropriate to enable flexible, open-ended inquiry without predefined hypotheses (Creswell & Poth, 2018). The combination of document analysis with triangulated theoretical lenses principle-based ethics, socio-technical systems theory and critical discourse analysis enhances analytical depth by integrating normative evaluation with contextual and linguistic analysis (Fairclough, 2013; Floridi et al., 2020; Baxter & Sommerville, 2011).

Population

The population of this study consisted of national and international policy documents, strategies, regulations and ethical guidelines published between 2018 and 2025 that offer guidance on AI ethics and governance. This period was selected because ethical AI discourse and formal governance initiatives rapidly expanded during these years with foundational frameworks emerging post 2018 (Jobin et al., 2019). The population included documents from key international organizations such as UNESCO, the OECD, the European Union and national governments with advanced AI strategies (e.g., United States, China, Singapore, Canada). Documents were included if they (a) explicitly address ethical considerations in AI development or deployment (b) articulate principles, values or guidelines related to AI technology (c) are accessible in English. This population ensured a comprehensive globally representative set of policy texts for analysis reflecting diverse socio-political contexts and governance philosophies.

Sample

A purposive sampling strategy was employed to select the most relevant and credible documents from the defined population. Purposive sampling is suitable for qualitative inquiry when the researcher seeks information rich sources that directly address the research problem (Palinkas et al., 2015; Etikan et al., 2016). The final sample comprised 25 documents selected based on the following criteria:

1. Relevance: Document explicitly addresses AI ethics or governance principles.
2. Authority: Issued by recognized governmental or intergovernmental institutions.
3. Regency: Published between 2018–2025 to reflect contemporary ethical discourse.
4. Language: Available in English to ensure consistent analytical interpretation.

The sample included influential international recommendations (e.g., UNESCO's Recommendation on the Ethics of AI) OECD principles, regional regulatory proposals (e.g., the EU AI Act) and national

AI strategies with ethics components. This purposive selection enabled cross contextual comparison and ensured that the analysis captured both normative frameworks and jurisdictional variations.

Inclusion and Exclusion Criteria

To ensure consistency and analytical relevance explicit inclusion and exclusion criteria were applied during document selection. Documents were included if they:

(a) Focused primarily on AI ethics, governance and regulation (b) policy was published between 2018 and 2025 (c) Originated from recognized governmental, intergovernmental or international organizations (d) contained explicit discussion of ethical principles, governance mechanisms or implementation guidance were available in English.

Documents were excluded if they

(a) Focused exclusively on technical AI development without governance or ethical considerations (b) represented duplicate versions of previously published frameworks (c) consisted of news articles, opinion pieces or non-official commentaries (d) lacked sufficient substantive content related to ethical governance. Applying these criteria enhanced transparency, reduced selection bias and ensured that all analyzed documents contributed meaningfully to the study's research objectives.

Instruments

The primary data collection instruments used in this study was (a) Document Retrieval Log and (b) Structured Coding Matrix.

Document Retrieval Log: This instrument recorded metadata for each document including authoring body, year of publication, title, type (policy/guideline/ strategy), geographic scope and key thematic keywords (e.g., transparency, accountability, fairness). This log ensured transparent tracking of all sampled documents and facilitated systematic comparison. Similar document cataloging techniques are recommended in policy analysis literature to ensure traceability and reproducibility (Bowen, 2009; Hsieh & Shannon, 2005).

Structured Coding Matrix: Based on the theoretical framework a coding matrix was developed to systematically extract and categorize content related to:

1. Ethical principles (e.g., fairness, transparency, accountability)
2. Operational implementation strategies
3. References to socio-technical factors
4. Emergent AI risks
5. Sectoral or contextual mentions
6. Discursive gaps and silences

Each coding category contained operational definitions and decision rules to ensure consistent interpretation across documents. Deductive codes were derived from the theoretical framework and existing AI ethics literature, including transparency, fairness, accountability, privacy and human oversight. Inductive codes were generated during document review to capture emerging themes, implementation challenges and contextual issues not fully represented in the initial framework. The coding framework was refined iteratively throughout the analysis to accommodate new patterns while maintaining alignment with the study objectives.

The matrix was designed in alignment with qualitative content analysis protocols (Hsieh & Shannon, 2005) and tailored for document comparison. NVivo software was used to facilitate efficient coding, query building and visualization of patterns, enabling methodological rigor and organization of large volumes of qualitative text.

Procedures

The document analysis procedure involved multiple stages to ensure depth, consistency and analytical integrity:

Stage 1 Document Retrieval: All documents were systematically gathered from official sources (government portals, organizational publications, regulatory repositories). Official retrieval minimized the risk of using outdated or unofficial versions.

Stage 2 Preliminary Review, each document was preliminarily read to validate its relevance and assign initial metadata entries. This stage ensured that only documents meeting inclusion criteria entered full analysis.

Stage 3 Coding and Thematic Extraction: Using the Structured Coding Matrix and NVivo, texts were coded inductively and deductively. Deductive codes informed by principle-based ethics and socio-technical theory provided conceptual anchors (e.g., fairness, privacy). Inductive coding allowed emergent themes to surface organically from the text (Saldaña, 2021).

Stage 4 Content Analysis, following coding, documents were analyzed to identify patterns in ethical principles, implementation strategies and contextual emphases. Content analysis allowed quantification of thematic prevalence while retaining qualitative nuance (Krippendorff, 2018).

For the frequency analysis presented in Table 1 each ethical principle was counted once per document when it was explicitly identified within the text, regardless of the number of times it appeared in that document. This approach was adopted to measure the prevalence of ethical themes across governance documents rather than the repetition of terms within individual texts. The resulting frequencies therefore represent the number of documents referencing a particular ethical principle and provide an indication of its relative prominence within the sampled AI governance frameworks.

Stage 5 Critical Discourse Analysis (CDA): was applied to explore how language constructs, frames or omits ethical commitments. This stage focused on identifying discursive silences ethical areas not overtly discussed and language that privileges particular governance narratives (Fairclough, 2013; Wodak & Meyer, 2016).

During this phase attention was directed toward lexical choices, modal expressions, institutional framing and patterns of inclusion and omission within policy texts. Particular emphasis was placed on how ethical responsibilities were assigned which stakeholders were positioned as decision-makers and which social groups or concerns received limited visibility. This analytical process enabled examination of underlying power relations and governance assumptions embedded within AI policy discourse rather than focusing solely on explicitly stated ethical principles.

Stage 6 Triangulation, Findings from principle-based coding, socio-technical interpretation and CDA were triangulated to ensure credibility and comprehensiveness (Denzin, 2017).

Trustworthiness and Research Rigor

Several strategies were employed to enhance the trustworthiness of the study. Credibility was strengthened through triangulation of three analytical lenses principle-based ethics socio-technical systems theory and critical discourse analysis. Dependability was supported through systematic documentation of coding procedures, analytical decisions and data management processes within NVivo. Confirmability was enhanced by maintaining an audit trail that recorded coding refinements and interpretive decisions throughout the analysis. Transferability was facilitated through detailed description of document characteristics governance contexts and analytical procedures enabling readers to assess the applicability of findings to other AI governance settings. Collectively these measures increased methodological transparency and analytical rigor.

Coding consistency was strengthened through the use of a structured coding matrix containing predefined categories and operational definitions derived from the study's theoretical framework. The coding process followed an iterative approach in which coded segments, category definitions and thematic interpretations were reviewed and refined throughout the analysis. Detailed records of coding decisions and category development were maintained to support transparency, consistency and analytical rigor.

In addition, periodic recoding and comparison of previously analyzed segments were conducted

throughout the analytical process to assess coding consistency and refine category definitions where necessary. This procedure provided an additional layer of coding validation and enhanced the reliability of thematic interpretation.

ETHICAL CONSIDERATIONS

Although the study analyzed publicly available documents and did not involve human subjects it adhered to stringent ethical standards in research conduct. As the research relied exclusively on publicly accessible policy documents and did not collect data from human participants, formal ethical approval and informed consent were not required. All documents were properly cited according to APA 7th Edition guidelines to respect intellectual property and academic integrity. Reflexivity was maintained to mitigate researcher bias and coding decisions were documented for transparency. Ethical rigor in document analysis is emphasized in qualitative methodology literature recognizing the importance of accurate representation and responsible interpretation (Tracy, 2019; Bowen, 2009). Ethical safeguards ensured that interpretations remained grounded in the text avoiding speculative or unjustified conclusions.

RESULTS AND DISCUSSION

The qualitative analysis identified five interrelate themes that capture both convergence and divergence in how ethical AI principles are articulated, operationalized and translated into practice across governance frameworks.

Theme 1: Convergence around Core Ethical Principles

This theme captures the strong normative convergence across national and international AI governance documents around a core set of ethical principles as identified through qualitative document analysis and revealed a high degree of convergence around a core set of ethical principles across national and international AI governance frameworks. Transparency, fairness, accountability, privacy and human oversight emerged as the most frequently articulated ethical commitments across the sampled documents. The coding process identified transparency as the most frequently referenced ethical principle, appearing in international and national documents. Transparency was commonly associated with explainability, public trust and disclosure obligations. For example, OECD principles emphasize transparency as a mechanism for promoting trustworthy AI whereas the EU AI Act links transparency to regulatory compliance and risk management. Similarly, UNESCO frames transparency within a broader human-rights-based approach demonstrating both convergence in ethical priorities and variation in policy interpretation. To illustrate these differences the analyzed documents framed transparency in distinct ways despite sharing a common ethical commitment. The (OECD, n.d.) AI Principles describe transparency and explainability as mechanisms for fostering trustworthy AI and enabling stakeholders to understand AI-driven outcomes. In contrast the (European Commission, 2021) AI Act operationalizes transparency through specific legal obligations requiring providers of certain AI systems to disclose relevant information and support regulatory oversight. UNESCO's Recommendation on the Ethics of Artificial Intelligence situates transparency within a broader human-rights framework emphasizing public awareness, democratic accountability and societal trust. These examples demonstrate that while transparency is widely endorsed across governance frameworks its interpretation and implementation vary according to institutional priorities and governance approaches.

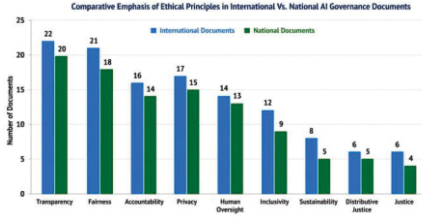
International frameworks such as (UNESCO, 2022; OECD, n.d) Recommendation on the Ethics of Artificial Intelligence and AI Principles consistently emphasize these values as foundational to responsible AI governance. Similarly regulatory instruments such as the (European Commission, 2021) AI Act embed these principles within formal regulatory structures for high-risk AI systems.

Table 1: Frequency and Emphasis of Ethical Principles across AI Governance Documents

Ethical Principle	Number of International Documents	Number of national Documents	Emphasis/High Medium /Low	Examples Document cited
Transparency	25	20	High	OECD (n.d.), EU AI Act (2021)
Fairness	21	18	High	UNESCO (2022), US AI Strategy
Accountability	16	14	Medium	OECD (n.d.), Canada AI Strategy
Privacy	17	15	Medium	EUAIAct (2021), Singapore
Human Oversight	14	13	Medium	UNESCO (2022), China AI Strategy
Inclusivity	12	9	Low	UNESCO (2022), OECD (n.d.)
Sustainability	8	5	Low	Limited mention
Distributive Justice	6	5	Low	Rarely explicitly addressed
Justice	6	4	Low	Rarely explicitly addressed

The table demonstrates that while fairness and transparency appear consistently across documents accountability and enforcement mechanisms are unevenly specified. Regulatory frameworks particularly the EU AI Act assign responsibility through legally binding obligations whereas international guidelines rely predominantly on voluntary compliance and moral persuasion.

Table shows clear patterns in ethical emphasis. Transparency and fairness appear in over 25 documents each reflecting broad normative alignment across governance frameworks. Accountability while emphasized in 16 international and 14 national documents is articulated less consistently and often with varying levels of operational detail. Principles such as sustainability, inclusivity and distributive justice are underrepresented appearing in fewer than 10 documents highlighting significant discursive silences in global and national AI ethics governance. These patterns suggest that governance frameworks prioritize procedural oversight and risk mitigation over broader socio political and redistributive concerns



"Figure 1 demonstrates a clear concentration of ethical discourse around transparency, fairness and accountability across both international and national AI governance frameworks. These principles consistently receive the highest levels of articulation indicating a strong normative focus on procedural ethics and trust-oriented governance mechanisms. Conversely principles associated with structural justice sustainability and distributive concerns receive comparatively limited attention particularly within national level documents. This pattern highlights selective ethical framing in which governability and system oversight are prioritized over broader socio political and redistributive considerations. The figure thus provides an empirical anchor for analyzing discursive silences and power dynamics within contemporary AI governance narratives.

Theme 2: Divergent Institutional Framing and Operationalization of Ethics

Critical discourse analysis revealed that ethical principles are framed through distinct institutional logics that reflect the authority and governance capacity of the issuing bodies. UNESCO's framework predominantly employs normative and aspirational language frequently using modal verbs such as "should" and "encouraged to" which position ethical responsibility as a shared moral obligation among stakeholders. For instance, UNESCO Emphasizes that Member States "should promote human centered AI systems that respect human rights and dignity" (UNESCO, 2022).

In contrast the EU AI Act adopts a legalistic and compliance-oriented discourse employing enforceable language such as "shall ensure" and "must comply". This framing constructs ethics as a matter of regulatory compliance rather than moral commitment assigning explicit accountability to AI providers and deployers. OECD guidelines occupy an intermediate position combining normative principles with risk management language aimed at institutional and economic stability.

This variation in discursive framing is not merely a stylistic difference but reflects underlying institutional mandates and political priorities. UNESCO as an intergovernmental normative organization without direct regulatory authority relies on persuasive and value-oriented discourse to encourage voluntary alignment among member states. In contrast the EU possesses legislative capacity and therefore frames ethical obligations through legally enforceable requirements that translate abstract ethical principles into compliance mechanisms.

The OECD's emphasis on innovation economic competitiveness and risk governance positions ethics within a framework that seeks to balance technological development with societal safeguards. These differences reveal how ethical discourse functions as a governance instrument shaping not only what constitutes responsible AI but also who is empowered to define, monitor and enforce ethical standards.

The comparison demonstrates a continuum of governance approaches ranging from UNESCO's value-based persuasion to the EU's legally binding regulation with the OECD occupying an intermediary position that links ethical guidance with economic and innovation policy objectives.

This contrast illustrates how similar ethical principles are operationalized differently according to institutional power regulatory authority and governance objectives. Rather than functioning as universally interpreted concepts ethical principles are embedded within specific political and legal contexts that influence their implementation and enforcement. From a critical discourse perspective ethics serves not only as a normative framework but also as a mechanism through which institutions exercise governance authority and establish legitimacy in the regulation of emerging technologies. Consequently, the operationalization of AI ethics reflects broader struggles over responsibility, accountability and control within the evolving global governance landscape of artificial intelligence

Theme 3: Accountability and Human Oversight as Contested Concepts

Accountability and human oversight emerged as central but unevenly articulated principles across the analyzed documents. While nearly all frameworks emphasize the importance of human involvement in AI decision making the nature and scope of oversight vary considerably. In several documents human oversight is framed ambiguously without clarifying whether oversight entails meaningful decision authority or merely symbolic monitoring.

For example, OECD documents refer to "appropriate human oversight" without specifying operational thresholds or decision rights effectively diffusing responsibility across institutions. In contrast the EU AI Act explicitly assigns accountability to identifiable actors requiring providers of high-risk systems to

implement oversight mechanisms documentation and post deployment monitoring

The differing treatment of accountability reflects contrasting governance philosophies and institutional capacities. Documents employing broad formulations such as “appropriate human oversight” provide flexibility for implementation across diverse contexts but simultaneously leave significant discretion to organizations responsible for AI deployment. This ambiguity may facilitate policy adoption among a wider range of stakeholders however it can also obscure responsibility when adverse outcomes occur. By contrast regulatory frameworks that specify accountable actors and oversight obligations reduce interpretive flexibility while increasing the potential for enforcement and legal scrutiny. This divergence reflects a broader tension between distributed ethical responsibility and centralized regulatory accountability.

From a critical discourse perspective this tension reflects broader struggles over the allocation of responsibility in AI governance. The use of flexible and indeterminate language may preserve institutional adaptability yet it can also dilute mechanisms for redress, transparency and liability. These findings suggest that accountability is not merely a technical or procedural requirement but a politically negotiated concept through which institutions define the boundaries of responsibility and control in the governance of artificial intelligence.

Theme 4: Discursive Silences and Marginalized Groups in AI Ethics Governance

Beyond what governance documents emphasize the analysis also identified significant discursive silences. Ethical concerns related to marginalized populations distributive justice and structural inequalities are frequently under-articulated or treated superficially. While inclusivity is rhetorically emphasized its operationalization remains weak suggesting a gap between normative commitments and enforceable governance mechanisms.

For example, UNESCO’s Recommendation on the Ethics of Artificial Intelligence explicitly highlights inclusivity, diversity and the protection of vulnerable groups. However, many national AI strategies focus primarily on innovation, economic competitiveness and risk management while providing limited discussion of distributive justice, structural inequalities or mechanisms for protecting marginalized populations. This disparity suggests that inclusivity is often acknowledged as an ethical aspiration but is less frequently translated into specific governance measures.

Additionally emerging AI technologies such as generative AI and autonomous decision-making systems are inconsistently addressed across frameworks. Environmental sustainability long term societal impact and power asymmetries between global technology producers and AI consuming regions remain largely absent from policy discourse. These omissions indicate that governance priorities are predominantly shaped by concerns related to innovation management regulatory feasibility and economic competitiveness rather than distributive justice or global equity. The selective visibility of ethical issues reflects underlying structural biases in global AI governance agendas.

These silences operate as discursive boundary setting mechanisms that determine which ethical concerns are legitimized within AI governance and which are excluded from formal consideration. Such exclusions are not incidental but reflect broader political economic and institutional hierarchies that shape global AI policy agendas.

Theme 5: The Policy Practice Translation Gap and Contextual Inequalities

Although AI governance documents widely articulate ethical principles their translation into enforceable and context-sensitive practices remains inconsistent due to

ambiguous accountability mechanisms and reliance on voluntary compliance. This reliance on soft governance instruments and non-binding guidelines reflects a broader preference for flexibility over enforceability in global AI governance but it also limits the ability to ensure consistent implementation across diverse institutional settings.

This implementation gap is particularly pronounced in low resource and regulatory limited contexts where the absence of technical infrastructure institutional capacity and localized policy adaptation mechanisms restricts the effective operationalization of global ethical frameworks. As a result, governance models risk reproducing structural inequalities by imposing uniform ethical expectations without accounting for contextual disparities in implementation capacity.

DISCUSSION

Ethics as Normative Commitment vs Regulatory Control

The discussion is structured around the five major themes identified in the Results section linking empirical findings to broader theoretical and governance debates, particularly normative commitment and ethics as a mechanism of regulatory control. International frameworks such as UNESCO and the OECD frame ethical AI primarily as a shared moral responsibility emphasizing values such as inclusivity human dignity and social good. This normative framing aligns with principle-based ethics, which prioritizes ethical ideals but relies heavily on voluntary adoption and institutional goodwill. In contrast the EU AI Act reframes ethics as a matter of legal obligation embedding ethical principles within enforceable regulatory requirements. This shift reflects a governance logic that prioritizes compliance risk mitigation and institutional accountability over ethical persuasion. The coexistence of these approaches suggests that ethical convergence at the principal level does not translate into uniform governance practices reinforcing the study's argument that ethical AI governance is shaped by institutional power and regulatory capacity rather than shared values alone.

For example, UNESCO's Recommendation on the Ethics of Artificial Intelligence emphasizes principles such as human dignity, inclusiveness and sustainability while encouraging member states to adopt these values through voluntary policy measures. Similarly, the OECD AI Principles promote responsible stewardship and human centered values but leave implementation mechanisms largely to national governments. In contrast, the EU AI Act translates comparable ethical concerns into legally binding obligations through risk-based requirements, conformity assessments and regulatory oversight. These examples illustrate how similar ethical principles are interpreted through distinct governance mechanisms ranging from voluntary commitment to enforceable compliance.

This duality reflects a deeper governance tension between ethics as a moral discourse and ethics as a regulatory technology. Rather than existing as opposing models normative and legalistic approaches function as complementary governance strategies that collectively enable both legitimacy and control within global AI governance regimes. However, this coexistence also produces structural inconsistencies in implementation, as shared ethical principles are reframed through different institutional logics.

Power, Institutional Authority and Governance Design

From a critical discourse perspective, the language used in AI governance documents reflects underlying power structures and institutional priorities. The reliance on aspirational language in international guidelines enables broad consensus but limits enforceability effectively diffusing responsibility across stakeholders. Conversely regulatory instruments concentrate power and accountability within formal legal systems which may enhance compliance but risk marginalizing non-compliant actors or regions with limited regulatory capacity.

This power asymmetry has significant implications for global AI governance. Countries and institutions with strong regulatory infrastructures are better positioned to implement enforceable ethical frameworks while regions with weaker institutional capacity may be left reliant on soft governance mechanisms. As a result, global AI ethics discourse may inadvertently reinforce existing geopolitical and technological inequalities.

This gap underscores a critical limitation in current AI ethics scholarship which often assumes that normative articulation is sufficient for ethical implementation. The findings instead demonstrate that without enforceable structures and context-sensitive governance mechanisms ethical commitments risk remaining performative rather than operational.

The Policy to Practice Translation Gap and contextual Inequalities

The study's findings provide empirical support for the existence of a policy to practice translation gap in AI ethics governance. While ethical principles are consistently articulated across governance documents, their operationalization remains uneven and context dependent. The reviewed documents reveal that transparency, accountability and human oversight are among the most frequently referenced principles however detailed implementation guidance varies considerably across frameworks. While regulatory instruments such as the EU AI Act specify compliance obligations and monitoring requirements, several international guidelines primarily articulate broad expectations

Without defining measurable indicators for implementation. This variation helps explain why organizations may endorse ethical principles in policy statements while facing difficulties in translating them into operational practices. Socio technical factors, including organizational readiness, sectoral complexity, institutional capacity and resource availability significantly influence the extent to which ethical commitments can be implemented in practice.

These findings are consistent with previous research highlighting the gap between high level ethical principles and their practical implementation within organizations and regulatory systems. While prior studies have identified concerns regarding ethics washing and limited operational guidance the present analysis demonstrates that these challenges persist across both international and national AI governance frameworks despite broad agreement on core ethical values.

This gap is further exacerbated by ambiguous definitions of accountability and human oversight which allow organizations to claim ethical compliance without demonstrating measurable outcomes. The findings extend prior critiques of "ethics washing" by demonstrating how governance language itself enables symbolic compliance while limiting practical accountability. This theme highlights the structural disconnect between ethical commitments at the policy level and their realization in practice.

Implications for Global and Global South AI Governance

Extending the policy practice translation gap, this theme emphasizes how contextual and regional inequalities shape the feasibility of ethical AI implementation. For Global South contexts where infrastructural regulatory and resource constraints may limit the adoption of complex compliance regimes. International guidelines that emphasize moral responsibility without providing implementation support risk placing disproportionate ethical burdens on institutions least equipped to fulfill them.

These findings suggest the need for context-sensitive governance models that balance normative ethical commitments with practical implementation pathways. Without such adaptation AI governance frameworks may unintentionally perpetuate global inequities undermining the inclusive and human centered values they claim to promote.

CONCLUSION

This study set out to critically examine how ethical principles are articulated, framed and operationalized within 25 national and international AI governance documents. Through an exploratory qualitative document analysis informed by principle-based ethics, socio-technical systems theory and critical discourse analysis. The study demonstrates that while global AI governance frameworks converge around core ethical principles such as transparency, fairness, accountability, privacy and human oversight their translation into actionable and enforceable practices remains uneven and fragmented.

However, despite this convergence, the study revealed persistent operational gaps and inconsistencies. Many documents provide high level normative guidance but offer limited mechanisms for implementation, monitoring and evaluation (Raji et al., 2021; McGregor et al., 2020). This disconnect between principle and practice continues to challenge organizations and policymakers seeking to operationalize ethical AI across diverse sectors.

A key contribution of this study lies in moving beyond descriptive mapping of ethical principles to interrogate the discursive and institutional dynamics shaping AI governance. The findings reveal that international frameworks frequently frame ethics as a moral or aspirational commitment relying on voluntary adoption and shared responsibility whereas regulatory instruments most notably the EU AI Act embed ethical principles within legally enforceable compliance structures. This divergence illustrates how ethical AI governance is influenced not only by normative values but also by institutional authority regulatory capacity and political priorities.

Building on these findings the study proposes the policy to practice translation gap as a key explanatory concept for understanding why ethical commitments often fail to produce consistent implementation outcomes. Although ethical principles are consistently articulated at the policy level their operationalization is constrained by ambiguous accountability mechanisms limited socio-technical guidance and insufficient consideration of contextual realities. Discursive silences particularly regarding marginalized populations distributive justice and emerging AI technologies further weaken the capacity of existing frameworks to deliver equitable and socially responsive outcomes.

From a theoretical perspective the integrated analytical framework employed in this study demonstrates the value of combining normative ethics with socio-technical and discourse-based approaches. This triangulation enables a more comprehensive understanding of AI governance documents as socio political artifacts that both reflect and reproduce power relations rather than neutral expressions of ethical consensus.

Looking forward the findings underscore the need for context-sensitive implementation-oriented AI governance models that bridge ethical aspiration and operational reality. Policymakers and institutions must move beyond high level ethical commitments toward frameworks that clearly assign responsibility, support implementation across diverse socio-technical contexts and explicitly address equity and inclusion. Without such efforts AI ethics risks remaining symbolic reinforcing institutional legitimacy while offering limited protection against real-world harms.

However, contemporary AI governance demonstrates significant progress in articulating shared ethical principles the effectiveness of these frameworks ultimately depends on their capacity to translate ethics into practice. By highlighting discursive framing power dynamics and operational gaps this study provides a conceptual and empirical foundation for advancing more accountable inclusive and socially grounded AI governance.

The findings have important implications for policy, practice and research. For policymakers and organizations, the results highlight the importance of strengthening accountability, implementation mechanisms and context-sensitive governance

approaches. For researchers, the study demonstrates the value of integrating normative, socio-technical and critical discourse perspectives in the analysis of AI governance frameworks. While global AI governance shows increasing convergence around foundational ethical principles greater attention to implementation, equity and emerging challenges remains essential. By highlighting both achievements and limitations within existing governance documents, this study contributes to ongoing efforts to advance more accountable, inclusive and socially responsive AI governance (Floridi et al., 2020; Fjeld et al., 2020; Raji et al., 2021; Obermeyer et al., 2019)

LIMITATION OF THE STUDY

This study is subject to several limitations that should be acknowledged when interpreting the findings. First the research is based on a qualitative document analysis of 25 national and international AI governance and policy documents. While this dataset provides a broad overview of dominant ethical framings it does not capture the full diversity of emerging or unpublished policy practices particularly at organizational or industry implementation levels.

Second the study focuses on formal policy and governance texts which reflect intended principles and regulatory aspirations rather than actual implementation outcomes. As a result, the analysis cannot directly assess how ethical principles are enacted in practice across different institutional or socio-technical contexts.

Third although the use of critical discourse analysis and socio-technical systems theory enables a deeper examination of language, power and institutional framing, interpretive analysis inherently involves a degree of subjectivity. While efforts were made to maintain analytical rigor through theoretical triangulation alternative interpretations of the same texts are possible.

Finally, the study's emphasis on national and international level governance documents limits its ability to fully account for localized or informal governance mechanisms particularly in rapidly evolving AI deployment contexts. These limitations suggest the need for future research to integrate empirical case studies stakeholder interviews or implementation level evaluations to complement document-based analyses.

RECOMMENDATIONS

Based on the findings of this study, several recommendations emerge for policymakers, AI developers, organizations and researchers to strengthen AI ethics governance and promote responsible, equitable and context-sensitive AI deployment.

Policy Recommendations

- **Operationalize Principles and Mitigate Risks:** Governments and regulatory bodies should develop clear, enforceable standards to translate high-level ethical principles (e.g. fairness, transparency, accountability) into practice with structured compliance monitoring (Raji et al., 2021; Sanderson et al., 2024). This regulatory framework must proactively adapt to account for rapid technological shifts like generative AI, autonomous systems and large-scale machine learning (McCormack et al., 2025; Tabbassum & Chintale, 2024).
- **Socio-Technical Integration and Equity:** Policymakers should consider organizational infrastructure, cultural contexts and sector-specific nuances rather than applying a one size fits all model [Baxter & Sommerville, 2011]. AI governance policies must explicitly incorporate equity-focused provisions to ensure systems do not reinforce existing social inequalities (Obermeyer et al., 2019; Barthwal et al., 2025).
- **Framework Harmonization:** Cross-border collaboration between international organizations and national governments should be prioritized to enhance consistency in AI ethics standards while

allowing flexibility for local adaptation (Fjeld et al., 2020; Corrêa et al., 2022).

PRACTICAL RECOMMENDATIONS

- **Monitoring and Sector Guidelines:** Organizations should deploy technical and procedural tools to monitor AI system performance, detect bias and ensure continuous alignment with ethical standards through regular audits (Floridi et al., 2020; Sanderson et al., 2024). Ethical guidelines should be developed for high-stakes sectors like healthcare, finance, law enforcement and education to address sector-specific risks directly (Raji et al., 2021).
- **Stakeholder Engagement and Capacity Building:** Organizations must involve multidisciplinary stakeholders including ethicists, engineers, social scientists and affected communities in the design deployment and evaluation phases to enhance accountability [Jobin et al., 2019]. This must be supported by investments in training programs for developers, managers and policymakers to understand AI ethics principles and socio-technical strategies (McGregor et al., 2020).

Research Recommendations and Future Directions

- **Empirical and Technical Evaluation:** Future research should assess the practical implementation and effectiveness of AI ethics frameworks in real world organizational settings through case studies and field experiments (Sanderson et al., 2024). Academics should focus heavily on the unique ethical implications of generative AI and autonomous decision-making systems regarding bias propagation and accountability gaps (McCormack et al., 2025).
- **Cross-Cultural and Contextual Studies:** Comparative studies across different countries and cultural landscapes are highly recommended to identify context-sensitive strategies for operationalizing ethical principles globally (Fjeld et al., 2020; Corrêa et al., 2022). Future research should explore mechanisms for ensuring equity, particularly how AI systems impact vulnerable populations, to build frameworks centered on social justice (Obermeyer et al., 2019; Barthwal et al., 2025).
- **Socio-Technical Integration:** Studies should continue to combine principle-based ethics with socio-technical systems theory and critical discourse analysis to examine not just what principles are stated, but how they are implemented and framed in practice (Baxter & Sommerville, 2011; Fairclough, 2013).

To expand the current academic scope, the following overall suggestions to guide future research:

- Development of metrics and evaluation tools to measure organizational adherence to AI ethics alongside long-term, longitudinal studies to monitor how frameworks evolve with rapid technological advancements.
- Exploration of multi-stakeholder governance models combining industry, academia and civil society, policy diffusion to reduce fragmentation and research on ethics by design approaches embedding principles directly into AI development processes.

REFERENCES

Amanah Principle in AI Governance. (2025). *AI and Ethics*. https://doi.org/10.1007/s43681-025-00780-y

Arar, K. H., Özen, H., Polat, G., & Turan, S. (2025). Artificial intelligence, generative AI, and research integrity: A hybrid systemic review. *Smart Learning Environments, 12*, 44. https://doi.org/10.1186/s40561-025-00403-3

Barthwal, R., Campbell, M., & Shrestha, S. (2025). Ethical challenges in large-scale AI implementation. *AI & Society, 40*, 1203–1220. https://doi.org/10.1007/s00146-025-01520-4

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering *Interacting with Computers, 23*(1), 4–17. https://doi.org/10. 1016/j.intcom. 2010.07.003

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W.

Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 316–334). Cambridge University Press.

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017).AI now 2017 report.AI Now Institute, New York University.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics, 24*(2), 505– 528. https://doi.org/10.1 007/s1 1948-017-9901-7

Chen, M., Mao, S., & Liu, Y. (2020). Big data: A survey. *Mobile Networks and Applications, 25*(1), 1–39. https://doi.org/10. 1007/s 11036-019-01383-8

Corrêa, C., Silva, R., & Santos, P. (2022). Comparative analysis of AI governance frameworks: International and national perspectives. *Technology in Society, 71*, 102100.
https://doi.org/10.1016/j.techsoc.2022.102100

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Massmann, D., & de Oliveira, N. (2022). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *arXiv*. https://arxiv.org/abs/ 2206.11922

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer.

European Commission. (2021). *Proposal for a regulation on artificial intelligence (AI Act)*.
https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Fairclough, N. (2013). *Critical discourse analysis: The critical study of language* (2nd ed.). Routledge.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2020). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, 689–707. https://doi.org/10.1007/s11023-020-09517-8

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. C., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, 2020-1. https://doi.org/10.2139 /ssrn.3518482

Frontiers Editorial Board. (2025). The evolving literature on the ethics of artificial intelligence for healthcare: A PRISMA scoping review. *Frontiers in Digital Health, 3*, 1701419. https://doi.org/10.3389/fdgth.2025.1701419

Hasan Rumlus, M. (2025). Governing AI: A narrative review of algorithmic accountability and legal frameworks. *Legalis – Journal of Law Review*. https://doi.org/10.61978/legalis.v3i2. 810

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

Kamali, J., Alpat, M. F., & Bozkurt, A. (2024).AI ethics as a complex and multifaceted challenge: Decoding educators' AI ethics alignment through activity theory. *International Journal of Educational Technology in Higher Education, 21*, Article 62. https://link.springeropen.com/articles/10. 1186/s41239-024-00496-9

Kaanskaia, A. N. (2025). *Ethical challenges in artificial intelligence: Fairness, transparency, and accountability*. *NEYA Global Journal of Non-Profit Studies*. https://doi.org/10.64357/ neya-gjnps-tech-ai-02

Khan, A. A., Akbar, M. A., Fahmideh, M., Liang, P., Waseem, M., *Ahmad, A., Niazi, M., & Abrahamsson,* P. (2022). AI ethics: An empirical study on the views of practitioners and lawmakers. *arXiv*. https://arxiv.org/ abs/2207.01493

McCormack, M., Gifford, T., & Vass, A. (2025). Generative AI and ethical risk assessment. *Ethics and Information Technology, 27*(2), 205–223. https://doi.org/10.1007/s10676-025-09740-9

McGregor, S., Murray, D., & Ng, V. (2020). Ethics washing in AI governance: The disconnect between principle and practice. *AI & Society, 35*, 611–622. https://doi.org/10.1007/s00146-020- 00962-0

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2) 1–21. https://doi.org/10.1177/205395 1716679679

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

OECD. (n.d.). *principles on artificial intelligence*. OECD. https://www.oecd.org/going-digital/ai/principles/

Oguru, M. F. (2025). Ethics of artificial intelligence: Balancing innovation with privacy, fairness, and accountability. *Asian Journal of Advanced Research and Reports, 19*(10), 42–50. https://doi.org/10.9734/ajarr /2025/v19i 101169

Pant, A., Hoda, R., Tantithamthavorn, C., & Turhan, B. (2022). Ethics in AI through the practitioner's view: A grounded theory literature review. *arXiv*. https://arxiv.org/ abs/2206.09514

Raji, I. D., Binns, R., & Mitchell, M. (2021). AI governance documents: An analysis of principles, practices, and gaps. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 170–180).ACM. https://doi.org/10.1145/3461702.3462603

Sanderson, H., Patel, R., & Liu, J. (2024). Operationalizing AI ethics: Challenges and strategies. *AI and Ethics, 5*(1), 89–105. https://doi.org/10.1007/s43681-024-00063-1

Tabbassum, A., & Chintale, P. (2024). Ethical considerations in AI developments. *International Journal of Geographical Information Science, 38*(2), 201–223. https://doi.org/10.1080/13658816.2024.1913472

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

UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000380455

Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and AI systems. *Philosophical Transactions of the Royal Society A, 376*, 20180085. https://doi.org/10.1098/rsta.2018. 0085

Zeng, Y., Lu, E., & Huangfu, C. (2019). Linking AI principles to practice: Implementations, challenges, and opportunities. *AI & Society, 34*, 441–456. https://doi.org/10.1007/s00146-018-0850-6.