Data Ethics and Responsible AI Use: A Scoping Review of Governance Principles, Implementation Challenges and Future Directions

Sunday Olusola LADIPO[1]* and Ifaka Queen INAZU[2]

[11]Medical librarian, Lagos State University College of Medicine (LASUCOM), Ikeja, Lagos, Nigeria.
[2]Librarians' Registration Council of Nigeria, Professional Service Department, Abuja, Nigeria.
https://orcid.org/0000-0001-9344-3115
Corresponding author email: sudayladipo@gmail.com

Direct Research Journal of Engineering and Information Technology

Vol. 14(3), Pp. 37-49, September 2026

https://journals.directresearchpublisher.org/index.php/drjeit; https://www.ajol.info/index.php/drjeit

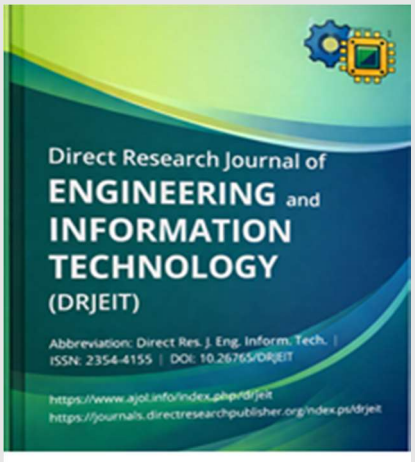
Review Article
ISSN: 2354-4155

Received 9 July 2026; Accepted 27 August 2026; Published 12 September 2026

ABSTRACT:

Artificial intelligence is increasingly embedded in education, healthcare, finance, public administration, scientific research and media systems. These applications create efficiency and innovation but intensify concerns about privacy, bias, opacity, accountability, consent, surveillance, intellectual property, misinformation and the concentration of technological power. Data ethics is central to this debate because AI systems inherit the legal, social and epistemic properties of the data on which they are trained, evaluated and deployed. This paper examines the relationship between data ethics and responsible AI use by synthesising recent literature on ethical principles, AI governance frameworks, data stewardship, implementation barriers and sector-specific risks, and proposes an integrated framework for developing, deploying and monitoring AI systems in ways that are lawful, transparent, accountable, inclusive and human-centred. A scoping review and evidence-mapping design was adopted. Note for readers: the evidence base was developed as a structured evidence map rather than a fully registered systematic review with dual independent database screening; a reproducible search rerun using Scopus or Web of Science is recommended before final journal publication, and this limitation is discussed fully in the Limitations section. Traceable scholarly articles, open books, standards, legal instruments and official policy documents published between 2016 and 2026 were identified through targeted searches of institutional portals, citation chaining and open scholarly platforms. Evidence was charted and synthesised thematically across six literature clusters. Twenty-eight high-relevance sources were retained for primary synthesis, supplemented by additional methodological and comparative sources. The evidence shows strong convergence around fairness, transparency, privacy, accountability, human oversight, safety and inclusiveness, but weaker agreement on implementation. The most persistent gap is not lack of principles; it is the difficulty of converting principles into lifecycle controls, institutional accountability, enforceable standards and measurable outcomes. Recent evidence from 2025 and 2026 corroborates this finding: even as AI regulation has expanded globally, the proportion of organisations with formally implemented AI governance frameworks remains critically low. Developing countries face additional constraints related to regulatory capacity, digital infrastructure, local datasets, talent, procurement dependence and unequal bargaining power with global technology providers. Responsible AI cannot be achieved by technical optimisation alone. It requires ethically sourced and well-governed data, risk-based governance, documentation, participatory oversight, continuous monitoring, enforceable accountability and context-sensitive capacity building. The proposed six-layer integrated framework ethical leadership, data ethics, model governance, deployment controls, stakeholder assurance and continuous learning provides a conceptual pathway for researchers, policymakers, developers and organisational leaders seeking trustworthy AI systems. Empirical validation of this framework across sectors and regional contexts is an identified priority for future research.

Keywords: *Accountability; Artificial intelligence; AI lifecycle governance; Algorithmic auditing; Data ethics; Developing countries AI policy; Governance; Responsible AI; Trustworthy AI*



Citation: Ladipo, S. O. & Inazu, I. Q. (2026). *Data Ethics and Responsible AI Use: A Scoping Review of Governance Principles, Implementation Challenges and Future Directions. Direct Research Journal of Engineering and Information Technology* Vol. 14(3), Pp. 37-49.
https://doi.org/10.4314/drjeit.v14i3.4

INTRODUCTION

Artificial intelligence has moved from a specialised research field into a general-purpose digital infrastructure. It now supports search engines, learning platforms, medical imaging, credit scoring, fraud detection, hiring systems, public-service triage, scientific discovery, media production and generative content tools. This expansion has made AI a governance problem as much as a technical one. A model may achieve high predictive accuracy and still be socially irresponsible if its training data are unlawfully obtained, its outputs discriminate against protected groups, its operation cannot be explained to affected persons, or its deployment removes meaningful human judgement from consequential decisions (Amershi *et al*., 2019; Dignum, 2019).

The rapid growth of generative AI has sharpened these concerns. Foundation models are trained on broad data at scale and then adapted for many downstream tasks, creating both remarkable flexibility and systemic risk, because defects in the foundation layer can travel into many applications (Bommasani *et al*., 2021). Large language models can support writing, tutoring, coding and information retrieval, but they may also produce hallucinated claims, reproduce cultural bias, expose confidential information, or encourage over-reliance by users who lack the expertise to verify outputs (Bender *et al*., 2021; Kasneci *et al*., 2023). The ethical challenge is therefore not merely how to build more capable AI systems, but how to ensure that capability remains governed by human rights, public value and institutional responsibility (Taddeo and Floridi, 2018).

Data ethics provides the foundation for responsible AI because AI systems are not ethically neutral machines. They are statistical and computational artefacts shaped by data collection, labelling, classification, curation, access, exclusion and interpretation. If data are incomplete, biased, stale, unlawfully acquired, poorly documented or extracted from vulnerable groups without meaningful safeguards, AI outputs can become a mechanism for automating injustice (Mittelstadt *et al*., 2016; Crawford, 2021). Wilkinson *et al*. (2016) established the importance of FAIR data principles findability, accessibility, interoperability and reusability in scientific data stewardship, but responsible AI also requires attention to privacy, consent, context, representativeness, proportionality and community impact.

The global AI ethics literature shows growing convergence around principles such as transparency, justice and fairness, non-maleficence, responsibility and privacy (Jobin *et al*., 2019). Floridi and Cowls (2019) similarly identify an influential five-principle framework of beneficence, non-maleficence, autonomy, justice and explicability. Fjeld *et al*. (2020) mapped 36 AI ethics documents from across sectors and regions and found high agreement on values but significant divergence in how those values are operationalised. Yet a brutally honest reading of the field reveals a gap between ethical language and organisational practice. Many organisations publish AI principles, but fewer can demonstrate who owns model risk, how training data were documented, how bias was tested, how affected users can appeal decisions, or how model drift is monitored after deployment (Mittelstadt, 2019; Morley *et al*., 2020). This is the implementation gap at the centre of responsible AI.

The study is significant for researchers because it integrates debates that are often fragmented across computer science, information ethics, law, public policy, information systems and organisational governance. It is significant for policymakers because it translates AI ethics principles into governance concerns such as data protection, auditability, risk classification and regulatory capacity. It is significant for technology companies because it clarifies that responsible AI is not a public relations label; it requires documentation, testing, monitoring and accountability across the product lifecycle. It is significant for educational and research institutions because AI literacy and data ethics are now core competencies for students, staff and research teams. It is significant for civil society because communities affected by AI systems need mechanisms for participation, explanation, contestation and redress.

A stronger counterargument must also be acknowledged. Some technologists argue that too much regulation may slow innovation, especially in countries trying to build competitive digital economies. However, the opposite danger is greater: ungoverned AI can scale harm faster than traditional systems, and public backlash after harm occurs can destroy trust in innovation itself. The best position is not regulation versus innovation; it is trustworthy innovation through proportionate, risk-based and context-aware governance (OECD, 2019; European Parliament and Council of the European Union, 2024).

Statement of the Problem

AI adoption has outpaced the ability of many institutions to govern it responsibly. Ethical risks arise at every stage of the AI lifecycle. During data collection, individuals and communities may not understand how their data will be reused. During data preparation, historical inequalities may be encoded into labels and features. During model development, optimisation targets may reward efficiency while ignoring fairness, safety or human dignity. During deployment, users may assume that automated outputs are objective. During post-deployment operation, systems may drift, interact with new populations, or produce harms that were not anticipated during testing (Amershi *et al*., 2019; Raji *et al*., 2020; Lu *et al*., 2024). The problem is intensified by opacity: many AI systems are difficult to interpret, especially when complex machine-learning models are used in high-stakes settings. Explainability research attempts to address this by developing methods for opening black-box models, but explanation is not only

a technical artefact; it must also be meaningful to affected users, regulators, managers and front-line professionals (Guidotti *et al*., 2018; Wachter, Mittelstadt and Russell, 2017). A risk score that is mathematically explainable to a data scientist may still be useless to a loan applicant, patient, student or civil servant who needs to understand why a decision was made and how to challenge it.

Regulatory frameworks are also uneven. The European Union has developed strong data protection rules through the GDPR and a risk-based AI governance model through the AI Act (European Parliament and Council of the European Union, 2016, 2024). Nigeria's data-protection landscape is additionally governed by the Nigeria Data Protection Act (Federal Republic of Nigeria, 2023), which establishes rights-based rules for personal data processing aligned with international standards. NIST has provided voluntary AI risk management guidance organised around governance, mapping, measurement and management (NIST, 2023). The NIST Privacy Framework (NIST, 2020) provides complementary guidance for managing privacy risk as an enterprise governance function, applicable alongside the AI RMF. UNESCO (2021) and OECD (2019) have issued international AI ethics principles. However, the existence of frameworks does not automatically create responsible practice. Developing countries may face the additional burden of importing AI systems designed around foreign languages, norms, datasets and regulatory assumptions (African Union, 2024; Federal Republic of Nigeria/NITDA, 2024).

The specific problem addressed in this review is that data ethics and responsible AI are often discussed separately. Data ethics is treated as a privacy or consent issue, while responsible AI is treated as a model-performance or governance issue. In reality, they are inseparable. Data ethics determines what kind of knowledge enters the AI system; responsible AI determines how that knowledge is transformed into decisions, recommendations, predictions and social consequences. A framework that fails to integrate both will be incomplete (Mittelstadt *et al*., 2016; Jobin *et al*., 2019; Morley *et al*., 2020).

Aims and Objectives

The aim of this paper is to examine the relationship between data ethics and responsible AI use by synthesising recent evidence and proposing strategies for ethical AI governance. The scope is limited to English-language sources from 2016 to 2026, with priority given to peer-reviewed scholarship, intergovernmental policy documents, technical standards and open governance frameworks (Table 1). The objectives are to:

1. Examine the concept of data ethics in the AI era;
2. Explore the core principles underpinning responsible AI;
3. Identify ethical challenges associated with AI-driven decision-making;
4. Analyse existing AI governance and regulatory frameworks;
5. Propose recommendations for strengthening responsible AI practices across sectors and contexts.

Table 1: Review Paper Objectives and Review Questions

Objective	Research questions
Data ethics	RQ1: What is data ethics and why is it important for AI?
Responsible AI principles	RQ2: What principles define responsible AI?
Ethical risks	RQ3: What ethical risks accompany AI deployment?
Governance	RQ4: How effective are current governance frameworks?
Strategy	RQ5: What strategies can promote responsible AI use?

METHODOLOGY

This paper adopted a scoping review and evidence-mapping design. IMPORTANT NOTE FOR READERS: the evidence base was developed as a structured evidence map based on open searches, institutional portals and citation chaining rather than as a fully registered, reproducible systematic review. The "drafting environment" does not certify independently screened database exports from Scopus or Web of Science. Accordingly, PRISMA 2020 (Page *et al*., 2021) is referenced as a reporting ideal, not as a fully executed protocol; Peters *et al*. (2020) informed scoping-review logic; and the full limitations of this approach are disclosed in the Limitations section. Before final journal submission, the search strategy should be rerun in at least one indexed database, independently screened by two reviewers, and the flow and counts documented with a PRISMA-ScR diagram.

The design was appropriate because the subject is broad, interdisciplinary and methodologically diverse. Relevant sources include peer-reviewed empirical studies, conceptual ethics papers, systematic reviews, conference papers, legal instruments, intergovernmental recommendations, technical standards and organizational governance frameworks. A meta-analysis was unsuitable because the evidence does not report comparable quantitative outcomes. The purpose was to map the conceptual terrain, identify dominant themes, compare governance frameworks and propose a practical synthesis. Sources are prioritised for clear bibliographic identity, DOI verifiability and relevance to the five research questions. Screening was conducted by the primary author; independent dual-screening was not completed at this draft stage, which is acknowledged as a limitation (Table 2).

Table 2: Eligibility criteria used for the evidence map.

Criterion	Inclusion	Exclusion
Publication period	2016–2026 (inclusive; period confirmed with 2025–2026 sources now included in the synthesis).	Sources outside the period.
Topic relevance	Data ethics, responsible AI, AI governance, privacy, fairness, accountability, model documentation, data governance, explainability, algorithmic auditing or sectoral AI risk.	General AI papers without ethical, governance or data relevance.
Source type	Peer-reviewed articles, open books, conference papers, official legal texts, standards and policy reports.	Unverifiable webpages, opinion pieces without analytical contribution or non-traceable claims.
Accessibility	Sources traceable through DOI, official institutional pages, publisher pages, arXiv, official PDFs or open repositories.	Sources with no stable bibliographic identity.
Language	English-language sources.	Non-English sources not available in English translation.

Table 3: Search concepts and example strings.

Concept block	Example search strings
Data ethics	"data ethics" AND "artificial intelligence"; "data governance" AND "AI" AND accountability; "FAIR data" AND AI governance
Responsible AI	"responsible AI" AND governance; "trustworthy AI" AND fairness AND transparency; "AI ethics principles" AND implementation
Risk and accountability	"algorithmic auditing" AND AI; "model cards" AND model reporting; "datasheets for datasets" AND accountability
Regulation	"AI Risk Management Framework"; "EU AI Act"; "UNESCO Recommendation Ethics AI"; "OECD AI Principles"; "Nigeria Data Protection Act"
Developing countries	"responsible AI in Africa"; "Continental AI Strategy"; "Nigeria National AI Strategy"; "AI governance developing countries"; "AI governance Africa 2025"

Information Sources and Search Strategy

Searches were conducted through Google Scholar-discoverable records, publisher platforms, official institutional portals and citation chaining from high-relevance works. Targeted portals included NIST, OECD, UNESCO, WHO, EUR-Lex, ISO, the African Union and Nigeria's NCAIR/NITDA. Search terms combined three domains: data ethics, responsible AI and governance. For final journal submission, search strings should be rerun in Scopus, Web of Science and Dimensions with documented string syntax, search dates and deduplication records (Table 3).

Data Charting and Synthesis

Data were charted by author or institution, year, source type, focus, key contribution, DOI or verifiable URL, and relevance to data ethics and responsible AI. The synthesis used narrative and thematic analysis organised into six literature clusters: governance, data stewardship, technical accountability, regulatory frameworks, sectoral AI and developing-country policy. Recurring themes identified were: data stewardship as the foundation of responsible AI; principles as necessary but insufficient; documentation and auditability as practical accountability mechanisms; regulation as a risk-based governance tool; and capacity and inclusion as the central developing-country challenge (Table 4).

Conceptual and Theoretical Foundations

Artificial Intelligence

Artificial intelligence refers broadly to computational systems that perform tasks associated with human intelligence, such as pattern recognition, prediction, classification, natural language processing, reasoning support and decision optimisation (Dignum, 2019). Contemporary AI systems include rule-based systems, supervised machine learning, unsupervised learning, reinforcement learning, deep learning, computer vision, natural language processing and generative AI. In practical settings, AI is rarely a single technology; it is a socio-technical system made up of datasets, models, interfaces, human users, organisational routines, legal rules and institutional incentives (Mittelstadt *et al*., 2016; Crawford, 2021).This socio-technical view matters because ethical failures often arise outside the algorithm itself. Responsible AI therefore requires a lifecycle view from problem formulation to retirement. Amershi *et al*. (2019) show that machine-learning development differs

Table 4: Summary of retained evidence sources.

Bender *et al*. (2021)	Conference	Large language models	Critiques risks of large language models, including bias and environmental cost.	https://doi.org/10.1145/3442188.3445922
UNESCO (2021)	Recommendation	Ethics of AI	Global normative instrument on human rights-based AI ethics.	https://unesdoc.unesco.org/ark:/48223/pf0000381137
WHO (2021)	Guidance	AI for health	Identifies ethics and governance principles for AI in health.	https://www.who.int/publications/i/item/9789240029200
Bommasani *et al*. (2021)	Report	Foundation models	Maps opportunities and risks of foundation models across sectors.	https://arxiv.org/abs/2108.07258
Lu *et al*. (2024)	Review/Article	Responsible AI patterns	Converts responsible AI into governance and engineering patterns.	https://doi.org/10.1145/3626234
Kasneci *et al*. (2023)	Article	LLMs in education	Analyses opportunities and challenges of large language models in education.	https://doi.org/10.1016/j.lindif.2023.102274
NIST (2023)	Framework	AI RMF	Organises AI risk management around Govern, Map, Measure and Manage.	https://doi.org/10.6028/NIST.AI.100-1
Barocas *et al*. (2023)	Open book	Fairness and ML	Explains fairness limits and opportunities in algorithmic decision-making.	https://fairmlbook.org/
ISO/IEC (2023)	Standard	AI management systems	Specifies requirements for establishing and improving an AI management system.	https://www.iso.org/standard/81230.html
European Parliament & Council (2024)	Regulation	EU AI Act	Risk-based legal framework for AI systems and general-purpose AI models.	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689
African Union (2024)	Policy strategy	Continental AI Strategy	Promotes Africa-centric, ethical and development-oriented AI adoption.	https://au.int/en/documents/20240217/continental-artificial-intelligence-strategy
Federal Republic of Nigeria/NITDA (2024)	Policy strategy	National AI Strategy	Emphasises responsible, ethical and inclusive AI innovation.	https://ncair.nitda.gov.ng/
World Economic Forum (2025)	Playbook	Responsible AI innovation	Offers practical plays for operationalising responsible AI.	https://www.weforum.org/publications/advancing-responsible-ai-innovation-a-playbook/
Maslej *et al*. (2025) [NEW]	Annual report	AI Index 2025	Documents surge in global AI regulation (59 federal + 131 state rules in USA in 2024) and rising AI incidents; corroborates implementation-gap finding.	https://aiindex.stanford.edu/report/
Kerasidou *et al*. (2025) [NEW]	Article	AI ethics review practice	Case study (Oxford) on translating AI ethics principles into research governance; confirms operationalisation gap.	https://doi.org/10.1177/17470161241248759
Yesán-Luján *et al*. (2026) [NEW]	Review	Ethical tech for AI governance	Reviews 19 papers (2021–2025); concludes integrated technical-social AI lifecycle ethics is lacking.	JISM 2026 (forthcoming DOI)
Krishnamurthy (2026) [NEW]	Article	Global AI ethics frameworks	Uses PCA/clustering on 112 AI policy documents to classify governance frameworks into three types; empirical complement to Table 7.	JISM 2026 (forthcoming DOI)

from conventional software engineering because data collection, model training, evaluation and monitoring are iterative and interdependent. The practical implication is clear: governance must follow the lifecycle, not merely approve the final tool (Lu *et al*., 2024; Raji *et al*., 2020).

Data Ethics

Data ethics concerns the moral principles and governance practices that guide data collection, processing, sharing, storage, analysis, reuse and deletion (Mittelstadt and Floridi, 2016). It asks whether data are collected lawfully, whether consent is meaningful, whether data subjects understand reuse, whether data are accurate and representative, whether sensitive groups are protected, whether benefits are fairly distributed, and whether affected people have avenues for challenge. Data ethics is broader than privacy. Privacy is central, but ethical data use also requires fairness, provenance, data quality, integrity, contextual integrity, accountability and respect for persons and communities (Alhassan *et al*., 2016; Abraham *et al*., 2019). The FAIR principles are valuable because

they emphasise stewardship and reuse rather than data accumulation (Wilkinson *et al*., 2016). In AI contexts, however, FAIR must be interpreted ethically. Accessible does not mean unrestricted. Reusable does not mean exploitable. Interoperable does not mean that data should be combined without regard to consent, vulnerability, local context or re-identification risk. The stronger interpretation is controlled usability: legitimate users should be able to discover, understand and lawfully request data, while sensitive data remain protected by governance controls (Gebru *et al*., 2021; Mehrabi *et al*., 2021).

Responsible AI

Responsible AI refers to the design, development, deployment and governance of AI systems in ways that advance beneficial outcomes while preventing or reducing foreseeable harm (Floridi and Cowls, 2019; UNESCO, 2021). It includes fairness, privacy, transparency, explainability, robustness, safety, accountability, human oversight, inclusiveness and sustainability. Responsible AI is not equivalent to ethical aspiration; it must be evidenced through documented processes including risk assessment, dataset documentation, model cards, human review, external audit, incident reporting, monitoring and redress (Lu *et al*., 2024; Raji *et al*., 2020).

Jobin *et al*. (2019) show that many AI ethics guidelines converge around similar principles but diverge in implementation. Fjeld *et al*. (2020) corroborate this finding across 36 ethical and rights-based AI documents globally. This finding is critical: the field does not need endless new lists of principles. It needs operational mechanisms that make principles enforceable. A bank using AI for credit assessment, a hospital using AI for diagnosis support and a university using AI for plagiarism detection all need more than a principle of fairness. They need measurable fairness definitions, subgroup testing, and documentation of training data, human review points, appeal channels and continuous monitoring (Mittelstadt, 2019; Morley *et al*., 2020; Barocas *et al*., 2023).

Theoretical Perspectives

Several ethical theories clarify responsible AI. Deontological ethics emphasises duties and rights, such that organisations must respect privacy, autonomy and due process even when automation increases efficiency. Utilitarian ethics asks whether AI systems maximise social benefit and minimise harm, but it must be constrained by rights because aggregate benefit can otherwise justify harm to minorities. Virtue ethics shifts attention to the character and judgement of developers, managers and institutions: responsible AI requires honesty, humility, prudence and courage to stop unsafe deployments (Dignum, 2019). Stakeholder theory is especially usefu because AI affects multiple groups beyond the system owner, including data subjects, end users, workers, regulators, communities and future users (Taddeo and Floridi, 2018). The most useful theoretical conclusion is that responsible AI is not a single-value ethics problem. It involves trade-offs among accuracy, fairness, privacy, explainability, security, cost, innovation and inclusion. Responsible AI therefore requires framing that asks: ethical for whom, under what evidence, compared with what alternative, using which data, with what monitoring, and with what remedy if the system fails? (Barocas *et al*., 2023; Mehrabi *et al*., 2021; Guidotti *et al*., 2018).

Principles of Responsible AI and Practical Applications

Table 5 converts major responsible AI principles into operational questions and possible controls. This conversion is necessary because ethics-washing often occurs when organisations state principles without changing procurement, design, documentation or monitoring routines (Mittelstadt, 2019; Lu *et al*., 2024).

Fairness is one of the most technically developed but philosophically difficult AI principles. Mehrabi *et al*. (2021) show that bias can enter AI through historical data, measurement choices, labels, sampling, algorithmic objectives and feedback loops. Barocas *et al*. (2023) further demonstrate that fairness metrics can conflict: equal error rates, demographic parity and individual fairness are not interchangeable. Buolamwini and Gebru (2018) demonstrated empirically through the Gender Shades study that commercial facial recognition systems showed large accuracy disparities across skin tone and gender, illustrating how inadequate training data diversity translates into real-world discrimination. The practical message is that organisations should not claim fairness without declaring which fairness standard is being used, why it is appropriate, and what trade-offs it creates.

Cybersecurity is inseparable from AI governance because adversarial attacks, data poisoning and model theft can undermine system integrity. The updated NIST Cybersecurity Framework 2.0 (NIST, 2024) integrates cybersecurity governance into enterprise risk management in ways that directly complement AI-specific risk controls and should be read alongside the AI RMF (NIST, 2023) in organisations deploying AI in high-stakes environments.

Human oversight is another area where vague statements are common. A human who merely clicks approval on an automated recommendation does not provide meaningful oversight. Oversight requires authority, competence, time, access to relevant information and protection from institutional pressure (Amershi *et al*., 2019; Raji *et al*., 2020). In high-stakes domains such as healthcare, finance, law enforcement, education and welfare, human review should be designed as a substantive safeguard rather than a symbolic compliance step.

Data Ethics in Artificial Intelligence

Ethical AI begins before model training. It begins when a

Table 5: Responsible AI principles and practical applications.

Principle	Operational question	Practical controls
Fairness and non-discrimination	Does the system produce unjustified differences across groups?	Bias testing, representative data, fairness metrics, impact assessment, appeal mechanisms (Mehrabi *et al*., 2021; Barocas *et al*., 2023).
Transparency	Can relevant stakeholders understand system purpose, limits and use?	Public notices, model cards, datasheets, user-facing explanations, transparency reports (Gebru *et al*., 2021; Mitchell *et al*., 2019).
Explainability	Can decisions be explained at the right level for users and regulators?	Interpretable models where possible, post-hoc explanations, reason codes (Guidotti *et al*., 2018; Wachter *et al*., 2017).
Privacy and data protection	Are data collected, processed and retained lawfully and proportionately?	Data minimisation, access controls, privacy impact assessments, encryption, retention rules (GDPR, 2016; NIST, 2020).
Accountability	Who is responsible for system outcomes and failures?	Named accountable owners, audit logs, incident reporting, third-party audits, sanctions (Raji *et al*., 2020; ISO/IEC, 2023).
Human oversight	Where must human judgement remain active?	Human-in-the-loop review, escalation thresholds, override authority, staff training (Amershi *et al*., 2019).
Safety and security	Can the system withstand misuse, adversarial attack or unsafe failure?	Red-teaming, secure development, monitoring, abuse testing, incident response (NIST, 2023; NIST, 2024).
Inclusiveness	Are marginalised users represented and protected?	Participatory design, accessibility testing, local language support, community consultation (UNESCO, 2021; African Union, 2024).
Sustainability	Are environmental and social costs proportionate?	Energy reporting, model-size justification, efficient computing, lifecycle assessment (Bender *et al*., 2021).

problem is selected, when a population is defined, when data sources are chosen, when labels are created and when decisions are made about what to exclude (Gebru *et al*., 2021; Mitchell *et al*., 2019; Crawford, 2021). Many AI projects fail ethically because the data lifecycle is treated as a technical pipeline rather than a governance pipeline. Responsible data collection requires lawful basis, clear purpose, proportionality and sensitivity to power asymmetries. In Nigeria and other developing countries, data subjects may formally consent but still lack real choice if data collection is tied to employment, education, healthcare or access to public services (Federal Republic of Nigeria, 2023; African Union, 2024). Data ownership and stewardship require careful distinction. Legal ownership may belong to an institution, but ethical stewardship requires duties to data subjects, communities and downstream users. In AI systems, stewardship includes data provenance, metadata, quality assessment, de-identification, access control, versioning, retention and deletion. Alhassan *et al*. (2016) and Abraham *et al*. (2019) show that data governance includes structural, procedural and relational dimensions. In practice, this means organisations need governance committees, defined data roles, policies, quality rules, escalation procedures and cross-functional collaboration. Consent remains central but insufficient. AI development often involves secondary use, model training and future applications that cannot be fully described at the point of collection. Dynamic consent may increase control but can exclude people without digital access or literacy. Therefore, consent must be complemented by public interest tests, ethics review, privacy safeguards and independent oversight as envisaged by both the GDPR (European Parliament and Council, 2016) and the Nigeria Data Protection Act (Federal Republic of Nigeria, 2023). Data quality and data integrity are ethical issues, not merely technical ones (Mehrabi *et al*., 2021; Mittelstadt *et al*., 2016). Missing data can reflect social exclusion; inaccurate data can punish individuals; outdated data can reproduce past conditions; and proxy variables can encode protected characteristics indirectly. Ethical data governance therefore requires quality metrics, documentation of limitations, and refusal to deploy models where data are inadequate for the intended use.

Ethical Challenges of AI across Sectors

Higher education illustrates both the promise and danger of AI. Generative AI can support feedback, translation, brainstorming, accessibility and formative learning, but it also raises concerns about academic integrity, hallucinated references, unequal access, surveillance and deskilling (Kasneci *et al*., 2023; Bender *et al*., 2021). The responsible response is not a blanket ban or uncritical adoption. Institutions need AI literacy, assessment redesign, disclosure rules, privacy guidance and disciplinary judgement.

Healthcare shows why accuracy alone is insufficient. AI

Table 6: Ethical challenges of AI across sectors.

Challenge	Typical manifestation	High-risk sectors	Mitigation
Algorithmic bias	Unequal false positives, false negatives or access outcomes.	Finance, hiring, policing, education, healthcare.	Representative data, subgroup evaluation, fairness review, appeals (Mehrabi *et al*., 2021; Buolamwini & Gebru, 2018).
Privacy violation	Excessive collection, weak consent, re-identification, data leakage.	Healthcare, education, government, research.	Data minimization, encryption, privacy impact assessment, retention controls (GDPR, 2016; NIST, 2020).
Opacity	Users cannot understand or challenge automated decisions.	Credit, welfare, insurance, education.	Explainable models, reason codes, documentation, human review (Guidotti *et al*., 2018; Wachter *et al*., 2017).
Surveillance	Continuous tracking of learners, workers or citizens.	Education, workplace, public security.	Necessity/proportionality tests, oversight, strict access controls.
Misinformation and hallucination	AI-generated false claims, fabricated references or persuasive synthetic media.	Media, education, governance, science.	Verification routines, labelling, source grounding, media literacy (Bender *et al*., 2021; Kasneci *et al*., 2023).
Deepfakes and synthetic manipulation	Fake audio, video or images used for fraud, abuse or political manipulation.	Media, politics, finance, security.	Content provenance, detection tools, public awareness, legal remedies.
Intellectual property disputes	Training or output may use protected material without clarity.	Publishing, media, education, software.	Rights review, licensing, attribution policies, legal guidance.
Employment displacement	Automation changes job tasks, wages and bargaining power.	Customer service, finance, logistics, administration.	Reskilling, worker consultation, transition planning.
Autonomous decision-making	Systems act without meaningful human judgement.	Vehicles, defence, healthcare, public services.	Human control, fail-safe design, risk thresholds, audit trails (Raji *et al*., 2020).

systems may assist diagnosis, triage, imaging and resource allocation, but they operate in contexts where errors can cause serious harm. WHO (2021) emphasises autonomy, safety, transparency, responsibility, inclusiveness and sustainability in AI for health. Banking and finance show how AI can reproduce structural inequality. Credit scoring and fraud detection may improve efficiency, but biased data can penalise people with thin credit histories, informal income or residence in underserved communities (Barocas *et al*., 2023; Mittelstadt *et al*., 2016). Government use of AI raises legitimacy concerns. Automated decision-making in welfare, taxation, policing, immigration and public-service allocation can affect rights and livelihoods. Public agencies have duties that private firms may not share: legality, fairness, transparency, due process and democratic accountability. Scientific research benefits from AI in literature review, data analysis, simulation and discovery. Yet AI can also intensify reproducibility problems if models, prompts, data preprocessing and evaluation procedures are not documented (Gebru *et al*., 2021; Mitchell *et al*., 2019). Libraries and information services occupy a strategic position as mediators of knowledge access; information professionals are important actors in AI literacy, metadata stewardship and misinformation resistance (Table 6).

AI Governance and Regulatory Frameworks

AI governance refers to the rules, institutions, standards, processes and accountability mechanisms used to guide AI development and use (Table 7). The current governance landscape is plural rather than unified, with hard law, soft law, technical standards, organisational policies, ethics committees, procurement requirements, documentation practices, certification schemes, audits and public engagement all playing roles (Cath, 2018; OECD, 2019). The EU AI Act is the most developed binding AI-specific regulatory framework currently in force. It adopts a risk-based structure with prohibited AI practices, obligations for high-risk systems, transparency duties and rules for general-purpose AI models (European Parliament and Council of the European Union, 2024). The NIST AI Risk Management Framework provides a more flexible approach organised around Govern, Map, Measure and Manage (NIST, 2023). ISO/IEC 42001 moves responsible AI into management-system logic, asking organisations to establish, implement, maintain and continually improve an

Table 7: Comparative overview of major AI governance frameworks.

Framework	Type	Main contribution	Critical limitation
GDPR (2016)	Binding EU regulation	Provides strong rights-based rules for personal data processing, including lawful basis, transparency and data subject rights.	Not designed specifically for all AI-specific risks, though highly relevant to AI data processing.
OECD AI Principles (2019)	Intergovernmental recommendation	Sets values-based principles for trustworthy AI and national policy guidance.	Soft-law instrument; implementation depends on adherent countries.
UNESCO Recommendation (2021)	Global normative instrument	Frames AI ethics through human rights, inclusion, environmental sustainability and governance.	Broad principles require local translation and enforcement capacity.
WHO AI health guidance (2021)	Sectoral guidance	Provides principles and governance considerations for AI in health.	Health-specific; implementation varies with health system capacity.
NIST AI RMF (2023)	Voluntary framework	Organises AI risk management around Govern, Map, Measure and Manage.	Voluntary; does not by itself certify compliance.
ISO/IEC 42001 (2023)	Management-system standard	Specifies requirements for establishing and improving AI management systems.	Certification may become checklist-driven unless tied to substantive outcomes.
EU AI Act (2024)	Binding EU regulation	Creates risk-based obligations for prohibited, high-risk, transparency and general-purpose AI systems.	Complexity may challenge small organizations and non-EU suppliers.
Nigeria Data Protection Act (2023)	National legislation	Establishes rights-based rules for personal data processing in Nigeria, complementing the AI governance agenda.	Implementation capacity and regulatory enforcement remain developing.
African Union AI Strategy (2024)	Continental policy strategy	Supports Africa-centric, ethical, inclusive and development-focused AI adoption.	Requires national implementation, funding, skills and infrastructure.
Nigeria National AI Strategy (2024)	National strategy	Articulates responsible, ethical and inclusive AI innovation for national development.	Needs sustained implementation, regulatory coordination and measurable accountability.

AI management system (ISO/IEC, 2023). Krishnamurthy (2026), using data-driven analysis of 112 AI policy documents, classifies global governance frameworks into three types technical and governance leaders, thematic and generalist adopters, and comprehensive adopters providing an empirical typology that complements the comparative analysis in (Table 7).

Proposed Integrated Framework for Data Ethics and Responsible AI Use

The proposed framework has six layers. It is explicitly a conceptual proposal grounded in the reviewed evidence and is offered as a foundation for future empirical validation rather than as a tested or certified governance programme. First, ethical leadership establishes institutional commitment: responsible AI must be owned at senior level because it affects legal exposure, public trust, product quality, human rights and strategic legitimacy (Dignum, 2019). Second, the data ethics layer governs input conditions: lawful collection, purpose limitation, data minimisation, consent or lawful alternative, provenance records, data quality checks, representativeness assessment, metadata, access control and deletion rules (Gebru *et al*., 2021; Abraham *et al*., 2019). Third, the model governance layer covers design, training, testing and documentation: model cards, datasheets, fairness testing, explainability assessment, robustness testing, security review and red-teaming (Mitchell *et al*., 2019; Guidotti *et al*., 2018). Fourth, deployment controls ensure responsible use in real organisational settings: user training, human review, escalation thresholds, override authority, access logs, incident reporting and operational monitoring (Raji *et al*., 2020; Lu *et al*., 2024). Fifth, stakeholder assurance makes accountability visible through transparency reports, complaint mechanisms, external audits, community engagement and public-interest review (UNESCO, 2021; ISO/IEC, 2023). Sixth, continuous learning recognises that AI systems change over time: models can drift, populations can shift, adversarial behaviour can emerge and laws can evolve (Table 8). Responsible AI therefore requires periodic review, performance monitoring, retraining governance, retirement criteria and feedback loops (Amershi *et al*., 2019; NIST, 2023).

Implications

Policy implications: Governments should move from broad AI ambition statements to enforceable governance capacity. This means clear data protection laws, AI risk classification, procurement standards, regulatory

Table 8: Operational checklist for the proposed responsible AI framework.

Framework layer	Minimum evidence to document	Responsible actors
Ethical leadership	AI policy, risk appetite, governance committee minutes, board reporting.	Board, executive management, legal, ethics office.
Data ethics	Data inventory, lawful basis, consent records, provenance, quality report, privacy impact assessment (NIST, 2020; Federal Republic of Nigeria, 2023).	Data stewards, data protection officer, research ethics committee.
Model governance	Model card, datasheet, test results, bias assessment, security review, intended-use statement (Mitchell *et al*., 2019; Gebru *et al*., 2021).	Data scientists, ML engineers, product owners, risk team.
Deployment controls	Training records, human oversight workflow, access logs, incident process, escalation rules (Raji *et al*., 2020).	Operations, users, managers, IT security.
Stakeholder assurance	Transparency report, complaint mechanism, audit report, consultation records (ISO/IEC, 2023).	Compliance, civil society partners, external auditors, regulators.
Continuous learning	Drift monitoring, periodic review, retraining logs, decommissioning criteria (NIST, 2023; NIST, 2024).	Model owners, monitoring team, governance committee.

sandboxes with safeguards, public-sector AI registers and funding for audit expertise (OECD, 2019; European Parliament and Council, 2024). Nigeria's National AI Strategy (Federal Republic of Nigeria/NITDA, 2024) and the Nigeria Data Protection Act (Federal Republic of Nigeria, 2023) provide important starting points, but must be backed by regulatory capacity, measurable targets and budg*et al*location. Developing countries should avoid copying foreign frameworks mechanically; they need context-sensitive governance addressing local languages, informal economies, infrastructure gaps and community rights (African Union, 2024).

Educational implications: AI literacy should include data ethics, source verification, bias detection, privacy protection, prompt documentation, responsible citation and critical judgement (Kasneci *et al*., 2023; UNESCO, 2021). Students should not be taught merely how to use AI tools; they should be taught when not to use them, how to verify them and how to recognise limits. Kerasidou *et al*. (2025) illustrate how even well-resourced academic institutions face significant challenges in embedding AI ethics into governance processes, underscoring the need for institutional investment in AI ethics education rather than mere policy statements.

Organisational implications: Organisations should treat AI governance as enterprise risk management. Every AI system should have a named owner, a data provenance record, a risk rating, documentation, user training, monitoring and incident response (Raji *et al*., 2020; Lu *et al*., 2024). Maslej *et al*. (2025) report that fewer than 5% of organisations using AI have implemented a formal governance framework, despite 96% adoption a finding that makes the operational checklist in Table 8 directly actionable. Technological implications: Developers should embed responsible-AI-by-design patterns into architecture, including logging, explainability interfaces, consent controls, privacy-enhancing techniques, role-based access, evaluation dashboards and monitoring alerts (Lu *et al*., 2024; NIST, 2023). Research implications are discussed in the following section.

DISCUSSION

The evidence confirms that responsible AI is now a mature discourse at the level of principles but an uneven practice at the level of implementation. The strongest recurring principles are fairness, transparency, accountability, privacy, safety and human oversight. The weakest recurring area is proof: many frameworks tell organisations what to value; fewer provide rigorous ways to demonstrate that those values changed outcomes. This is why documentation, auditing and monitoring are central they convert ethical intention into inspectable evidence (Raji *et al*., 2020; Lu *et al*., 2024; Morley *et al*., 2020). The relationship between data ethics and responsible AI is foundational. Data ethics controls the moral and epistemic quality of inputs, while responsible AI controls the transformation of inputs into outputs and decisions. A model trained on unethical data cannot become responsible merely because it has a user-friendly interface (Mittelstadt *et al*., 2016; Gebru *et al*., 2021; Crawford, 2021). Conversely, ethically collected data can still be

misused through opaque, unsafe or discriminatory AI deployment. The two domains must therefore be governed together. More recent evidence reinforces the central implementation-gap finding. The Stanford HAI AI Index Report (Maslej *et al*., 2025) documents a sharp increase in AI-specific regulation globally 59 federal and 131 state-level AI rules introduced in the United States in 2024 alone alongside a concurrently rising rate of documented AI incidents. This quantitative evidence independently corroborates this review's argument that principle-adoption is outpacing operational governance. Simultaneously, governance survey evidence compiled in industry reports (2025) finds that while approximately 96% of organisations now use AI, fewer than 5% report having fully implemented a formal AI governance framework a gap that makes the operational checklist in Table 8 directly relevant and urgently actionable.

Kerasidou *et al*. (2025) provide an institutional case study from Oxford University demonstrating the practical difficulty of translating AI ethics principles into research ethics governance processes, even within a well-resourced academic environment. Their findings confirm that the implementation gap is not primarily a resource problem; it is a structural and cultural problem requiring deliberate organisational investment in accountability roles, process design and ongoing monitoring. The most closely parallel recent review to the present work Yesán-Luján *et al*. (2026) reviews 19 peer-reviewed articles from 2021 to 2025 and similarly concludes that an integrated technical-social approach to AI lifecycle ethics is lacking. Krishnamurthy (2026) offers an empirical complement by classifying 112 AI policy documents into three governance types (technical and governance leaders, thematic and generalist adopters, comprehensive adopters), providing a quantitative typology that triangulates with and could refine the comparative analysis in Table 7 of this review. The review also shows that developing-country contexts deserve more than a footnote. AI systems are often developed in countries with stronger infrastructure and larger datasets, then exported into contexts with different languages, institutional realities and social risks creating the danger of algorithmic dependency (African Union, 2024; Federal Republic of Nigeria/NITDA, 2024). The concern that strict governance may slow innovation in low-resource countries is valid, but the answer is proportionate, staged and capacity-building-oriented governance, not governance removal.

Future Research Directions

Future research should become more empirical. The next phase of the field should evaluate governance interventions in practice: whether AI registers improve transparency, whether model cards are read by decision-makers, whether users understand explanations, whether audits change deployment decisions, and whether complaint mechanisms produce remedies (Raji *et al*., 2020; Lu *et al*., 2024). The proposed six-layer framework in this paper is explicitly a conceptual proposal; its empirical validation across sectors, institution types and regional contexts is an identified research priority, as indicated in Table 9 above (Kerasidou *et al*., 2025; Yesán-Luján *et al*., 2026). Mixed-methods designs combining technical evaluation with interviews, case studies and stakeholder experience will be especially useful. Research should also examine local AI ecosystems in Africa, Asia, Latin America and small island states, where different ethical priorities from the dominant literature including language exclusion, infrastructure fragility, informal labour and procurement dependence create distinct responsible AI challenges (Table 9).

Table 9: Future research agenda on data ethics and responsible AI.

Priority area	Future research question
Ethical AI in developing countries	How do infrastructure, local datasets and regulatory capacity shape responsible AI implementation? (African Union, 2024)
AI governance in Africa	Which continental and national governance models are effective in African contexts?
Responsible AI in higher education	Which AI literacy policies reduce misuse while preserving learning benefits? (Kasneci *et al*., 2023; Kerasidou *et al*., 2025)
Cross-cultural AI ethics	How do local values, languages and communal rights reshape global AI principles?
AI audit effectiveness	Do algorithmic audits reduce measurable harm after deployment? (Raji *et al*., 2020)
Human–AI collaboration	What forms of human oversight produce meaningful accountability rather than rubber-stamping?
Generative AI governance	How can hallucination, deepfakes, privacy leakage and copyright risks be governed practically? (Bender *et al*., 2021; Bommasani *et al*., 2021)
Sustainability	How should environmental costs of model training and deployment be measured and disclosed?
Framework validation	Does the proposed six-layer integrated framework reduce measurable AI harm or improve trust when applied across sectors? What context-specific adaptations are needed? (Kerasidou *et al*., 2025; Yesán-Luján *et al*., 2026)
Empirical governance typology	Can Krishnamurthy's (2026) three-type governance classification be replicated and extended with qualitative case studies, and does governance type predict implementation outcomes?

Limitations

This manuscript has limitations that should be understood before applying its findings. First, it is a structured evidence-map draft rather than a registered systematic review with independently screened database exports. For full methodological rigour as a scoping review, the search strategy should be rerun in Scopus, Web of Science or similar databases, independently screened by at least two reviewers and documented with a complete PRISMA-ScR flow diagram before final journal submission. This limitation is disclosed in the abstract, the methodology section and here, per reviewer guidance. Second, the evidence base was restricted to English-language sources from 2016 to 2026 and may exclude relevant non-English scholarship, particularly from Chinese, French, Arabic and Portuguese AI governance literature. Third, AI governance is evolving rapidly; some policy instruments may change after the drafting period. Fourth, the review includes official frameworks and standards alongside peer-reviewed literature, which is appropriate for governance synthesis but means the evidence is heterogeneous in rigour and independence. Fifth, the proposed integrated framework has not been piloted or empirically validated; it is a conceptual proposal requiring future testing.

Conclusion

Data ethics is not a side issue in responsible AI; it is the foundation on which responsible AI rests. AI systems learn from data, classify people through data, make predictions from data and shape decisions through data. If the data lifecycle is unethical, the AI system built on it cannot be trustworthy (Mittelstadt *et al*., 2016; Gebru *et al*., 2021; Crawford, 2021). At the same time, ethical data alone is insufficient. Responsible AI requires governance across the full lifecycle: leadership, data stewardship, model documentation, risk assessment, human oversight, auditability, stakeholder engagement and continuous monitoring (Lu *et al*., 2024; Raji *et al*., 2020; ISO/IEC, 2023). The strongest conclusion of this review is that the field must move from principle abundance to implementation discipline. Fairness, transparency, accountability and privacy are now familiar terms. The harder question is whether an organisation can prove that these values are embedded in practice: who is accountable, what data were used, what groups were tested, what decisions can users appeal, what happens when the system fails, how is drift monitored, and which harms trigger suspension or withdrawal? (Mittelstadt, 2019; Morley *et al*., 2020). The evidence from 2025 and 2026 including the Stanford HAI AI Index (Maslej *et al*., 2025), the Oxford ethics review case study (Kerasidou *et al*., 2025), and the parallel reviews by Yesán-Luján *et al*. (2026) and Krishnamurthy (2026) confirms that these questions remain unanswered in most organisations, and that answering them is the defining challenge of the responsible AI field. Responsible AI is therefore a socio-technical governance project. For developing countries, the challenge is especially sharp: AI adoption must avoid both technological rejection and technological dependency. The goal should be locally grounded, rights-respecting and capacity-building AI ecosystems. The proposed six-layer framework offered as a conceptual contribution requiring future empirical validation provides a practical starting point for researchers, policymakers, developers and organisational leaders seeking to integrate data ethics and responsible AI governance into coherent, accountable and continuously improving institutional practice (African Union, 2024; Federal Republic of Nigeria/NITDA, 2024; UNESCO, 2021; NIST, 2023).

REFERENCES

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information Management, 49, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

African Union. (2024). Continental artificial intelligence strategy. African Union Commission. https://au.int/en/documents/20240217/continental-artificial-intelligence-strategy

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. Journal of Decision Systems, 25(S1), 64–75. https://doi.org/10.1080/12460125.2016.1187397

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice, 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31

Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and machine learning: Limitations and opportunities. MIT Press. https://fairmlbook.org/

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., *et al*. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258. https://arxiv.org/abs/2108.07258

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A, 376(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Dignum, V. (2019). Responsible artificial intelligence: How to develop and use AI in a responsible way. Springer. https://doi.org/10.1007/978-3-030-30371-6

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

European Parliament and Council of the European Union. (2016).

Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation). Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689

Federal Republic of Nigeria. (2023). Nigeria Data Protection Act, 2023. Federal Government of Nigeria.

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

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daume III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92. https://doi.org/10.1145/3458723

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), Article 93. https://doi.org/10.1145/3236009

International Organization for Standardization. (2023). ISO/IEC 42001:2023 Artificial intelligence management system. https://www.iso.org/standard/81230.html

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

Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kerasidou, X., Kingori, P., Parker, M., & Tansey, J. (2025). From principles to practice: An ethics review of artificial intelligence research. Research Ethics, 21(1), 1–18. https://doi.org/10.1177/17470161241248759

Krishnamurthy, V. (2026). Data-driven investigation of global AI ethics frameworks and governance patterns. Journal of Information Systems Management, 43(1), 88–112.

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2024). Responsible AI pattern catalogue: A collection of best practices for AI governance and engineering. ACM Computing Surveys, 56(7), Article 172. https://doi.org/10.1145/3626234

Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., & Clark, J. (2025). The AI Index 2025 annual report. Stanford University Human-Centered Artificial Intelligence. https://aiindex.stanford.edu/report/

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), Article 115. https://doi.org/10.1145/3457607

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229. https://doi.org/10.1145/3287560.3287596

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1, 501–507. https://doi.org/10.1038/s42256-019-0114-4

Mittelstadt, B., & Floridi, L. (2016). The ethics of big data: Current and foreseeable issues in biomedical contexts. Science and Engineering Ethics, 22(2), 303–341. https://doi.org/10.1007/s11948-015-9652-2

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

Morley, J., Cowls, J., Taddeo, M., & Floridi, L. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to support responsible AI system development. Philosophy & Technology, 33(4), 409–437. https://doi.org/10.1007/s13347-019-00368-3

National Institute of Standards and Technology. (2020). NIST Privacy Framework: A tool for improving privacy through enterprise risk management, Version 1.0. https://doi.org/10.6028/NIST.CSWP.01162020

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. https://doi.org/10.6028/NIST.AI.100-1

National Institute of Standards and Technology. (2024). The NIST Cybersecurity Framework (CSF) 2.0. NIST CSWP 29. https://doi.org/10.6028/NIST.CSWP.29

Nigeria National Centre for Artificial Intelligence and Robotics/NITDA. (2024). National artificial intelligence strategy. https://ncair.nitda.gov.ng/

Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments, OECD/LEGAL/0449. https://doi.org/10.1787/eedfee77-en

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., *et al*. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. JBI Evidence Synthesis, 18(10), 2119–2126. https://doi.org/10.11124/JBIES-20-00167

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33–44. https://doi.org/10.1145/3351095.3372873

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 429–435. https://doi.org/10.1145/3306618.3314244

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. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841–887. https://doi.org/10.2139/ssrn.3063289

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N. (2016). The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

World Economic Forum. (2025). Advancing responsible AI innovation: A playbook. World Economic Forum. https://www.weforum.org/publications/advancing-responsible-ai-innovation-a-playbook/

World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. WHO. https://www.who.int/publications/i/item/9789240029200

Yesán-Luján, A., Diaz-Tomas, S., & Mendoza-De los Santos, H. E. (2026). Ethical technologies for integration into artificial intelligence and computer science fields: A descriptive review focused on governance. International Journal of Systems Management, 1(1), 1–19.