Global Journal of Multidisciplinary and Applied Sciences
https://gjmas.com/index.php/gjmas
E- ISSN: 2313-6685
VOL 01 ISSUE 01 2023

Ethical and Governance Challenges in Artificial Intelligence and Emerging Technologies

Ayesha Riaz
Department of Computer Science Quaid-i-Azam University, Islamabad, Pakistan
Email: *ayesha.riaz@qau.edu.pk*
Muhammad Bilal Khan
School of Electrical Engineering and Computer Science (SEECS) National University of Sciences and Technology (NUST), Islamabad, Pakistan
Email: *bilal.khan@seecs.nust.edu.pk*

Abstract:

Artificial Intelligence (AI) and emerging technologies (e.g., Internet of Things, biometrics, and autonomous systems) are rapidly reshaping decision-making in government, industry, and everyday life. Alongside benefits, these systems introduce ethical risks—privacy intrusion, bias and discrimination, opacity, safety failures, labor displacement, and concentration of power—while exposing governance gaps in accountability, auditing, and enforcement. This paper synthesizes major ethical challenges across the AI lifecycle (data, model design, deployment, monitoring) and proposes a practical governance framing grounded in internationally recognized standards and policy instruments. We argue that effective governance requires risk-based regulation, organizational controls (impact assessments, audits, incident reporting), and socio-technical safeguards (human oversight, contestability, and transparency). We conclude with an implementation-oriented roadmap that integrates AI risk management practices with rights-based ethics to support trustworthy, socially beneficial innovation.

Keywords: AI ethics, governance, accountability, algorithmic bias, privacy, transparency, risk management, regulation

INTRODUCTION

AI systems increasingly mediate access to jobs, credit, education, healthcare, and public services domains where errors and unfairness translate into real harm. The governance problem is not only technical; it is institutional and political: who sets objectives, whose values are embedded, who bears risk, and who is accountable when harm occurs. Global frameworks increasingly converge on "trustworthy AI," emphasizing human rights, fairness, transparency, robustness, and accountability (e.g., UNESCO's ethics recommendation; OECD principles; NIST's AI risk framework; and risk-based regulation such as the EU AI Act). At the same time, the pace of emerging technologies—foundation models, connected sensors, biometric identification, and automated decision systems—creates new threat surfaces (cybersecurity, model misuse, surveillance) and new governance needs (auditable logs, oversight capacity, and

cross-border coordination). This article maps ethical risks, clarifies accountability across the AI lifecycle, and presents a governance toolkit that institutions can apply in both public and private sectors.

Privacy, Surveillance, and Data Governance:

AI-driven systems significantly intensify data extraction and surveillance capacities, particularly in digitally interconnected environments such as smart cities, wearable health devices, biometric identification systems, and Internet of Things (IoT) infrastructures. These technologies enable continuous, passive, and large-scale data collection, often extending beyond explicitly provided information to inferred attributes such as behavioral patterns, health conditions, social relationships, or political inclinations. Ethical risks become acute when data is repurposed across contexts without meaningful user consent, when anonymization techniques fail under advanced re-identification methods, or when multiple datasets are combined to construct detailed personal profiles. Traditional "notice-and-consent" models are increasingly inadequate in this setting, as individuals lack real understanding or control over complex data flows. Consequently, governance frameworks must prioritize purpose limitation, data minimization, proportionality, and lifecycle-based oversight to prevent function creep and systemic privacy erosion. UNESCO's AI ethics framework situates privacy within a broader human rights perspective, emphasizing human dignity, autonomy, transparency, and safeguards against mass surveillance, discriminatory profiling, and misuse by both state and private actors.From an implementation perspective, robust privacy and data governance requires institutionalized mechanisms rather than ad hoc compliance. Organizations deploying AI should maintain comprehensive data inventories and lineage documentation to ensure traceability and accountability across the data lifecycle. Differential access controls, strong encryption standards, and cybersecurity measures are essential to protect sensitive information from unauthorized use or breaches. Governance must also extend to third parties through vendor management and data-sharing agreements that include audit rights, security obligations, and limitations on secondary use. For high-stakes or large-scale deployments, mandatory privacy and algorithmic impact assessments can identify risks to individual rights before harm occurs. Finally, effective redress mechanisms—such as complaint procedures, independent oversight bodies, and avenues for human review—are critical to restoring trust and ensuring that individuals can challenge misuse, errors, or unfair surveillance practices in AI-enabled systems.

Bias, Discrimination, and Fairness in Automated Decisions:

Bias and discrimination in automated decision-making systems arise from multiple sources across the AI lifecycle, including historically skewed datasets, biased labeling practices, measurement errors, and the use of proxy variables that indirectly encode protected characteristics such as gender, ethnicity, socioeconomic status, or geography. In high-stakes domains—such as hiring, credit scoring, predictive policing, healthcare triage, and welfare eligibility—reliance on aggregate performance metrics like overall accuracy can obscure systematic harms experienced by specific groups, including elevated false-negative or false-positive rates. Ethical governance therefore requires context-sensitive definitions of fairness, recognizing that different applications may demand distinct fairness objectives (e.g., equality of opportunity, demographic parity, or error-rate balance). Importantly, fairness is not solely a technical optimization problem but also a normative and policy choice that must align with

legal standards, social values, and institutional responsibilities. The OECD's AI Principles underscore this human-centred approach, emphasizing fairness, inclusiveness, and accountability as foundational elements of trustworthy AI.

Effective mitigation of algorithmic bias demands sustained organizational commitment rather than one-time technical fixes. Best practices include the use of representative and high-quality training data, participatory data collection where feasible, and rigorous external validation across diverse populations and operational contexts. Bias mitigation techniques—such as data reweighting, algorithmic constraints, and post-processing adjustments—should be complemented by transparency about trade-offs and residual risks. Continuous monitoring is essential, as models can become unfair over time due to data drift, behavioral change, or shifting institutional practices. Governance mechanisms such as periodic audits, bias impact assessments, documentation of fairness decisions, and clear accountability for corrective action are critical to ensuring that automated systems remain equitable, lawful, and socially acceptable throughout their deployment lifecycle.

Transparency, Explainability, and Contestability:

Transparency and explainability are foundational to ethical governance of AI systems, particularly as complex deep learning architectures and foundation models increasingly operate as "black boxes" whose internal logic is difficult even for developers to fully interpret. However, transparency in governance does not require full technical disclosure of source code or model weights; rather, it demands that AI systems be understandable at an appropriate level for different stakeholders, including policymakers, auditors, domain experts, and affected individuals. This includes clarity about a system's purpose, decision logic, data sources, performance limitations, and known risks. UNESCO's ethical framework emphasizes transparency and human oversight as essential safeguards to preserve human dignity, prevent arbitrary decision-making, and ensure that AI augments rather than replaces human judgment. Without such safeguards, opaque systems risk eroding trust, enabling unaccountable power, and undermining procedural fairness, particularly in public administration and high-stakes regulatory contexts.

A governance-ready approach operationalizes transparency through standardized documentation and institutional processes rather than ad hoc explanations. Tools such as model cards, data sheets for datasets, and algorithmic registers provide structured disclosures about training data, intended use, evaluation metrics, bias considerations, and operational constraints. Clear communication of uncertainty—such as confidence intervals, error rates, and conditions under which outputs may be unreliable—helps decision-makers avoid overreliance on automated recommendations. Maintaining logs of automated decisions enables traceability, post-hoc auditing, and incident investigation. Crucially, contestability mechanisms must be embedded into system design and organizational policy, allowing individuals to challenge automated outcomes, request human review, and receive meaningful explanations that support due process. In regulated sectors and public services, such mechanisms are essential to upholding accountability, protecting rights, and ensuring that AI-enabled decisions remain subject to human judgment and democratic oversight.

Accountability, Liability, and Institutional Oversight:

Accountability and liability represent some of the most complex governance challenges in AI deployment, as automated systems often involve multiple actors across the value chain, including data providers, model developers, platform vendors, system integrators, and end-user

organizations. This multi-actor ecosystem frequently produces "responsibility gaps," where harm caused by an AI system cannot be clearly attributed to a single accountable party. Effective AI governance therefore requires explicit allocation of roles and responsibilities across the entire system lifecycle—from design and data acquisition to deployment, monitoring, and retirement. These responsibilities must be operationalized through concrete controls such as approval and escalation gates, formal risk acceptance by accountable executives, clearly defined monitoring obligations, and mandatory incident reporting and remediation processes. The NIST Artificial Intelligence Risk Management Framework (AI RMF) reinforces this approach by framing AI governance as a continuous process of mapping risks, measuring impacts, and managing residual risk, rather than a one-time compliance exercise. Institutional oversight mechanisms are essential to translate accountability principles into enforceable practice. Organizations can establish dedicated AI governance boards or ethics committees with cross-functional representation from technical, legal, operational, and domain experts to oversee high-risk deployments and adjudicate trade-offs between innovation and risk. Mandatory algorithmic or impact assessments prior to deployment can help identify potential harms, affected stakeholders, and mitigation strategies, while independent audits provide external validation of claims related to fairness, robustness, security, and compliance. Procurement policies also play a critical governance role by requiring vendors to provide documentation, testing evidence, transparency artifacts, and ongoing monitoring commitments as contractual obligations. Together, these institutional mechanisms help ensure that accountability is embedded into organizational structures, that liability is traceable and enforceable, and that AI systems remain subject to meaningful human oversight throughout their operational lifespan

Regulation, Safety, and the Governance of General-Purpose:

Regulation of artificial intelligence is increasingly converging around a risk-based governance model that calibrates legal obligations to the potential severity and scale of harm posed by different AI applications. A prominent example is the European Union's Artificial Intelligence Act, which entered into force on 1 August 2024 and introduces a tiered regulatory structure distinguishing between unacceptable-risk, high-risk, limited-risk, and minimal-risk AI systems, with phased implementation timelines for compliance. This approach reflects a broader international consensus that blanket regulation is neither effective nor innovation-friendly; instead, governance must focus regulatory scrutiny on systems that affect fundamental rights, public safety, or critical infrastructure. By embedding requirements such as risk management, data governance, human oversight, and post-market monitoring, the EU AI Act signals a shift toward lifecycle-based regulation that extends beyond pre-deployment certification to ongoing operational accountability.

The governance of emerging and general-purpose AI systems—including large foundation models capable of multi-domain deployment—poses additional challenges related to scale, dual-use potential, and systemic risk. Such systems can be repurposed across sectors, increasing the likelihood of misuse, unintended consequences, and cascading failures. Effective governance therefore requires safeguards such as abuse monitoring, controlled access to high-risk capabilities, robust cybersecurity protections, and standardized evaluation protocols to assess safety, bias, robustness, and alignment before and after deployment. At the policy level, ongoing debates highlight tensions between rapid innovation and regulatory burden, particularly for small and medium-sized enterprises and public-sector institutions with

limited compliance capacity. These tensions underscore the importance of clear regulatory guidance, proportional obligations, regulatory sandboxes, and technical assistance mechanisms that enable responsible experimentation while maintaining public trust, safety, and democratic oversight in the governance of advanced AI technologies.

Summary:

Artificial Intelligence and emerging technologies present significant ethical and governance challenges that extend beyond technical performance to fundamental questions of rights, accountability, and societal trust. As AI systems increasingly shape decisions in public services, markets, and everyday life, risks related to privacy intrusion, surveillance, bias, discrimination, opacity, and safety become more pronounced. Traditional governance approaches—such as notice-and-consent for data use or one-time compliance checks—are insufficient in the face of continuous data collection, complex model behavior, and rapidly evolving deployment contexts. Effective governance therefore requires a shift toward lifecycle-based, risk-oriented frameworks that integrate ethical principles with operational controls.This article highlights five core governance dimensions: privacy and data governance, fairness and non-discrimination, transparency and contestability, accountability and institutional oversight, and regulation and safety for general-purpose AI. International frameworks from UNESCO, the OECD, NIST, and the European Union increasingly converge on human-centred, rights-based principles, while emphasizing practical mechanisms such as impact assessments, documentation, audits, monitoring, and human oversight. Risk-based regulation—exemplified by the EU AI Act—seeks to balance innovation with protection by focusing obligations on high-risk and systemic AI uses. Ultimately, trustworthy AI depends on embedding ethical governance into organizational structures, procurement practices, and regulatory systems, ensuring that technological advancement proceeds alongside social responsibility, legal accountability, and public trust.

References:

UNESCO, Recommendation on the Ethics of Artificial Intelligence, Paris, France: United Nations Educational, Scientific and Cultural Organization, 2021.

UNESCO, Ethics of Artificial Intelligence: Human Rights and Governance Framework, Paris, France: United Nations Educational, Scientific and Cultural Organization, 2023.

Organisation for Economic Co-operation and Development (OECD), OECD Principles on Artificial Intelligence, Paris, France, 2019.

Organisation for Economic Co-operation and Development (OECD), Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449), Paris, France, 2019.

National Institute of Standards and Technology (NIST), Artificial Intelligence Risk Management Framework (AI RMF 1.0), Gaithersburg, MD, USA: U.S. Department of Commerce, 2023.

National Institute of Standards and Technology (NIST), AI Risk Management Framework Playbook, Gaithersburg, MD, USA: U.S. Department of Commerce, 2023.

European Commission, Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), ,2024.

European Commission, The EU Artificial Intelligence Act: Risk-Based Approach and Implementation Timeline, 2024.

Associated Press, “EU releases code of practice for general-purpose AI ahead of AI Act enforcement,” AP News, 2025.

Reuters, “EU rejects calls to delay AI rules and confirms phased rollout,” Reuters, 2025.

World Economic Forum, Global AI Governance: Aligning Innovation with Responsibility, Geneva, Switzerland, 2023.