RECONCEPTUALISING ACCOUNTABILITY IN ARTIFICIAL INTELLIGENCE INSTITUTIONAL FRAMEWORKS, GAPS, AND GOVERNANCE MECHANISM

Gauri Lamba, Student, BBA LL.B. (Hons.), Symbiosis Law School, Pune

ABSTRACT

With the advent of artificial intelligence integration in every sector today, concerns about accountability are of paramount importance. The paper delves into the concepts of reliability and Accountability[1] that of an entity for the parts they play across the project that they have worked on[2]. Accountability also requires that the results of this work are verifiable and not hallucinated. This paper examines the current state of AI accountability, its scope and significance, the key institutions shaping it, and the existing practices being implemented. AI accountability emphasizes risk-management systems that assess risks to human rights, fairness, transparency, and safety, and adopt proportionate measures to mitigate harm. The integration of artificial intelligence (AI) across sectors has intensified concerns regarding accountability. This paper examines the concept of accountability in AI systems, focusing on its conceptual foundations, institutional frameworks, and existing governance practices. It argues that while current mechanisms emphasize transparency, risk management, and oversight, significant accountability gaps persist due to the complexity and opacity of AI systems. By analyzing global regulatory approaches and institutional actors, the paper highlights the limitations of existing frameworks and underscores the need for more structured, enforceable, and human-centric accountability mechanisms.

Key Words: - *Artificial Intelligence, Machine Learning, Accountability*

[1]Organisation for Economic Co-operation and Development (OECD), *OECD Principles on Artificial Intelligence* (2019).
[2]David Leslie et al., *AI Accountability in Practice* (Alan Turing Institute 2024) 10–11.

1. INTRODUCTION

Artificial Intelligence is gaining indulgence[3]in diverse professional fields. This rapidly advancing technology offers users a range of benefits that enable them to work more efficiently and produce high-quality output. This has been made possible through advancements in machine learning [ML], which is the core method for building AI systems that can make independent decisions. However, it has often been observed that AI suffers from a lack of accountability[4]due to the large language models and large sample data feeding, which in turn affects the output data it generates. This can lead to inaccuracies and inefficiencies, which subsequently cause a deterioration in productivity and result in inaccurate outputs.One of the ten OECD AI Principles[5] refers to the accountability that AI actors bear for the proper functioning of the AI systems they develop and deploy. This means that AI actors must take measures to ensure their AI systems are trustworthy, transparent, and safe to use[6]. To achieve this, actors need tomanage risks throughout the lifecycle of their AI systems, from planning to data collection and processing, model building and validation, to deployment, operation, and monitoring[7].

According to the OECD AI Principles, trustworthy[8] AI refers to systems that promote human well-being and sustainable development, respect human rights, fairness, democracy, and the rule of law, and remain transparent, explainable, robust, secure, and safe throughout their lifecycle. AI actors are expected to ensure accountability for the proper functioning of AI systems and enable human oversight and redress where necessary.[9]

2. ACCOUNTABILITY IN AI

2.1 Conceptual meaning

[3]S. Chakravarthy Naik, *Exploring the Role of Technology in Ensuring Accountability of Artificial Intelligence*, 2 IPR J. Maharashtra Nat'l L. Univ. Nagpur 76, 77–79 (2024)
[4]Id. at 80–82; OECD, *Advancing Accountability in Artificial Intelligence*, supra note 1, at 17
[5]*Governing and Managing Risks Throughout the Lifecycle for Trustworthy AI* 16–18 (OECD Digital Econ. Papers No. 349, 2023)
[6]OECD, *OECD AI Principles* (2019); OECD, *Advancing Accountability in Artificial Intelligence*, supra note 1, at 16–20
[7]OECD, *Advancing Accountability in Artificial Intelligence*, supra note 1, at 19–21
[8]OECD, *OECD AI Principles* (2019); OECD, *Advancing Accountability in Artificial Intelligence*, supra note 1, at 16–20
[9]OECD, *Advancing Accountability in AI*, supra note, at 16–18

Accountability emphasizes a relationship of reliability[10], where the entity is obligated to take responsibility for its actions and justify them to a higher authority or the general public. Accountability in AI extends[11] beyond ethical responsibility to encompass legal duties and technical traceability.

2.2 Accountability gaps in AI systems

AI systems create accountability gaps, where decisions are made by automated systems that cannot be legally held accountable. AI systems operate through complex technical processes, making it challenging to determine who should be held responsible when something goes wrong. Responsibility isspread across several human actors, including system designers, developers, and policymakers.AI systems are often described as "black boxes" because their decision-making processes are difficult for humans to fully understand or explain[12]. This lack of transparency has direct implications for AI accountability, as decisions are generated through complex machine-learning models, such as deep neural networks and large language models, whose reasoning is not traceable. Since the basis of an AI-driven outcome cannot be clearly explained, it becomes challenging to identify responsibility, providejustification, and address the concerns of those affected by such decisions[13].

3. FACETS OF AI ACCOUNTABILITY

3.1 Answerability

Answerability requires that actors involved in AI systems are able to justify or defend decisionsproduced by AI. This includesdocumentation that explains the system's logic and addresses stakeholders who are affected by the s ystem[14].

3.2 Auditability

Auditability refers to the capacity of AI systems to be reviewedand assessed. This includes maintaining log books and keeping records of training data. Auditability is central to regulatory oversight and enforcement mechanisms[15].

---

[10]Naik, supra note 3, at 79–80
[11]OECD, *Advancing Accountability in Artificial Intelligence*, supra note 1, at 17; Alan Turing Inst., *AI Accountability in Practice* 10–13 (2024)
[12]Alan Turing Inst., *AI Accountability in Practice*, supra note 8, at 11–12
[13]David Leslie et al., *AI Accountability in Practice*10–12 (Alan Turing Inst. 2024)
[14]Alan Turing Inst., *AI Accountability in Practice*, supra note 8, at 12–14
[15]OECD, *Advancing Accountability in Artificial Intelligence*, supra note 1, at 44–48

3.2 Remedial Accountability

Remedial accountability focuses on corrective mechanisms in response to harm that has occurred[16]. This includes providing redress to affected parties, modifying the AI systems to prevent future issues, withdrawing AI deployment if necessary, or imposing legal liability on the responsible actors. The aim is to ensure that those harmed by AI decisions have appropriate remedies and that the systems are adjusted to mitigate further risks.

4. INSTITUTIONS ENGAGED IN AI ACCOUNTABILITY

4.1 Intergovernmental & Official Bodies

4.1.1 Organisation for Economic Co-operation and Development (OECD)

The Organisation for Economic Co-operation and Development (OECD) is an authority that establishes global standards for AI accountability. Its report, "Advancing Accountability in AI," articulates that accountability is a risk management process that encompasses the entire lifecycle[17] and requires ongoing monitoring and governance.

The OECD outlines four interconnected stages in this process:

1. Define the scope, actors, and context
2. Assess risks
3. Address risks
4. Govern through continuous monitoring and documentation

4.1.2 United Nations Educational, Scientific, and Cultural Organization (UNESCO)

UNESCO[18] establishes a global, ethical, and human rights-based framework for artificial intelligence through its recommendations on the Ethics of AI, which underscore the accountability of AI as a responsibility shared by states, institutions, and developers. The organization emphasizes the importance of remedies for harm resulting from the use of AI.

4.1.3 International Telecommunication Union (ITU)

[16]Alan Turing Inst., *AI Accountability in Practice*, supra note 8, at 16–18
[17]OECD, *Advancing Accountability in Artificial Intelligence*, supra note 1, at 19–23
[18]UNESCO, *Recommendation on the Ethics of Artificial Intelligence* arts. 8–14 (2021)

The ITU leads the development[19] of AI standards and governance in telecommunications and digital infrastructure, ensuring the safe and reliable operation of systems through cross-border coordination, particularly in smart cities. ITU develops standards for interoperable AI in networks, 5G, energy efficiency, and digital health, promoting accountability via benchmarks and multi-stakeholder talks.

Key Focus Areas

- AI for networks, smart services, and public infrastructure.
- Global programs like AI for Good for safe deployment.
- Capacity building through the ITU Academy.

4.1.4 European Union (EU: AI Act and Policy Framework)

The EU leads with the world's first comprehensive AI Act, which enforces accountability through risk-based rules, documentation, assessments, and human oversight, and has become the global regulatory standard.

Key Obligations

- Risk-based rules: Bans unacceptable risks; high-risk needs documentation and audits.
- Human oversight: Required for critical decisions to prevent harm.

Enforcement: National authorities handle compliance with multimillion-euro penalties.

4.2 NATIONAL RESEARCH & POLICY INSTITUTES

These institutions translate principles into practical governance tools, exerting a significant influence on policymakers, regulators, and courts.

4.2.1 Alan Turing Institute

The Alan Turing Institute operationalises AI accountability through applied governance frameworks[20]. Its work emphasises answerability, auditability, and process-based governance across the AI lifecycle. It provides practical tools for public-sector accountability, including workflow mapping and documentation. Its research is frequently cited in government policy and regulatory guidance.

[19]Int'l Telecomm. Union, *AI for Good Global Summit: Governance and Standards Outputs* (2022)
[20]Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), COM (2021) 206 final

4.2.2 Ada Lovelace Institute

The Ada Lovelace Institute addresses the societal and legal impacts of AI, advocating for accountability through transparency, public input, and democratic oversight of power. It bridges ethics, law, and policy to make AI more contestable and responsibly governed. The institute examines how AI systems wield power, advocating institutional duties, transparency in algorithms, and ways for people to challenge decisions. Its research highlights the importance of public participation in ensuring democratic oversight.

Key Principles

- Transparency and contestability: Tools for auditing and questioning AI outcomes.
- Power and ethics: Studies on balancing innovation with societal safeguards.
- Policy bridging: Influences laws through evidence-based reports on accountability.

4.3 Centre for the Governance of AI

The Centre for the Governance of AI (GovAI) examines long-term risks associated with advanced AI, emphasizing institutional accountability, global coordination, and systemic safeguards as systems become increasingly autonomous. Its research shapes how responsibility is assigned in policy debates, without direct enforcement.GovAI analyzes governance challenges, such as power concentration in AI development, advocating for clear institutional roles and international cooperation to mitigate existential risks. Work highlights accountability gaps in autonomous systems.

Key Areas

- Institutional accountability: Frameworks for assigning responsibility among developers, deployers, and regulators.
- Global coordination: Strategies for aligned international standards on powerful AI.
- Systemic risk: Research on long-term impacts and policy tools for oversight

4.4 Algorithmic Justice League

The Algorithmic Justice League (AJL) fights algorithmic bias and discrimination by spotlighting accountability failures in facial recognition and automated decision-making. Through research, advocacy, and public campaigns, it pushes institutions to adopt fairer, more responsible AI—especially in terms of equality.AJL exposes how biased algorithms harm marginalized groups,

using data-driven reports and stories to demand transparency and fixes in high-stakes systems like hiring or policing.

Key Tactics

- Research and audits: Mapping biases in tools like facial tech.
- Advocacy: Campaigns pressuring companies and governments for reforms.
- Public engagement: Films, talks, and tools to build awareness and action.

4.5 Access Now

What it does: Advocates for digital rights and human rights in tech governance, producing policy recommendations and accountability reports on AI transparency, due process, and anti-surveillance protections.

Key goals: Combat AI threats to rights through critiques and campaigns for governments and corporations

4.6 Privacy International

What it does: Critiques opaque AI and surveillance tech undermining privacy and liberties, pushing for oversight and legal safeguards.

Key goals: Strengthen transparency and data governance in state and corporate AI use

4.7 ISO (International Organization for Standardization)

What it does: Develops global standards for AI risk management, governance, and reliability to enable compliance and certification.

Key goals: Translate accountability into structured, auditable processes for organizations.

4.8 IEEE (Institute of Electrical and Electronics Engineers)

What it does: Creates ethical and technical standards for responsible AI design, focusing on transparency and traceability.

Key goals: Embed accountability-by-design into system architecture for engineers.

4.9 NIST (National Institute of Standards and Technology)

What it does: Provides AI risk management frameworks emphasizing governance, documentation, and monitoring tools.

Key goals: Offer practical risk assessment and mitigation for regulatory compliance.

4.10 Microsoft Responsible AI

What it does: Runs internal governance with risk assessments, review boards, and deployment controls.

Key goals: Promote corporate self-regulation through documentation and monitoring.

4.11 Accenture Responsible AI

What it does: Delivers advisory services for enterprise audits, governance models, and risk assessments.

Key goals: Help organizations operationalize accountability in AI practices.

4.12 Alan Turing Institute

What it does: Advances accountability via Process-Based Governance (PBG) workbook for integrating it into workflows.

Key goals: Enable role-mapping, proportionate actions, and transparency in daily AI operations.

V.CONCLUSION

Artificial intelligence does not merely introduce new tools into existing systems; it fundamentally reshapes how decisions are made, how risks are distributed, and how responsibility is assigned. As AI systems become increasingly autonomous, data-driven, and embedded in critical sectors, the traditional foundations of accountabilitybased on clear human agency and traceable decision-makingare placed under considerable strain. This paper has argued that while contemporary frameworks attempt to address these concerns through principles of transparency, risk management, and human oversight, they remain fragmented and often insufficient in practice. The persistence of accountability gapsdriven by technological opacity, diffusion of responsibility, and jurisdictional inconsistenciesreveals that current approaches are more reactive than structurally robust. In many cases, accountability is not absent but diluted, making enforcement uncertain and, at times, ineffective.

A key insight emerging from this analysis is that accountability in AI cannot remain confined to abstract ethical commitments or soft-law instruments. It must be embedded into the very architecture of AI systems through enforceable legal obligations, audit mechanisms, and clearly defined lines of responsibility across the AI lifecycle. This requires a shift from principle-based governance to process-based and outcome-oriented accountability, where actors are not only expected to act responsibly but are also demonstrably held accountable for failures.

Moreover, the centrality of human responsibility must be preserved. AI systems, regardless of their sophistication, do not possess moral agency. Allowing technological complexity to obscure or dilute accountability risks creating a regulatory vacuum in which harm occurs without consequence. Legal systems must therefore resist the temptation to attribute autonomy to machines in a way that displaces liability, and instead reaffirm that responsibility ultimately rests with the human actors who design, deploy, and benefit from these systems. Looking forward, effective AI governance will depend on three interrelated developments: the harmonization of international regulatory standards, the integration of technical and legal expertise in oversight mechanisms, and the institutionalization of transparency and auditability as non-negotiable requirements. Without these, accountability will remain aspirational rather than operational.

In conclusion, the challenge is not merely to regulate artificial intelligence, but to ensure that its deployment remains aligned with fundamental legal principles of fairness, justice, and responsibility. The future of AI governance will be defined not by how advanced these systems become, but by how effectively societies ensure that accountability keeps pace with innovation[21].

[21]John Rawls, *A Theory of Justice* (Revised edn., Harvard University Press 1999).