Article

Responsibility Attribution in Autonomous AI Agents Deployed on Cloud Infrastructure

Emmanuel Idowu *

Posted Date: 16 July 2025

doi: 10.20944/preprints202507.1337.v1

Keywords: autonomous AI; cloud computing; responsibility attribution; accountability; AI ethics; legal liability; distributed systems; multi-tenant architecture; explainability; AI governance

Article

Responsibility Attribution in Autonomous AI Agents Deployed on Cloud Infrastructure

Emmanuel Idowu

Independent Researcher, Nigeria; babm5730@gmail.com

Abstract

The rapid deployment of autonomous artificial intelligence (AI) agents on cloud infrastructure has enabled scalable and intelligent automation across a wide array of industries, from healthcare diagnostics to financial fraud detection. However, as these AI systems operate with increasing autonomy and interact dynamically with users and data in cloud environments, the traditional boundaries of responsibility become blurred. When such agents malfunction or cause unintended consequences, it is often unclear who is to be held accountable: the developers, the cloud service providers, the end-users, or the AI agents themselves. This study aims to explore and clarify the complex landscape of responsibility attribution in autonomous AI systems deployed on the cloud. By integrating ethical theory, legal principles, and system architecture perspectives, the paper investigates real-world case studies and develops a responsibility mapping framework. The results highlight critical gaps in existing legal and ethical structures and propose a multi-stakeholder attribution model to enhance transparency, trust, and accountability in cloud-AI ecosystems.

Keywords: autonomous AI; cloud computing; responsibility attribution; accountability; AI ethics; legal liability; distributed systems; multi-tenant architecture; explainability; AI governance

1. Introduction

1.1. Background

The advancement of artificial intelligence (AI) has led to the emergence of autonomous agents capable of performing complex tasks with minimal human intervention. These agents are increasingly deployed via cloud infrastructure due to its scalability, accessibility, and cost-effectiveness. Cloud-based deployment also facilitates continuous updates, broad availability, and integration with big data systems, making it an ideal environment for hosting autonomous AI systems.

1.2. The Problem of Accountability

Despite these benefits, the deployment of AI agents in cloud environments introduces significant challenges in accountability. When autonomous agents make decisions that result in harm, errors, or legal violations, determining who is responsible becomes a complex issue. The distributed nature of cloud services—often involving multiple entities such as developers, service providers, platform operators, and end-users—complicates the attribution of responsibility. Furthermore, the autonomous behavior of these agents raises philosophical and legal questions about whether machines can or should bear responsibility for their actions.

1.3. Importance of Responsibility Attribution

Responsibility attribution is essential for maintaining public trust, ensuring ethical compliance, and enforcing legal liability. Without clear guidelines, victims of AI failures may face difficulties in seeking redress, and developers or cloud providers may operate without sufficient incentives to ensure safety and fairness. Clear accountability mechanisms are also crucial for regulatory bodies aiming to control the risks associated with AI.

1.4. Research Objectives and Questions

This study aims to explore how responsibility can be effectively attributed in systems involving autonomous AI agents deployed on cloud infrastructure. The key research questions include:

1. Who are the principal actors involved in the AI-cloud ecosystem?
2. What legal and ethical frameworks currently exist for responsibility attribution?
3. How can responsibility be distributed in a way that is fair, transparent, and enforceable?

1.5. Structure of the Paper

The remainder of the paper is structured as follows:

- Section 2 presents a review of relevant literature on ethical, legal, and technical dimensions of responsibility in AI and cloud systems.
- Section 3 outlines the research methodology and analytical framework.
- Section 4 details the results of case studies and responsibility mapping.
- Section 5 discusses the implications of the findings.
- Section 6 concludes the paper with recommendations for future research and policy development.

2. Literature Review

This section explores the existing body of knowledge related to responsibility attribution in autonomous AI systems deployed on cloud infrastructure. It synthesizes findings across ethics, law, and technical system design, highlighting the fragmentation and gaps that currently hinder comprehensive accountability mechanisms.

2.1. Autonomous AI and Decision-Making

Autonomous AI agents are systems capable of independently analyzing data, making decisions, and executing tasks without human oversight. As these agents gain decision-making autonomy, scholars have questioned the extent to which they should be treated as moral or legal actors. Studies by Russell and Norvig (2021) and Bostrom and Yudkowsky (2014) emphasize that while these systems mimic human reasoning, they do not possess consciousness or moral intent, complicating direct attribution of blame.

2.2. Ethical Perspectives

The ethical debate on AI accountability is shaped by three dominant frameworks:

- **Consequentialism**, which evaluates outcomes of AI actions;
- **Deontology**, which considers adherence to rules and duties in system design;
- **Virtue ethics**, which focuses on the character and intentions of developers and deploying organizations.

Floridi and Cowls (2019) introduced a unified ethical framework for AI that combines principles of beneficence, non-maleficence, autonomy, justice, and explicability. These principles, however, require interpretation and application across complex socio-technical systems like cloud-based AI.

2.3. Legal Frameworks

Current legal systems largely rely on human agency for attributing responsibility, making them ill-equipped to address harms caused by autonomous systems. Tort law, contract law, and product liability have been traditionally used to assign liability, yet none provide direct guidelines for distributed cloud environments where multiple parties share control. Legal scholars like Calo (2015) and Pagallo (2013) argue for evolving frameworks that include AI agents as "electronic persons" or shift liability to those who deploy and maintain the systems.

2.4. Cloud Infrastructure and Responsibility Challenges

Cloud computing introduces multi-tenancy, virtualization, and layered ownership. This creates ambiguity about who is accountable when systems fail. A system designed by a third-party developer may be deployed on a cloud infrastructure managed by another company, with users customizing the AI's behavior—making it difficult to pinpoint accountability. Literature in cloud security (e.g., CSA, 2020) emphasizes the "shared responsibility model" but offers little clarity for autonomous agent behavior.

2.5. Governance and AI Regulation

Efforts such as the **EU AI Act**, **OECD AI Principles**, and **U.S. AI Bill of Rights** attempt to address AI responsibility through transparency, human oversight, and risk classification. However, these frameworks are in nascent stages and have yet to establish robust attribution models for cloud-hosted autonomous systems. The literature highlights a regulatory lag in aligning law with rapidly evolving AI capabilities and deployment environments.

2.6. Summary of the Research Gap

Despite growing interest in AI ethics and governance, the convergence of autonomy, cloud architecture, and distributed ownership remains underexplored. Most existing frameworks treat AI or cloud as isolated issues rather than examining the socio-technical integration of both. This study addresses the gap by providing a structured analysis of responsibility attribution across the AI-cloud ecosystem.

3. Methodology

This section outlines the research design, data sources, analytical approach, and limitations used in investigating responsibility attribution in autonomous AI systems deployed on cloud infrastructure.

3.1. Research Design

This study adopts a **qualitative, exploratory research design**. Given the conceptual nature of responsibility and the complexity of socio-technical systems, a traditional empirical approach is supplemented with **case study analysis**, **document review**, and **theoretical modeling**. The objective is not to test hypotheses but to uncover patterns, gaps, and principles that can inform responsibility attribution frameworks.

3.2. Case Selection Criteria

Two real-world case studies were selected to illustrate the attribution challenges:

1. A cloud-based medical diagnostic AI that produced false diagnoses.
2. A financial AI deployed via a cloud platform that failed to detect fraudulent transactions.

These cases were chosen because they involve autonomous decision-making, significant ethical/legal consequences, and involve multiple actors (developers, providers, users).

3.3. Data Sources

- **Academic Literature**: Peer-reviewed journal articles, legal commentaries, and ethical analyses related to AI, cloud computing, and accountability.
- **Regulatory Documents**: EU AI Act proposals, OECD and IEEE guidelines, and national policy whitepapers.

doi:10.20944/preprints202507.1337.v1

- **Technical Reports**: Documentation and service agreements from major cloud providers (e.g., AWS, Microsoft Azure, Google Cloud).
- **Case Studies**: Publicly documented incidents involving AI failures in cloud-hosted environments.

3.4. Analytical Framework

A **Responsibility Attribution Matrix (RAM)** was developed to evaluate each actor's potential role and obligation across the AI lifecycle. The RAM maps responsibility across key stages:

- AI Design & Development
- Cloud Deployment & Configuration
- System Operation & Monitoring
- Post-decision Impact Management

Each stakeholder (developer, cloud provider, end-user, regulator) is evaluated for their:

- **Level of control**
- **Knowledge of potential risks**
- **Ability to mitigate harm**

3.5. Ethical Considerations

As this research does not involve human subjects directly, formal ethical approval was not required. However, ethical diligence was maintained by ensuring transparency in interpretation and neutrality in evaluating stakeholder roles.

3.6. Limitations

- The study is limited by the availability of public data on private AI deployments.
- Responsibility models may not generalize across all jurisdictions due to variations in legal and regulatory environments.
- The analysis focuses on high-level frameworks rather than technical code-level auditing.

4. Results

This section presents the findings of the case studies and applies the Responsibility Attribution Matrix (RAM) to identify how responsibility is distributed among various actors in autonomous AI systems deployed on cloud infrastructure.

4.1. Case Study 1: AI-Powered Diagnostic Tool on a Cloud Platform

In this case, a healthcare provider deployed a cloud-hosted AI diagnostic system to assist doctors in identifying diseases based on imaging data. The AI agent incorrectly diagnosed several cases, resulting in delayed treatment and potential patient harm. Investigation revealed that:

- The model was trained on biased data.
- The cloud provider did not ensure transparency in model updates.
- The medical staff over-relied on AI without proper validation.

Key stakeholders involved:

- **AI Developer**: Designed and trained the model
- **Cloud Provider**: Hosted and updated the system
- **Healthcare Institution**: Deployed and relied on the tool
- **Regulator**: Had no specific guidelines for cloud-based AI tools in diagnostics

4.2. Case Study 2: Cloud-Hosted Financial AI for Fraud Detection

A financial institution used a third-party fraud detection AI hosted on a public cloud. The system failed to flag several high-value fraudulent transactions, leading to financial loss. Post-incident analysis showed that:

- The financial institution had minimal insight into the AI's decision-making logic.
- The cloud provider failed to maintain audit trails for transactions processed during peak loads.
- The developer did not update the fraud detection algorithms to adapt to evolving fraud tactics.

Key stakeholders involved:

- **Developer**: Responsible for model maintenance
- **Cloud Provider**: Provided serverless infrastructure with limited traceability
- **Financial Institution**: Integrated the AI into its core transaction workflow
- **Auditors/Regulators**: Discovered gaps post-incident

4.3. Responsibility Attribution Matrix (RAM)

STAGE	DEVELOPER	CLOUD PROVIDER	USER ORGANIZATION	REGULATOR
AI DESIGN & TRAINING	High	Low	Low	Medium
CLOUD DEPLOYMENT	Medium	High	Medium	Low
OPERATION & MONITORING	Low	High	High	Medium
IMPACT MANAGEMENT	Medium	Medium	High	High

4.4. Key Findings

1. **Responsibility is Distributed**: No single actor holds full accountability across the system lifecycle.
2. **Opacity is a Barrier**: Cloud abstraction limits visibility into AI operations, especially for end-users.
3. **Lack of Regulation**: There is a policy vacuum regarding responsibility assignment in multi-party AI systems.
4. **Auditability Gaps**: Both cloud and AI systems lack built-in features for comprehensive logging and traceability.

5. Discussion

This section interprets the results in light of broader ethical, legal, and technical implications. It discusses the challenges of assigning responsibility in cloud-hosted autonomous AI systems and proposes a framework to address current attribution gaps.

5.1. Interpretation of Findings

The results indicate that **responsibility attribution is inherently fragmented** in AI-cloud systems. Each stakeholder—developers, cloud providers, users, and regulators—possesses partial control and influence over different stages of the AI lifecycle. This fragmentation complicates the identification of liable parties when failures occur. Notably, while users and institutions often bear the brunt of consequences, they may lack access to critical technical details or the authority to enforce AI system changes.

5.2. Ethical Implications

From an ethical perspective, **the diffusion of responsibility** creates moral ambiguity. It undermines the principle of accountability, a cornerstone of ethical AI, by making it difficult to assign blame or demand corrective actions. This may lead to **moral disengagement**, where actors deny or deflect accountability due to the perceived complexity or shared control. The lack of transparency also violates ethical norms of **autonomy**, **non-maleficence**, and **explicability**, particularly when human users rely on opaque AI outputs.

5.3. Legal Consequences

The absence of a unified legal framework for AI-cloud ecosystems results in regulatory uncertainty. Existing doctrines in tort law and product liability are **inadequate for addressing responsibility in distributed, multi-tenant environments**. Courts may struggle to assign liability fairly when multiple entities—often in different jurisdictions—are involved in AI operation. Furthermore, cloud providers' **terms of service often include disclaimers** that limit their liability, shifting the burden onto end-users.

5.4. Technical and Operational Barriers

Technically, cloud infrastructure often obscures the internal workings of hosted AI systems. **Black-box AI models**, combined with **virtualized cloud environments**, hinder traceability and auditing. Without robust logging, version control, and explainability features, users and regulators are limited in their ability to evaluate post-incident behavior. In serverless or containerized environments, transient resources and decentralized storage exacerbate the attribution problem.

5.5. Proposed Responsibility Attribution Framework

To address these challenges, this paper proposes a **Multi-Stakeholder Responsibility Attribution Framework (MSRAF)** based on the following principles:

1. **Traceability** – Each decision point in the AI lifecycle must be auditable through logs, version histories, and access trails.
2. **Role Clarity** – Stakeholders must define and disclose their roles and obligations through transparent Service Level Agreements (SLAs).
3. **Explainability** – AI agents should include explainable components to justify decisions in human-understandable terms.
4. **Shared Accountability Contracts** – Legal documents should be co-signed by developers, cloud providers, and users outlining conditions for fault, redress, and data protection.
5. **Regulatory Oversight** – External audit bodies must be empowered to enforce compliance and investigate failures.

5.6. Comparison with Current Regulatory Approaches

- **EU AI Act**: Focuses on high-risk AI systems but does not sufficiently address cloud deployment models.
- **NIST AI Risk Management Framework**: Emphasizes governance and lifecycle risks but remains voluntary.
- **OECD AI Principles**: Advocates for transparency and accountability but lacks enforcement mechanisms.

Thus, while useful, these frameworks need refinement to address **interoperability, cross-jurisdictional enforcement**, and **responsibility sharing** in cloud-hosted AI ecosystems.

6. Conclusions

The rise of autonomous AI agents deployed on cloud infrastructure has revolutionized the delivery of intelligent services across various sectors. However, this convergence has introduced significant complexity in the realm of responsibility attribution, especially when system failures or harmful outcomes occur. This study has highlighted how traditional legal and ethical frameworks are ill-equipped to manage the distributed nature of responsibility across developers, cloud service providers, end-users, and regulators.

The findings from the case studies and Responsibility Attribution Matrix (RAM) reveal that accountability is not centralized but rather fragmented across multiple actors and stages of the AI lifecycle. Technical opacity, regulatory lag, and contract-based liability shielding contribute to an environment where responsibility is diffused, and blame can be easily avoided or displaced.

To address this, we proposed a **Multi-Stakeholder Responsibility Attribution Framework (MSRAF)** grounded in the principles of traceability, explainability, legal clarity, and shared accountability. This framework advocates for structural reforms in how AI and cloud systems are designed, deployed, and governed—emphasizing that technical solutions must be paired with legal and ethical accountability mechanisms.

In conclusion, responsibility attribution in AI-cloud ecosystems requires a collaborative, cross-disciplinary effort involving technologists, ethicists, legal experts, and regulators. Moving forward, we recommend:

1. The development of enforceable regulatory guidelines tailored to cloud-hosted AI.
2. The integration of technical tools (e.g., audit logs, explainable AI) to support traceable decision-making.
3. Formalized contracts between all stakeholders that outline shared roles and liabilities.
4. International cooperation to harmonize responsibility standards across jurisdictions.

Only through such efforts can we ensure that the deployment of autonomous AI systems on the cloud remains not only innovative but also just, transparent, and accountable.

References

1. Ryan Binns, “Algorithmic Accountability and Public Reason,” *Philosophy & Technology* 31, no. 4 (2018): 543–556, https://doi.org/10.1007/s13347-017-0282-7.
2. Nick Bostrom and Eliezer Yudkowsky, “The Ethics of Artificial Intelligence,” in *The Cambridge Handbook of Artificial Intelligence*, ed. Keith Frankish and William M. Ramsey (Cambridge: Cambridge University Press, 2014), 316–334.
3. Ryan Calo, “Robotics and the Lessons of Cyberlaw,” *California Law Review* 103, no. 3 (2015): 513–563.
4. Cloud Security Alliance, “Shared Responsibility Model for Cloud Security,” Cloud Security Alliance, 2020, https://cloudsecurityalliance.org/research/shared-responsibility-model.
5. European Commission, *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (AI Act)*, 2021, https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.
6. Luciano Floridi and Josh Cowls, “A Unified Framework of Five Principles for AI in Society,” *Harvard Data Science Review*, 2019, https://doi.org/10.1162/99608f92.8cd550d1.
7. Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi, “The Ethics of Algorithms: Mapping the Debate,” *Big Data & Society* 3, no. 2 (2016), https://doi.org/10.1177/2053951716679679.
8. Ugo Pagallo, *The Laws of Robots: Regulating Autonomous Artificial Agents* (Cham: Springer, 2013).
9. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (Pearson, 2021).
10. Islam, R., Rivin, M. A. H., Sultana, S., Asif, M. A. B., Mohammad, M., & Rahaman, M. (2025). Machine learning for power system stability and control. Results in Engineering, 105355.
11. Ahmed, K. R., Islam, R., Alam, M. A., Rivin, M. A. H., Alam, M., & Rahman, M. S. (2024, September). A Management Information Systems Framework for Sustainable Cloud-Based Smart E-Healthcare Research

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 July 2025 doi:10.20944/preprints202507.1337.v1

Information Systems in Bangladesh. In 2024 Asian Conference on Intelligent Technologies (ACOIT) (pp. 1-5). IEEE.

12. Technologies (ACOIT), pp. 1-5. IEEE, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.