# REGULATORY AND POLICY DISCUSSIONS ON LLM AUDITING: CHALLENGES, FRAMEWORKS AND FUTURE DIRECTIONS

Kailash Thiyagarajan

Independent Researcher, Austin, TX - USA

## *ABSTRACT*

*The rapid rise of Large Language Models (LLMs) has revolutionized AI-driven applications but has also raised critical concerns regarding bias, misinformation, security, and accountability. Recognizing these challenges, governments and regulatory bodies are formulating structured policies to ensure the responsible deployment of LLMs. This paper provides a comprehensive analysis of the global regulatory landscape, examining key legislative efforts such as the EU AI Act, the NIST AI Risk Management Framework, and industry-led auditing initiatives. We highlight the gaps in current frameworks and propose a structured policy approach that promotes both innovation and accountability. To achieve this, we introduce a multi-stakeholder governance model that integrates regulatory, technical, and ethical perspectives. The paper concludes by discussing the future trajectory of AI regulation and the critical role of standardized auditing in enhancing transparency and fairness in LLMs.*

## *KEYWORDS*

*LLM Auditing, AI regulation, Ethical AI, Algorithmic Transparency, Bias and Fairness in AI, Explainability*

## 1. INTRODUCTION

Large Language Models (LLMs) such as GPT-4, LLaMA, and Claude have significantly reshaped human-computer interactions, powering applications in content generation, search, and automation. However, as these models become more sophisticated, concerns surrounding their **transparency, bias, and ethical implications** continue to grow. The opaque nature of LLMs makes it challenging to assess their decision-making processes, increasing the need for structured **regulatory oversight**.

In response, global regulatory bodies are actively working to develop **policies and auditing frameworks** to ensure responsible AI deployment. This paper explores the key aspects of LLM auditing and governance, addressing the following critical questions:

- What regulatory mechanisms currently exist for LLM auditing?
- How do legal frameworks promote transparency and accountability in AI models?
- What are the major challenges in implementing effective LLM audits?
- How can policy recommendations enhance governance and mitigate AI-related risks?

David C. Wyld et al. (Eds): NLCAI, AIFU, CCSEA, BIoT, SEA, SIPRO, BDML, CLOUD – 2025
pp. 105-112, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.151008

By examining existing AI regulations and industry-led auditing initiatives, this paper aims to present a **comprehensive policy framework** that promotes fairness, compliance, and user trust in LLM development and deployment.

## 2. LITERATURE REVIEW: GLOBAL REGULATORY LANDSCAPE

As the adoption of Large Language Models (LLMs) accelerates, regulatory efforts worldwide are adapting to address concerns surrounding **transparency, accountability, and ethical governance**. Various regions have introduced policies to oversee AI deployment, yet these frameworks differ significantly in their **scope, enforceability, and regulatory mechanisms**.

### 2.1. EU AI Act: A Risk-Based Regulatory Approach

The **EU AI Act** is the first comprehensive AI regulation, categorizing AI systems into different risk levels, ranging from minimal to high risk. For LLMs deployed in high-risk domains such as finance, healthcare, and legal applications, the Act mandates:

- Rigorous auditing and documentation of AI models before deployment.
- Implementation of bias testing, algorithmic transparency, and human oversightto ensure accountability.
- A requirement for AI providers to disclose training data sources and explainability measures**.**

Despite its structured approach, enforcement challenges remain, particularly regarding cross-border compliance and proprietary AI models that companies may be reluctant to disclose.

### 2.2. NIST AI Risk Management Framework (USA)

Unlike the EU's mandatory regulations, the **NIST AI Risk Management Framework** serves as a **voluntary guideline** for assessing AI risks in the United States. Key components include:

- Encouraging bias auditing and adversarial robustness testing.
- Promoting explainability metrics to assess AI decision-making transparency.
- Establishing best practices for AI risk mitigation in organizations.

While widely adopted in industry, its non-mandatory nature limits its regulatory impact, making compliance dependent on individual organizations**.**

### 2.3. China's AI Regulations: Strict Compliance Measures

China has adopted stringent AI governance policies, particularly for generative AI models like LLMs. Regulatory guidelines mandate:

- Security assessments and real-time monitoring of AI-generated content.
- Preventive measures to ensure AI does not generate content harmful to public security**.**
- Strong oversight mechanisms for LLM providers operating within China**.**

China's approach reflects a more centralized and restrictive model**,** with an emphasis on state oversight and compliance enforcement**.**

### 2.4. Industry Standards: ISO, IEEE, and Organizational Guidelines

In addition to governmental policies, industry standards play a crucial role in shaping AI auditing practices:

- **ISO/IEC 42001:** Establishes AI management system standards for organizational AI governance.
- **IEEE 7001-2021:** Defines AI transparency and governance principles**.**
- OpenAI's voluntary model cards and transparency reports: Provide insights into LLM risks, capabilities, and ethical considerations**.**

While these standards offer guidance, lack of universal enforcement and adoption disparities hinder their effectiveness. A globally recognized AI certification system could bridge these gaps by ensuring consistent auditing practices across jurisdictions**.**

## 3. CHALLENGES IN LLM AUDITING AND POLICY IMPLEMENTATION

Despite the increasing focus on **LLM auditing**, several key challenges continue to hinder effective regulatory oversight and policy enforcement. These challenges span technical, legal, economic, and ethical dimensions, making it difficult to establish universally accepted auditing standards.

### 3.1. Absence of Standardized Auditing Metrics

One of the primary obstacles in LLM auditing is the lack of a universally accepted evaluation framework for key risks such as bias, fairness, hallucinations, and robustness. Current auditing practices often:

- Rely on ad-hoc, organization-specific methodologies developed by AI vendors.
- Lack consistent benchmarks, making cross-model comparisons difficult.
- Struggle with measuring model drift and real-world performance degradation.

Developing global auditing standards that define quantifiable thresholds for fairness, accuracy, and transparency remains a critical challenge.

### 3.2. Model Opacity and Proprietary Restrictions

Most leading LLMs, such as **GPT-4, Claude, and Gemini**, operate as black-box systems**,** limiting external auditing efforts. AI developers often withhold internal model details due to:

- **Trade secret protections**, citing competitive risks.
- **Proprietary architectures**, making third-party assessments difficult.
- **Limited transparency in training data sources**, restricting bias evaluation.

Without clear mandates for explainability and documentation disclosure, regulators struggle to conduct independent and reliable audits.

### 3.3. Cross-Jurisdictional Compliance Conflicts

Global AI providers face legal inconsistencies when deploying LLMs across different regulatory environments**:**

- The **EU AI Act** mandates risk assessments for high-risk AI applications.
- In contrast, the United States follows a voluntary compliance model**.**
- China enforces strict content moderation policies on generative AI models.

This lack of regulatory alignment creates compliance mismatches and legal uncertainties**,** making it difficult for AI developers to adhere to a single, globally recognized auditing framework.

### 3.4. High Computational Costs of Auditing

Effective LLM auditing demands large-scale data evaluations**,** extensive bias testing, and adversarial robustness checks. These resource-intensive processes pose several challenges:

- **High infrastructure costs** associated with GPU/TPU computations.
- **Limited accessibility for independent auditors** due to financial constraints.
- **Scalability concerns** in conducting audits across constantly evolving models.

A potential solution is the development of publicly funded AI auditing infrastructures that allow third-party researchers to conduct evaluations without prohibitive costs**.**

### 3.5. Ethical and Societal Risks in Auditing

AI auditing raises **ethical dilemmas**, particularly regarding:

- **Censorship vs. Free Speech**: Regulatory intervention in LLM content moderation raises questions about who defines harmful content **and** where to draw the linebetween ethical oversight and restriction of expression.
- **Bias in Regulatory Auditing**: Government bodies and regulatory agencies may inadvertently introduce biases into auditing procedures, influencing AI governance outcomes.
- **Potential for Regulatory Capture**: Over-reliance on industry-led auditing frameworks may result in biased self-regulation**,** undermining objective compliance assessments.

Addressing these ethical concerns requires multi-stakeholder collaboration**,** ensuring that LLM auditing remains transparent, impartial, and socially responsible.

## 4. Proposed Policy Framework for LLM Auditing

As LLMs continue to play a crucial role in various domains, a structured and enforceable policy framework is necessary to balance technological innovation, transparency, and accountability. While existing regulations such as the EU AI Act, NIST AI Risk Management Framework, and IEEE AI Standards provide a foundational structure, gaps remain in auditing consistency, regulatory enforcement, and global compliance mechanisms**.**

To address these challenges, we propose a multi-layered regulatory framework that ensures standardized auditing procedures, independent oversight mechanisms, and ethical compliance measures. This framework is structured into two key components:

- Transparency and Standardization in LLM Auditing
- Regulatory Oversight, Bias Mitigation, and Ethical Safeguards

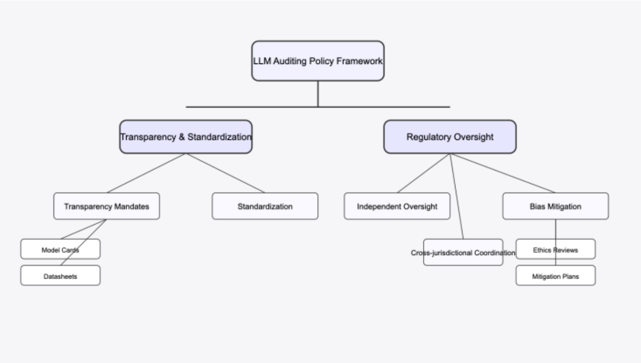

Figure 1

## 4.1. Transparency & Standardization in LLM Auditing

Transparency is fundamental to ensuring trustworthy and fair AI systems. To achieve this, policymakers should enforce mandatory disclosure mandates requiring AI developers to provide detailed documentation on:

- **Training data sources and model fine-tuning methods** to ensure traceability.
- **Bias mitigation strategies** implemented during model development.
- **Known risks and limitations**, including model drift and adversarial vulnerabilities.

These disclosures should be standardized in the form of model cards, datasheets, and risk assessments**,** similar to industry-led initiatives by OpenAI, Meta, and Anthropic. However, voluntary disclosure lacks enforceability; thus, global regulatory bodies should mandate transparency reports as part of AI compliance requirements**.**

**Standardized Auditing Metrics:**

A critical challenge in AI regulation is the lack of uniform auditing benchmarks. At present, different organizations use inconsistent methodologies to assess bias, fairness, and robustness. To address this, we propose the adoption of globally recognized AI certification standards, developed by ISO, IEEE, and other standardization bodies, which define compliance thresholds for:

- **Fairness and bias testing** to minimize algorithmic discrimination.
- **Explainability requirements** ensuring AI decision-making transparency.
- **Security and adversarial robustness** assessments for risk mitigation.

Additionally, to **reduce auditing costs**, policymakers should develop publicly accessible AI audit infrastructures, allowing independent researchers and regulatory bodies to evaluate models without excessive computational expenses**.**

## 4.2. Regulatory Oversight, Bias Mitigation, and Ethical Safeguards

While transparency and standardized audits enhance accountability, they must be supported by robust regulatory enforcement and ethical safeguards. One of the key weaknesses of existing AI governance models is the reliance on self-regulation, where AI companies assess their own compliance.

**Independent AI Auditing Bodies:**

To mitigate the risks of biased self-assessments, governments should establish independent AI oversight agencies responsible for:

- Evaluating high-risk AI models before deployment in sensitive applications (e.g., healthcare, finance, law enforcement).
- Imposing regulatory penalties on non-compliant AI providers.
- Ensuring adherence to AI ethics guidelines through interdisciplinary review panels.

These agencies should function similarly to financial regulatory bodies (e.g., the **Securities and Exchange Commission (SEC)**) by enforcing AI auditing through legal mandates rather than voluntary compliance.

**Bias Mitigation and Ethical AI Development:**

AI fairness must be integrated proactively into the development pipeline rather than as a post-deployment fix. To achieve this, we recommend:

- **Pre-deployment ethics reviews**, where interdisciplinary panels (including AI ethicists, legal experts, and social scientists) assess models for bias and societal risks.
- **Bias-resistant training data initiatives**, ensuring AI models are trained on diverse and inclusive datasets to prevent discriminatory outcomes.
- **Cross-jurisdictional regulatory cooperation**, fostering international agreements to harmonize AI governance policies**.**

Since AI technologies **evolve rapidly**, AI regulation must be dynamic and adaptable**.** Periodic review cycles, impact assessments, and stakeholder engagement should be integrated into the regulatory process to ensure that policies remain effective in mitigating emerging risks**.**

In summary, our proposed policy framework combines transparency mandates, standardized auditing mechanisms, and independent oversight institutions to ensure that LLMs are developed and deployed in a fair, accountable, and ethical manner**.** By establishing globally aligned regulatory frameworks**,** policymakers can enhance trust in AI technologies while minimizing risks associated with bias, opacity, and non-compliance**.**

## 5. Future Directions in AI Regulation and Auditing

As AI governance continues to develop, regulatory frameworks must remain adaptable to emerging risks and technological advancements. Several key areas require further research and policy refinement:

- **Explainability and Interpretability Standards:** Future regulations should establish clear guidelines for model transparency, ensuring that LLMs can provide interpretable justifications for their decisions, particularly in high-risk domains such as finance, healthcare, and law.

- **Decentralized and Secure Auditing Mechanisms:** Traditional AI auditing is computationally expensive and often controlled by large organizations. Federated auditing models using secure cryptographic verification techniques could allow independent researchers and regulators to evaluate AI systems without centralized control**.**

- **Public AI Benchmarking Datasets:** Regulatory bodies should invest in the creation of open-source datasets designed for bias testing, fairness evaluation, and robustness assessments. These datasets should be regularly updated to reflect real-world challenges in AI deployment.

- **Adaptive AI Regulations:** AI threats evolve rapidly, requiring dynamic regulatory updates**.** Governments should implement periodic policy reviews, risk assessments, and multi-stakeholder discussions to ensure that legal frameworks stay relevant in addressing new risks, such as adversarial attacks, prompt injections, and AI-generated misinformation**.**

By focusing on these areas, AI governance can move towards a more transparent, accountable, and ethically responsible future**.**

## 6. Conclusion

The increasing adoption of large language models presents both opportunities and challenges, making effective auditing and regulation essential. While existing frameworks such as the EU AI Act and NIST guidelines provide initial steps, they lack globally unified enforcement mechanisms. A comprehensive approach to LLM auditing should integrate transparency mandates, standardized evaluation criteria, and independent oversight to ensure fairness and accountability.

As AI continues to evolve, regulatory frameworks must be adaptable to emerging risks while maintaining a balance between innovation and responsible deployment. Establishing clear auditing standards and governance structures will be key to fostering public trust in AI systems and ensuring their ethical and reliable use across industries.

### Acknowledgments

We would like to thank the open-source community and researchers from large companies whose work in pioneering transformers and LLMs has made this advanced recommender system possible. We also want to thank e-commerce data teams who contributed real-world feedback on scaling, data cleaning, and model deployment strategies.

### References

[1] Mökander, J., Schuett, J., Kirk, H. R., &Floridi, L. (2023). *Auditing Large Language Models: A Three-Layered Approach.*arXiv preprint arXiv:2302.08500

[2] Gaebler, J. D., Goel, S., Huq, A., & Tambe, P. (2024). *Auditing the Use of Language Models to Guide Hiring Decisions.*arXiv preprint arXiv:2404.03086.

[3] McIntosh, T. R., Susnjak, T., Liu, T., Watters, P., Nowrozy, R., &Halgamuge, M. N. (2024). *From COBIT to ISO 42001: Evaluating Cybersecurity Frameworks for Opportunities, Risks, and Regulatory Compliance in Commercializing Large Language Models.*arXiv preprint arXiv:2402.15770.

[4] Hassani, S. (2024). *Enhancing Legal Compliance and Regulation Analysis with Large Language Models.*arXiv preprint arXiv:2404.17522.

[5] Kim, A. G., Muhn, M., Nikolaev, V. V., & Tan, I. (2024). *Large Language Models and Financial Reporting Oversight.* Public Company Accounting Oversight Board (PCAOB).

[6] World Health Organization (WHO). (2024). *WHO Releases AI Ethics and Governance Guidance for Large Multi-Modal Models.*

[7] Holistic AI. (2024). *LLM Auditing Guide: What It Is, Why It's Necessary, and How to Execute It.* Retrieved from https://www.holisticai.com/papers/llm-auditing-guide

[8] Resaro. (2024). *Resaro's Bias Audit: Evaluating Fairness of LLM-Generated Testimonials.* UK Government Digital Service.

[9] Mökander, J., Schuett, J., Kirk, H. R., &Floridi, L. (2023). *Auditing Large Language Models: A Three-Layered Approach.* AI and Ethics, 3, 1-14. DOI: 10.1007/s43681-023-00234-2

[10] Gaebler, J. D., Goel, S., Huq, A., & Tambe, P. (2024). *Auditing the Use of Language Models to Guide Hiring Decisions.* Proceedingsof the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT). DOI: 10.1145/3490215.3503723

## AUTHORS

**Kailash Thiyagarajan** is a Senior Machine Learning Engineer with over 18 years of experience in IT, specializing in scalable ML solutions, recommendation systems, and real-time inference. He has a strong background in MLOps, patent applications, and mentoring early-career engineers. Passionate about innovation, he actively participates in hackathons and applies AI to real-world problems‘