Balancing Accountability and Accessibility in Artificial Intelligence: A Proposal for Structural Transparency

Author Information:
Name: Piyush Kolekar
School: Dublin Jerome
Country: USA
Contact email: kolekarpiyush23@gmail.com
Year: 2026

Abstract: Artificial intelligence (AI) systems now play a crucial role in high-stakes decisions across healthcare, finance, education, and employment. However, most modern AI systems function as highly opaque "black boxes," lacking clear mechanisms for accountability. This paper explores structural interventions to make AI more transparent, trustworthy, and accountable without compromising its effectiveness or global accessibility. By analyzing current regulatory frameworks and the measurable harms of AI bias, this paper proposes a global AI Transparency Certification system to ensure ethical deployment while fostering innovation.

1. Introduction and Research Question

Research Question: How can artificial intelligence be made more transparent, trustworthy, and accountable without compromising its effectiveness or accessibility?

Currently, artificial intelligence plays a significant role in critical societal processes, from prioritizing job applications and considering loan requests to providing patient care. Economically, AI is expected to contribute up to $15.7 trillion to global GDP growth over the

next two decades (PwC, 2023). Almost every person seeking employment, financing, or healthcare faces AI-driven decisions.

A primary issue with modern AI systems—especially those based on deep learning—is that they operate as "black boxes." While the inputs and outputs are visible, the internal decision-making process remains unknown, often even to the engineers who designed the algorithms (IBM, 2024). Consequently, training data bias goes undetected, deepfakes become harder to address, and accountability is nearly impossible to establish after adverse outcomes.

For instance, facial recognition systems have historically misidentified darker-skinned women at significantly higher rates than lighter-skinned men (UNESCO, 2021), and hiring algorithms have filtered out applicants based on racial and gender proxies (World Economic Forum, 2024). Millions of people are subjected to AI-based decisions without understanding how those decisions are made or how to challenge them. To address this, international bodies like the European Union, UNESCO, and the World Economic Forum are actively promoting ethical and regulatory solutions.

2. Hypothesis

If governments and technology companies adopt explainable AI systems and conduct regular audits, then users will be more likely to develop trust in AI without compromising its effectiveness and accessibility.

This hypothesis relies on the logic that user trust is cultivated when the decision-making process is comprehensible and the responsible parties are clearly identified. Conversely, opaque AI directly fosters mistrust and a reluctance to adopt beneficial technologies. By ensuring users can comprehend the logic behind AI outputs and confirm accountability, we can bridge the gap between innovation and public trust.

3. Key Findings

3.1 Finding 1: Lack of Transparency Directly Reduces Trust

Many users fail to grasp how AI systems generate decisions, undermining overall confidence in the technology. According to a 2022 Pew Research Center study, 45% of Americans experience conflicting feelings about AI, citing a lack of clarity regarding how it works as a major concern. Furthermore, IBM's Global AI Adoption Index (2023) reports that over 50% of enterprise IT leaders view the inability to interpret an AI's actions as the primary barrier to scaling its use within their organizations. Consequently, AI explainability has become an operational necessity rather than an optional feature.

3.2 Finding 2: AI Bias Causes Documented, Measurable Harm

Studies consistently demonstrate that facial recognition software and algorithmic screening tools can discriminate against minorities and women due to biased training datasets. UNESCO noted an investigation revealing a 34.7% error rate in facial recognition for darker-skinned women, compared to just 0.8% for lighter-skinned men. Additionally, widely used credit algorithms have been shown to offer significantly lower credit limits to women than to men with identical

financial attributes. These instances of algorithmic bias stem from unaudited training data, underscoring the urgent need for mandatory transparency in AI models.

3.3 Finding 3: Regulation Improves Accountability Without Stifling Innovation

Passed in June 2024, the EU's AI Act (Regulation (EU) 2024/1689) established the world's first comprehensive legal AI governance regime (European Commission, 2024). Utilizing a risk-based approach, the legislation mandates transparency, human oversight, and independent auditing for high-risk applications, while applying minimal requirements to low-risk technologies like spam filters. This demonstrates that targeted regulation can successfully allocate responsibility where it is most needed while preserving a flexible environment for innovation.

3.4 Finding 4: Open and Accessible AI Expands Global Opportunity

The availability of free, multilingual AI services empowers students, researchers, and small businesses worldwide, particularly in developing nations, by providing access to state-of-the-art technology (OpenAI, 2025). Recognizing that accessibility is a core tenet of ethical AI, UNESCO has called for the preservation of this access in future regulations (UNESCO, 2021). Therefore, any newly implemented AI certification or regulation must include compliance tiers to ensure global access is not inadvertently restricted.

4. Proposed Solution: A Global AI Transparency Certification

To address the findings above, this paper proposes the development of a **Global AI Transparency Certification system**. Under this framework, any AI system intended for

high-stakes, crucial functions would require international certification prior to deployment. This system should be co-developed by technology companies and international regulatory bodies.

Certified systems must comply with four key criteria:

1. **Explainability:** Explain how the model functions in plain language and identify which variables heavily influence decision-making.

2. **Full Disclosure:** Publicly disclose system limitations, known error rates, and potentially impacted demographics.

3. **Independent Auditing:** Pass annual, independent bias audits conducted by certified third-party organizations.

4. **Human Oversight:** Ensure a mandatory human review process for any AI decision that carries severe consequences for an individual's life or livelihood.

International standards organizations (e.g., ISO, the United Nations) could oversee this certification process, mirroring the rigorous safety certifications currently utilized in the aviation and pharmaceutical industries.

This framework aligns directly with the research findings. Disclosing operational mechanics increases trust (Finding 1), while independent audits directly combat systemic bias (Finding 2).

A risk-proportionate, tiered approach ensures that innovation remains unhindered (Finding 3) and global accessibility is preserved (Finding 4). Furthermore, because the certification focuses on outputs, audits, and operational transparency rather than forcing developers to reveal proprietary source code or model weights, it safely balances commercial interests with public safety.

5. Conclusion

Artificial intelligence is rapidly becoming an unavoidable part of modern reality. Current research unequivocally shows that poor AI explainability, algorithmic bias, and a lack of proper regulation erode public trust. However, adopting explainable AI standards, enforcing ethical considerations, and requiring independent auditing can promote reliability without compromising technical efficiency. Implementing a standardized **AI Transparency Certification system** will empower companies to develop more trustworthy products while safeguarding public access to transformative technology.

References

- **European Commission.** (2024). EU Artificial Intelligence Act (Regulation (EU) 2024/1689). Official Journal of the European Union. https://artificialintelligenceact.eu

- **IBM.** (2023). Global AI Adoption Index 2023. IBM Institute for Business Value. https://www.ibm.com/think/topics/black-box-ai

- **OpenAI.** (2025). AI safety and alignment research. OpenAI. https://openai.com/safety

- **Pew Research Center.** (2022). AI and human enhancement: Americans' openness is tempered by a range of concerns. https://www.pewresearch.org

- **PwC.** (2023). Sizing the prize: What's the real value of AI for your business and how can you capitalise? Pricewaterhouse Coopers. https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

- **UNESCO.** (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

- **World Economic Forum.** (2024). Global Risks Report 2024. World Economic Forum. https://www.weforum.org/reports/global-risks-report-2024