AI, Ethics and Society

From Accountability to Trust

Whitepaper by Asutosh Hota
Director Consulting Expert, CGI

CGI

From Accountability to Trust

Artificial intelligence is rapidly becoming embedded in decision-making across society. Algorithms influence decisions in healthcare, finance, employment, and many other domains. While these systems promise efficiency and insight, their deployment raises fundamental questions about rexsponsibility, governance, and trust.

This paper examines five interconnected challenges that define the societal impact of AI. First, when AI systems make decisions, organizations must determine who is accountable for the outcomes. Second, ethical principles alone are insufficient unless they are translated into operational practices embedded in development processes. Third, weak governance structures create operational, reputational, and financial risks that often emerge only after AI systems are deployed.

The challenge becomes even more complex in the global environment, where competing regulatory models create a fragmented governance landscape. Finally, the long-term success of AI initiatives depends not only on technological capability but on trust. AI systems can only scale when users, regulators, and stakeholders believe that the systems are fair, transparent, and accountable.

> Responsible AI is not merely a technical matter. It is a question of organizational design, governance structures, and societal legitimacy.

Contents

Who is accountable when AI decides?

People are willing to trust AI, but only up to a point. When algorithms deny services or produce flawed outcomes, trust quickly disappears unless someone is accountable.

Assigning accountability, however, is far from simple. Public attitudes toward algorithmic decision-making remain cautious. Research shows that people are significantly less comfortable with AI replacing human decision-makers in domains like hiring or legal judgments than with AI assisting in personal health decisions. Many believe certain decisions should remain fundamentally human unless strong safeguards exist.

At the same time, experts themselves disagree on what responsible AI governance should look like. Some emphasize existential risks and argue for strict control. Others advocate rapid innovation with minimal regulation. Still others focus on present-day harms and structural inequalities embedded in AI systems.

Within organizations, accountability often becomes fragmented. AI systems are typically built through modular development processes resembling supply chains: data teams, model developers, system integrators, and product teams each contribute components. While individual teams may carefully address issues within their domain, responsibility for the overall system often remains unclear.

As a result, a system can pass internal checks at each stage and still cause unexpected harm in practice. No single team may have evaluated the full socio-technical system—the interaction between technology, organizational processes, and human decision-making.

For leaders, the implication is clear: responsibility cannot be delegated to the algorithm. If an AI system denies someone's loan or produces a flawed diagnosis, regulators and the public will hold the organization accountable.

Accountability must therefore be designed into the organization itself—through clear ownership, transparent decision pathways, and oversight structures embedded in the development and deployment of AI systems.

What this means for leaders?

AI governance is no longer a downstream compliance task. Organizations must treat responsibility, oversight, and trust as design principles from the start.

Leaders should ensure that:

- AI systems have clear ownership and accountability.
- Ethical principles translate into operational governance structures.
- AI risks are monitored continuously throughout the system lifecycle.
- Trust is treated as a strategic objective, not a communication exercise.

From AI principles to AI reality

In response to these concerns, many organizations have adopted AI ethics principles. Companies and governments alike commit to fairness, transparency, and accountability.

Yet AI failures continue to dominate headlines. The gap between principles and practice remains one of the most persistent challenges in responsible AI.

Organizations publish principles, but these rarely influence day-to-day development decisions. Studies show that simply exposing developers to ethical codes has little measurable impact on their behavior.

Part of the difficulty lies in conflicting values. AI systems often serve multiple stakeholders whose expectations differ significantly. What one group considers fair or acceptable may appear problematic to another.

Technical safeguards alone are also insufficient. Developers frequently implement moderation rules or fairness constraints to guide AI behavior. However, these guardrails can be fragile and sometimes introduce new unintended biases.

Ultimately, turning ethical principles into operational practice requires more than policy statements. It demands governance structures, incentives, and accountability mechanisms that embed responsible AI into everyday decision-making.

Why accountability breaks down in AI systems?

AI development often resembles a supply chain. Data teams, model developers, system integrators, and product teams each build separate components.

While every team may address risks within its own domain, responsibility for the overall system can remain unclear. As a result, an AI system can pass internal checks yet still cause harm once deployed in real-world environments.

The hidden cost of ungoverned AI

When governance structures are weak or absent, the consequences extend far beyond technical errors. Unchecked AI creates operational disruption, reputational damage, and financial risk.

Studies of real-world generative AI incidents reveal that many failures arise not from algorithmic flaws but from how systems are deployed, used, and monitored.

Misuse, insufficient oversight, and poor incident reporting often lie at the heart of AI breakdowns.

Small issues frequently escalate into larger crises because governance mechanisms fail to detect and contain them early.

Reputational risks illustrate this dynamic clearly. Even when AI performs correctly, the way it is used can undermine trust.

For example, communications labeled as AI-assisted may appear less sincere, altering how recipients perceive the sender.

Operational risks also grow when incident reporting and oversight systems are missing. Without transparent monitoring and structured response mechanisms, organizations may fail to detect system failures until they cause significant disruption.

"Improving technical robustness alone is rarely sufficient. Effective AI governance requires layered oversight, clear guardrails, and human involvement in critical decision points."

Fragmented global AI governance

The governance challenge becomes even more complex at the global level.

AI does not fail because it lacks regulation. It fails because it is governed differently across jurisdictions.

Today's global AI landscape reflects competing models of oversight:

1. **Market-driven systems** emphasize innovation and voluntary standards.
2. **Rights-based frameworks** prioritize legal safeguards and pre-deployment regulation.
3. **State-led models** integrate AI governance with national policy objectives.

These approaches reflect different assumptions about power, responsibility, and the role of institutions.

For organizations operating internationally, this fragmentation creates operational complexity. A system approved in one jurisdiction may face additional scrutiny or prohibition in another.

AI governance has therefore become a design constraint rather than a downstream compliance task. Organizations must anticipate regulatory diversity and embed governance into system architecture from the start.

Trust is the real AI advantage

Ultimately, the success of AI initiatives depends on trust. Organizations may deploy powerful technologies, but if users, regulators, and employees do not trust the system, adoption will stall.

Trust emerges when AI systems demonstrate fairness, transparency, and alignment with human values. Failures in these areas quickly undermine confidence, even if the technology itself functions as intended.

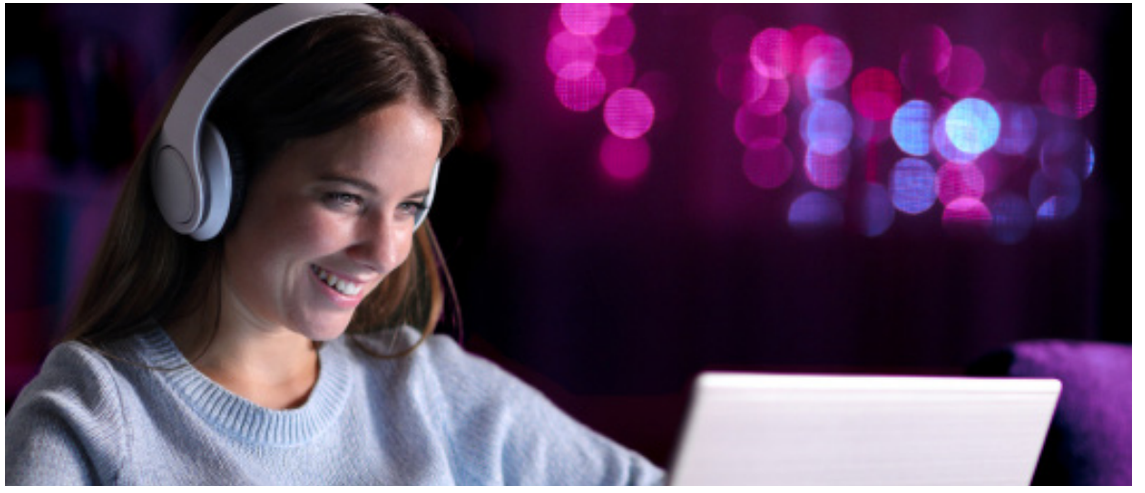
Many AI controversies are therefore not technical failures but trust failures. The technology may work, but people reject it because it conflicts with their expectations or values.

For leaders, trust must become a central design objective. Responsible AI is not simply about avoiding harm—it is about building systems that people believe are legitimate, fair, and accountable.

Organizations that recognize this early will be better positioned to scale AI responsibly across industries and societies.

Let’s continue the conversation

Asutosh Hota
Director Consulting Expert
asutosh.hota@cgi.com



Further Reading

AI Governance and Regulation

Ericson, P., Carli, R., Tucker, J. & Dignum, V. (2025). AI Policy for Whom? Reclaiming Governance from Capitalist Capture. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, pp. 838–849.

Nannini, L. (2025). When Less Regulation Means More Complexity: The EU AI Liability Directive Withdrawal and Its Impact on European Technological Competitiveness. AIES 2025, Vol. 8(2), pp. 1848–1861.

Schmitt, L. (2022). Mapping Global AI Governance: A Nascent Regime in a Fragmented Landscape. AI and Ethics, 2, 303–314.

Ruster, L.P. (2025). Responsible AI Practices: Histories, Definitions, Barriers and Future Directions. AIES 2025.

AI Ethics, Alignment and Moral Decision-Making

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines, 30(1), 99–120.

Rismani, S., Shelby, R., Davis, L., Rostamzadeh, N. & Moon, A. (2025). Measuring What Matters: Connecting AI Ethics Evaluations to System Attributes, Hazards, and Harms. AIES 2025.

AI Risk and Organizational Resilience

Li et al. (2025). A Closer Look at the Existing Risks of Generative AI: Mapping the Who, What, and How of Real-World Incidents. AIES 2025.

Antar, A.D., Huan, X. & Banovic, N. (2025). Do Your Guardrails Even Guard? Method for Evaluating Effectiveness of Moderation Guardrails in Aligning LLM Outputs with Expert User Expectations. AIES 2025.

Raff et al. (2025). You Don’t Need Robust Machine Learning to Manage Adversarial Attack Risks. AIES 2025, Vol. 8(3), pp. 2094–2106.

Bias, Trust and Human Adoption

Mun, J., Au Yeong, W.B., Deng, W.H., Schaich Borg, J. & Sap, M. (2025). Why (not) use AI? Analyzing People’s Reasoning and Conditions for AI Acceptability. arXiv:2502.07287.

Wilson, K. et al. (2025). No Thoughts Just AI: Biased LLM Hiring Recommendations Alter Human Decision Making and Limit Human Autonomy. AIES 2025.

Khadpe et al. (2025). Explaining the Reputational Risks of AI-Mediated Communication: Messages Labeled as AI-Assisted Are Viewed as Less Diagnostic of the Sender’s Moral Character. AIES 2025.

About CGI

Insights you can act on

Founded in 1976, CGI is among the largest technology and professional services companies in the world.

We are insights-driven and outcomes-focused to help accelerate returns on your investments. Across hundreds of locations worldwide, we provide comprehensive and scalable technology and professional services that are informed globally and delivered locally.

For more information, visit www.cgi.fi/tekoaly