A First Read

India's AI Governance Guidelines and What Comes Next

ARTIFICIAL INTELLIGENCE & THE FUTURE OF WORK · December 2025

Executive Summary

India has made its choice: guidelines now, statute later — if at all. Whether that sequencing holds up will depend on how quickly "later" actually arrives.

This November, MeitY released the India AI Governance Guidelines under the IndiaAI Mission — the clearest articulation yet of how the government intends to govern artificial intelligence in India. It is not a law. It is a voluntary, principles-based framework built around seven core principles (Trust, Human-Centricity, Fairness and Equity, Accountability, Understandable by Design, Safety and Robustness, and Inclusive and Sustainable Innovation), backed by three new institutional bodies: an AI Governance and Economic Group, a Technology and Policy Expert Committee, and an IndiaAI Safety Institute. This brief offers our first assessment of what the Guidelines actually establish, what they leave open, and what would need to happen next for a voluntary framework to function as genuine governance rather than an aspirational document.

1. What the Guidelines Establish

- **Seven principles, not seven rules.** The Guidelines' core content is a set of principles — Trust, Human-Centricity, Fairness and Equity, Accountability, Understandable by Design, Safety and Robustness, and Inclusive and Sustainable Innovation — that are explicitly framed as a foundational reference rather than enforceable obligations. No penalty attaches to non-compliance, because compliance itself is not currently mandatory.
- **A three-body institutional architecture.** The AI Governance and Economic Group, described as a high-level inter-ministerial body, is tasked with steering national AI governance strategy; a Technology and Policy Expert Committee supports it with technical and policy recommendations; and the IndiaAI Safety Institute is charged with safety research, standards development, system testing and risk evaluation. All three are new, and none has yet published operational detail on staffing, timelines or specific deliverables at the time of this brief.
- **An explicit choice against a binding AI statute, for now.** The Guidelines apply existing law — the IT Act, the Digital Personal Data Protection Act — to AI-specific harms, rather than creating new AI-specific legal obligations. This is a deliberate design choice, not an oversight: it prioritises speed and flexibility over the comprehensive, binding approach taken by the European Union's AI Act.

2. What Remains Genuinely Unclear

A guidelines-first approach is only as good as the credibility of the institutions meant to give it teeth, and on that score there is considerably more to be resolved than the Guidelines' release announcement suggests.

- **Testing and certification specifics:** The IndiaAI Safety Institute's mandate to "test AI systems" and "develop standards" is, so far, described only in general terms. What gets tested, on what timeline, against what published standard, and with what consequence for failure, are all open questions.
- **The path from voluntary to binding, if any:** The Guidelines do not specify conditions under which voluntary principles might convert into binding rule — whether that would require evidence of specific harms, a policy review after a fixed period, or simply ongoing political judgment.
- **Interaction with the DPDP Act's consent requirements:** The Digital Personal Data Protection Act already restricts using personal data to train AI models without explicit consent. How the new Guidelines' "Accountability" and "Understandable by Design" principles interact with, or add to, that existing statutory requirement is not yet clarified in the Guidelines themselves.

3. The Case For and Against This Approach

The case for a guidelines-first model is genuinely strong: it lets government, industry and civil society calibrate the framework against real-world AI harms as they emerge, rather than freezing assumptions about risk into binding statute before those risks are fully understood — a critique frequently levelled, fairly or not, at more prescriptive frameworks like the EU's. It also avoids imposing compliance costs on India's AI startup ecosystem at a moment when the government's own IndiaAI Mission is trying to lower barriers to entry, not raise them.

The case against it is equally straightforward: principles without enforcement mechanisms have a well-documented tendency, across many policy domains, to remain principles. NITI Aayog's own Responsible AI framework, released in 2020–21, made broadly similar commitments to safety, fairness and transparency — and those commitments did not, on their own, prevent the considerable growth of AI deployment in India over the intervening four years without much visible governance infrastructure behind them.

> *India has chosen to govern AI the way it governs many things: with principles first, capacity second, and enforcement only once the first two have had time to mature. The open question is whether AI's pace of change will wait.*

4. International Context

The Guidelines' voluntary, principles-first design places India at a deliberate distance from the two dominant global reference points for AI regulation, and that positioning is itself a policy signal worth unpacking.

- **The European Union's AI Act,** in force since 2024, is a comprehensive, binding, risk-tiered statute — the model India has explicitly chosen not to follow, at least for now, prioritising speed and flexibility over the EU's more prescriptive, compliance-heavy approach.

- **The United States' approach,** relying on executive orders and sector-specific regulation rather than comprehensive federal legislation, is closer in spirit to India's guidelines-first posture, though without India's explicit central coordinating architecture (the Governance Group, Expert Committee and Safety Institute).

India's decision to host the 2026 AI Impact Summit gives this positioning outsized importance: other Global South economies weighing their own AI governance path will likely look to how India's guidelines-first bet performs over the coming year as a reference case, for better or worse.

5. Stakeholder Implications

- **For AI developers and startups:** The absence of binding compliance obligations lowers near-term costs, but also leaves developers without a clear, stable compliance target to build toward — a trade-off some may prefer and others may find creates its own uncertainty.
- **For civil society and digital rights groups:** The Guidelines' voluntary nature gives advocacy groups a continued, live opening to push for statutory backing, rather than facing a settled framework that would foreclose further debate.
- **For other Global South governments:** India's guidelines-first experiment offers a lower-cost reference model than the EU's comprehensive statute, and its success or failure over the next year will likely influence how several other developing economies approach their own AI governance choices.

Key Terms

The seven sutras: The seven guiding principles of the India AI Governance Guidelines: Trust, Human-Centricity, Fairness & Equity, Accountability, Understandable by Design, Safety & Robustness, and Inclusive & Sustainable Innovation.

AI Governance and Economic Group: The high-level, inter-ministerial body tasked with steering the development of India's national AI governance strategy under the new Guidelines.

IndiaAI Safety Institute: The technical body charged with AI safety research, standards development, system testing and risk evaluation under the Guidelines' institutional architecture.

6. Recommendations

1. Publish a concrete operational timeline for the IndiaAI Safety Institute within the next two quarters.

Specific testing protocols, standards-development milestones and staffing commitments would meaningfully strengthen the credibility of the guidelines-first approach, converting a general mandate into a trackable one.

2. Define, in advance, the conditions under which the Guidelines would be escalated toward binding rule.

A pre-committed review trigger — for instance, documented harm above a defined threshold, or a fixed post-implementation review date — would give the voluntary framework real credibility without abandoning its current flexibility.

3. Clarify explicitly how the Guidelines' principles interact with the DPDP Act's existing consent requirements for AI training data.

Developers currently have to interpret this interaction themselves; explicit guidance would reduce both compliance uncertainty and the risk of inconsistent enforcement.

4. Establish a public reporting mechanism for AI-related harms,

even in the absence of binding rule, so that the evidentiary basis for any future move toward statute is being built now rather than assembled retroactively once harms have already accumulated.

7. Conclusion

The India AI Governance Guidelines are a genuine first step — considerably more developed than the Responsible AI principles that preceded them, and backed by institutional structures that, if resourced properly, could give a voluntary framework real substance. But a guidelines-first strategy is a bet on institutional follow-through, not a settled governance model, and the coming year — starting with whether the IndiaAI Safety Institute produces anything concrete — will be the first real test of whether that bet is paying off.

References

Ministry of Electronics and Information Technology / Press Information Bureau — India AI Governance Guidelines (November 2025)

NITI Aayog — Responsible AI framework (2020–2021), for comparison

Digital Personal Data Protection Act, 2023 — consent provisions relevant to AI training data

Contemporaneous reporting and early expert commentary on the Guidelines' release (November–December 2025)