

Executive Brief – Co-Creating the Future: South–North Collaboration for AI Impact

From Principles to Practice:

Operationalizing Trusted AI for the Global South

Key Outcomes from the India AI Summit DC Pre-Summit

India AI Impact Pre-Summit | Washington, D.C.
January 14, 2026

Convened by: Beju Rao, Uvika Sharma, Padmini Soni

1. Strategic Context and Purpose

The India AI Summit DC Pre-Summit was convened as a strategic prelude to the India AI Impact Summit 2026, bringing together policymakers, technologists, academics, industry leaders, and civil society actors from the Global South and North. The objective was not to restate high-level principles, but to examine how AI is being operationalized in real-world, high-stakes contexts, and what design, governance, and institutional choices are required to ensure AI advances responsibly and inclusively.

Across six hours of keynotes, panels, workshops, showcases, and research presentations, a clear shift emerged: the global AI conversation is moving decisively

from aspiration to execution. As AI systems become more agentic, embedded, and autonomous, the central questions are no longer whether AI can be deployed, but how it is governed, monitored, explained, and integrated into human systems at scale.

India is uniquely positioned at this moment. Its Digital Public Infrastructure (DPI), population-scale deployment experience, and “AI for All” orientation provide a foundation for demonstrating how trusted AI can be implemented in practice, particularly for the Global South. This brief synthesizes the Pre-Summit’s validated insights and outcomes to inform ongoing deliberations leading up to the India AI Impact Summit 2026 and to identify pathways for continued collaboration.

2. A Cross-Cutting Insight: Trust in AI Is Operational, Not Philosophical

A unifying theme across all sessions was that **trust in AI is earned through operations, not intent statements**. Trust does not emerge from model performance alone, nor from abstract ethical commitments. It is the product of concrete design decisions, governance mechanisms, and ongoing oversight.

Participants consistently identified predictable failure modes that undermine trust in deployed systems, including prompt injection, data leakage, shadow AI usage, hallucinations, weak grounding, and governance structures that cannot keep pace with system drift. Of particular concern was the risk of silent failure, situations in which AI systems degrade over time due to distribution shifts or feedback loops, without triggering obvious alarms.

The consensus view was that responsible AI requires design-time and run-time **governance**, with continuous monitoring, adversarial testing, and human accountability embedded into systems that operate in real-world environments.

3. Validated Outcome I: Trust and Safety in AI Deployment (DPI-Ready)

Outcome: Concrete recommendations for governance and responsible deployment, particularly in Digital Public Infrastructure and other high-stakes public systems.

Several practical approaches gained strong alignment:

- **The GIPA Framework (Generality, Intelligence, Power, Agency)** was proposed as a pragmatic way to assess potential harm. Rather than treating risk as linear, the framework recognizes that risk increases non-linearly as AI systems gain autonomy, intelligence, and decision-making power. This enables

policymakers and system designers to calibrate governance intensity to actual risk.

- **Decentralized, agentic architectures** were favored over highly centralized models. In this vision, citizens interact through personal AI agents that are authenticated, traceable, and auditable, reducing systemic risk while increasing individual agency.
- **Separation of planning and execution** emerged as a critical design pattern. AI systems may propose plans or recommendations, but humans must approve intent, with final execution remaining deterministic. This preserves accountability while enabling efficiency.
- **Proactive monitoring for silent failure** was identified as essential. Rigorous adversarial testing, early-warning systems for distribution shifts, and feedback-loop monitoring are necessary to ensure systems remain safe after deployment.

These approaches are directly applicable to DPI-enabled services, where scale magnifies both benefit and risk.

4. Validated Outcome II: AI Literacy and Explainability as Infrastructure

Outcome: Community-informed guidance to strengthen public AI literacy for educators, civil society, and service providers.

The AI literacy discussions emphasized that explainability must be meaningful to the audience, not merely technically correct. Transparency requirements differ across users, citizens, clinicians, administrators, and regulators, and systems must be designed accordingly.

Key insights included:

- **Audience-specific transparency:** Explainability should be tailored to user context, enabling informed trust rather than overwhelming users with technical detail.
- **Human-centric health support:** AI systems show particular promise in preventive and personalized care, especially in regions with limited healthcare access, provided they are designed to empower individuals rather than replace professionals.

- **Regulatory and liability literacy:** Public understanding of legal protections, responsibilities, and recourse mechanisms is essential for responsible adoption.
- **Global standards harmonization:** While full regulatory convergence is unlikely, baseline international standards for data quality and system integrity are critical to sustaining trust across borders.

AI literacy was consistently framed not as a one-time educational effort, but as **foundational infrastructure** for long-term adoption.

5. Validated Outcome III: Global South Differentiation and Collaboration

Outcome: Transferable design patterns and strategic adaptations for the Global South.

Participants cautioned that as generative AI commoditizes "knowledge at scale," economies that previously competed on scale alone face increasing pressure. Durable advantage will come from value-added thinking, creativity, and domain-specific expertise.

Three patterns stood out:

- **Value-Added Thinking at Scale:** Shifting from execution-only models toward human-AI collaboration that emphasizes judgment, creativity, and contextual understanding.
- **Pragmatic Regulatory Alignment:** Recognizing that regulatory harmonization often follows crises, and in the interim relying on bilateral and multilateral engagement, standards forums, and trade-based collaboration rather than waiting for a single global framework.
- **Strategic Democratization of AI:** Treating AI access as a tool for human-capital development, while selecting regulatory choices that protect trust without freezing local innovation.

This framing positions the Global South not as a passive recipient of AI, but as an active co-architect of globally relevant models.

6. Validated Outcome IV: Digital Public Infrastructure as a Population-Scale AI Operating System

Outcome: Scalable, modular use cases demonstrating social empowerment and economic participation.

The DPI showcase reinforced that trust at scale is enabled not by software alone, but by **standards, governance, interoperability, and human oversight**. Identity, payment, and consented data layers together function as a population-scale operating system that lowers barriers to entry for services such as financial inclusion.

Illustrative applications included:

- **Micro-lending and risk reduction** through DPI-enabled data, combining automation with human oversight.
- **Citizen services** such as health triage and grievance redressal, where maintaining humans in the loop is essential to prevent silent automation without accountability.

DPI was repeatedly cited as one of India's most powerful contributions to the global AI ecosystem.

7. Validated Outcome V: Building AI Co-Workers

Outcome: Design principles for culturally adaptive AI agents and safe deployment.

The "AI co-worker" framing emphasized that responsible scale depends less on raw model capability and more on **workflow integration and operational discipline**. Key principles included:

- Workflow integration over model performance
- Ethics-by-design and institutional capacity building
- Application-led inclusion in areas such as agriculture, rural health, and education
- Measuring success through friction reduction and cycle time, while maintaining auditability, data sovereignty, and security

8. Emerging Innovation and Talent Signals

Student and researcher presentations highlighted early-stage innovation and intergenerational collaboration, including civic agents for public-sector capacity, predictive healthcare using machine learning digital twins, fiduciary AI for investment advice,infrastructure as software using generative AI, and system-level optimization comparing LLM chatbots with flow-based systems.

These signals point to a growing pipeline of globally relevant innovation aligned with India's priorities.

9. Actionable Pathways for Collaboration

Based on these outcomes, participants identified several collaborative pathways that could support the India AI Impact Summit 2026, including deploying AI literacy and explainability across stakeholder populations, transporting the US government-funded innovations to Global South and India, co-developing trust and safety playbooks for DPI and expanding AI-based DPI globally, refining GIPA-aligned risk assessment guidance, supporting working groups or pilots, and contributing Global South perspectives to international standards discussions.

10. Closing

The DC Pre-Summit demonstrated that trusted, inclusive AI is achievable when principles are translated into operational practice. The insights synthesized here are offered as a living input to the India AI Impact Summit 2026 and as a basis for continued collaboration in advancing responsible AI for India and the Global South.

Additional session-level summaries, research artifacts, and use-case documentation are available upon request.