Doot: The AI Agent for Every Indian Citizen

National Architecture of “Agent One”

Citizen AI Agents, DPI-Scale Deployment, and the Agentic Trust Stack

Foreword

India is hosting the AI Impact Summit in February 2026, building on the global momentum that began at Bletchley Park in November 2023, followed by the Seoul AI Summit in May 2024 and the Paris AI Action Summit in February 2025, which India co-hosted. As part of this continuum, India has the unique opportunity and responsibility to shape an inclusive, impact-driven agenda for artificial intelligence, one rooted in the principles of People, Planet, and Progress.

The India AI Mission has organized this effort around seven thematic working groups, our seven Chakras, spanning AI for economic growth and social good, democratization of AI resources, inclusion for social empowerment, human capital, science, resilience, and innovation and efficiency. Over 100 countries and leading international organizations participate in these working groups, whose outputs will feed directly into the Leaders' Declaration. We are expecting representation from 15 heads of government, over 50 ministers, more than 40 CEOs of leading global and Indian companies, and around 500 prominent figures from the global AI ecosystem.

A key priority for the India AI Mission has been building sovereign compute capacity. Our initial target of 10,000 GPUs has been far exceeded, and we continue to scale through an empanelment model that we believe can be replicated across the Global South. We are also supporting the development of indigenous foundational language models, with basic versions expected to be operational before the summit and advanced versions by end of 2026.

Among the summit's Chakras, the Inclusion for Social Empowerment theme focuses specifically on enabling citizen-centric AI solutions that strengthen last-mile service delivery. This white paper, contributed by a scientific panel drawn from the academic and industry community, engages directly with that agenda. It explores how the next layer of India's digital public infrastructure might bring AI to citizens who face the highest barriers of literacy, language, and complexity, through intelligent agents built on existing DPI rails such as Aadhaar, DigiLocker, and UPI. These ideas are consistent with the summit's emphasis on multilingual, accessible, and locally relevant AI systems that translate global AI capabilities into measurable development outcomes.

Our governance model seeks to balance innovation and safety, not limiting innovation on the basis of safety alone, but placing innovation at the forefront while creating a responsible balance through self-regulation. We welcome contributions from the global research and technology community that advance this vision of AI as an instrument for inclusive growth and citizen empowerment.

K. Mohammed Y. Safirulla, IAS
Director, IndiaAI Mission
Ministry of Electronics and Information Technology (MeitY)
Government of India
February 2026

Table of Contents

Executive Summary: AI Agent as Envoy for All

We propose a national-scale architecture in which every citizen possesses a personal AI agent 'Doot': a digital twin running predominantly on the citizen's device and empowered by authenticated identity and documents already available through India's Digital Public Infrastructure (DPI) rails such as Aadhaar identity, DigiLocker paperless governance, Unified Payment Interface (UPI), and others. We call the architecture "Agent One". The Doot agent enables citizens to discover, access, and complete government and private services through natural language, voice, and local context—especially improving usability for citizens who face literacy, language, and complexity barriers.

The core conclusion is that Agent One is technically feasible but limited by governance and adoption constraints. The strongest insights converge on the need for:

- **Trust Infrastructure**: End-to-end privacy with agent actions that are attributable, bounded, and safe.
- **Agentic DPI Mindset**: At a scale of billions of users and potentially trillions of interactions.
- **Resilience-first Engineering:** Specification before intelligence, contract-bound APIs, auditability, and observability to prevent cascades.
- **Edge-first Compute:** Edge compute and efficient models enable on-device inference and lightweight personalization.
- **Adoption and Legitimacy Design:** The gap is often capability versus confidence; citizens must know when to trust versus verify, and must not feel surveilled.
- **Default Governance:** At population scale, the real policy is the set of defaults shipped in devices and platforms.

1. Introduction: Why Doot via Agent One, Why Now?

India has shown that it can use digital infrastructure on a large scale, such as Aadhaar, UPI, DigiLocker, and other DPI components. These rails have made it possible for billions of transactions per year, with almost everyone able to use them. They have also made it possible to store identification, payments, and data.

India is still a high-friction economy, even though it has gone digital. Service delivery is still broken up between apps, forms, websites, and schemes. Digital public infrastructure has its limits because it places significant mental and procedural load on citizens. Finding the correct scheme, figuring out if you're eligible, filling out paperwork, keeping track of your status, and dealing with exceptions all take time, talent, and determination.

AI has made it possible to engage with complicated systems using natural language, which is a new way to do things. These abilities now allow for autonomous agents that can watch, think, plan, and act—things that weren't possible just two years ago.

India has a chance to build the next level of digital public infrastructure: an agent layer that lets people get services by asking for what they need and lets the government offer them on a larger scale with much more efficiency.

2. Doot Vision: Citizen Agent as a "Personal Digital Twin"

Doot is not a new app, portal, or platform. It is an agent layer that sits above India's existing digital rails, converting them into a single, natural-language interface. Instead of navigating fragmented systems, citizens interact with their agent by stating their intent: "I need a birth certificate," "Apply for this subsidy," "Track my pension." The agent then plans and executes the required actions on the citizen's behalf.

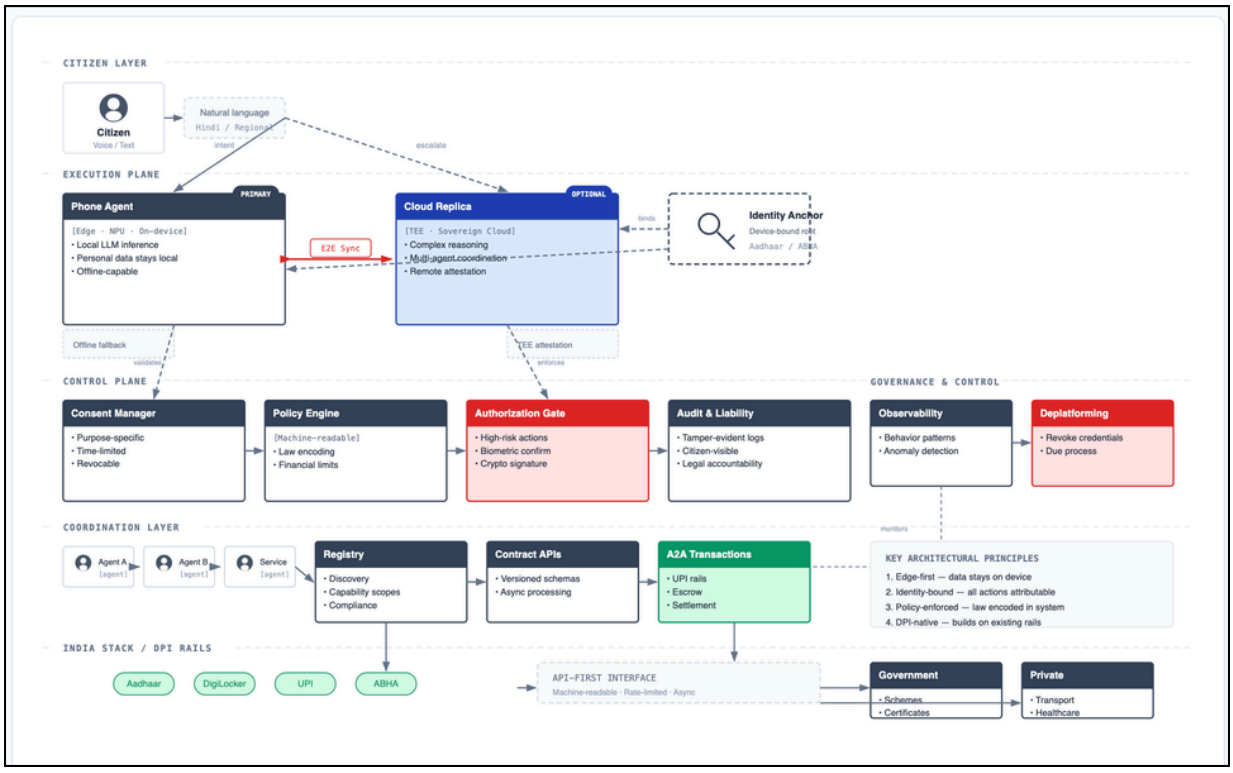

Figure 1: *Agent One: Full System Architecture—from Citizen Layer through DPI Rails*

Each agent is rooted in the citizen's legal and digital identity, acting strictly under that citizen's authority. It is not a government agent or a platform-owned bot. It is the citizen's representative in the digital world, capable of speaking to public and private service providers in a way that is authenticated and auditable. Agents are designed for inclusion, optimized for users who face the highest friction: rural populations, the elderly, migrants, multilingual users, and those with low literacy.

What Agent One is not:

- Centralized model trained on private data. There is no database where citizens' conversations or histories are pooled for training or surveillance.
- Autonomous entity with unchecked authority. The agent cannot invent goals or act beyond the citizen's intent and legally permitted capabilities.

- Surveillance layer. The agent does not monitor the citizen or extract data without explicit consent. All actions are visible, logged, and attributable to the citizen.

3. Motivating Use Case: Sitabai and Kumbh Mela

Sitabai is a 70-year-old pre-diabetic citizen with a limited budget and limited digital literacy. She wants to travel from Patna to Nashik for Kumbh Mela, requiring travel planning, medical guidance, scheme discovery, housing verification, event registration, language support, and care circle coordination.
All with a single voice command: "*Mujhe toh Nasik jana hai Kumbh Mela ke liye. Main kya karu?*"

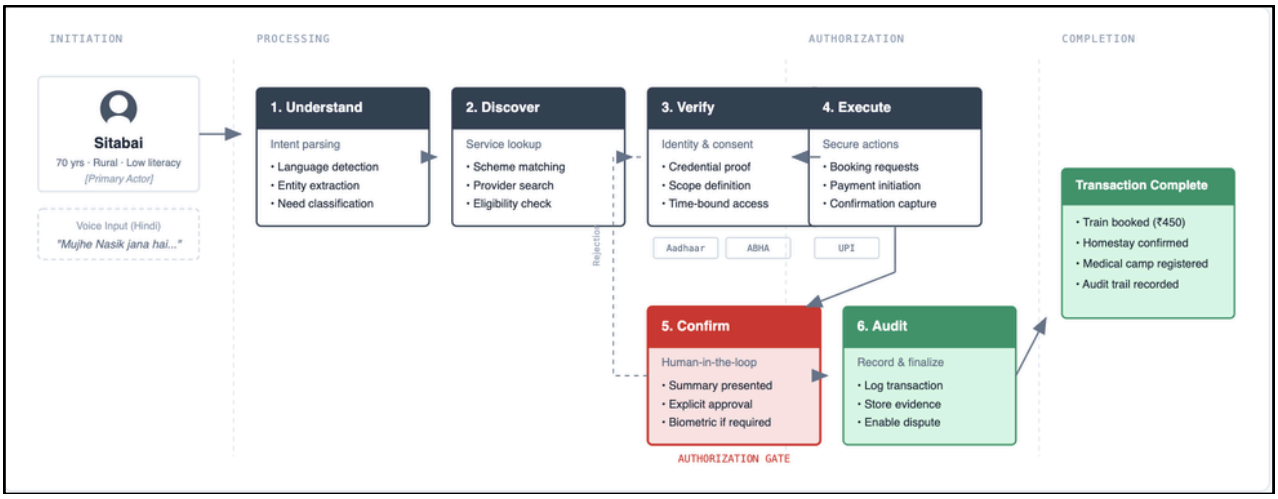

Figure 2: *Sitabai's Journey: End-to-End Agent Flow for Kumbh Mela Pilgrimage*

How the Trust Infrastructure Serves Sitabai

- **Identity**: Her agent proves she is a senior citizen eligible for railway concessions without revealing her birthdate or Aadhaar number—only the verified claim.
- **Consent**: She grants temporary health data access to the Kumbh medical camp, scoped to the event duration and revocable at any time.
- **Transaction**: Her agent pays ₹450 for a verified homestay via UPI, executed through the secure tool layer with full logging.
- **Discovery**: Her agent finds verified transport agents and accommodation providers serving the Nasik route, filtering for accessibility and her budget.
- **Human-in-the-Loop**: Before booking, the agent summarizes the itinerary in Hindi speech and waits for her voice confirmation.
- **Governance**: In the event of a last-minute homestay cancellation, her agent initiates dispute resolution via auditable logs, resulting in an automatic refund.

The agent enables an end-to-end journey while proving Sitabai's authenticity only when necessary and protecting her data by default.

3.1 Prototype: Kumbh Doot Application

A prototype application, Kumbh Doot, has been developed to demonstrate the citizen-facing interface of the Doot agent in the context of the Kumbh Mela use case. The application is accessible at kumbh-connect-now.base44.app and illustrates how natural-language interactions can be surfaced through a mobile-first interface.

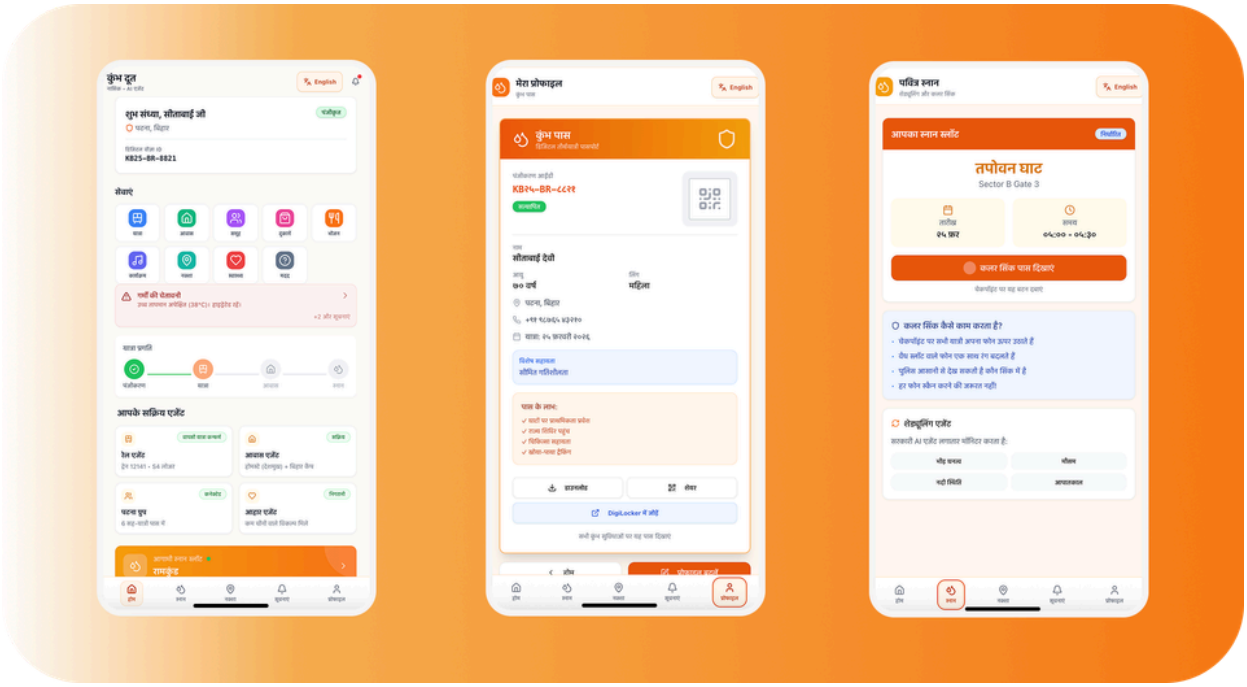

Figure 3: *Kumbh Doot: Prototype Citizen-Facing Application*

4. The Trust Infrastructure

India has grounded its Digital Public Infrastructure (DPI) in trust through techno-legal frameworks that tightly couple code with law. AI agents introduce new risks: opaque decision-making, a large attack surface, and unclear liability. The approach must be grounded in delegated authority: the agent holds no intrinsic power; all authority flows from the citizen, constrained by law, and enforced by technology.

4.1 Identity-bound Delegation

Each agent is cryptographically bound to the citizen's digital identity, anchored in Aadhaar or sector-specific identities, such as ABHA. When the agent interacts with any external service, it presents verifiable credentials: Agent A is acting for Citizen C, for Purpose P, under Scope S. For the high-stakes actions that carry legal or financial consequences, Personhood Credentials (PHC) is defined as a privacy-preserving digital credential that allows a user to prove to an online service that they are a real human person, not an AI bot without disclosing any additional personal information

4.2 Consent and Least Privilege

All authority is governed through fine-grained, revocable consent. Permissions are purpose-specific, time-limited, and capability-scoped. A citizen may allow their agent to access health records for a hospital visit, but nothing beyond that scope.

4.3 Secure Tool Execution

Agents cannot directly execute actions in the outside world. All network calls, form submissions, payments, and data retrieval occur through a secure tool layer that validates parameters, enforces policy, and logs outcomes. The language model can propose actions, but only policy-compliant tools can execute them. All external inputs are treated as hostile by default.

4.4 Human-in-the-Loop for High-Risk Actions

For critical or irreversible actions—financial transfers, legally binding documents, identity record changes, data sharing with third parties—the agent must obtain explicit human confirmation via biometrics, PIN, or secure confirmation before the system issues cryptographic authorization.

4.5 Policy-Bound Reasoning

Agent One is constrained by a machine-readable policy layer encoding law, financial limits, and data-protection rules. Even if the model proposes an unsafe plan, the policy engine prevents execution.

4.6 Privacy

Agents and models run on-device or within Trusted Execution Environments (TEEs). Within a TEE, the agent's memory, inputs, and reasoning state are encrypted in use, inaccessible even to infrastructure operators. The citizen and state can verify that only approved code is running through remote attestation.

4.7 Auditability and Forensics

Every data access, tool invocation, and transaction is recorded in tamper-evident logs visible to the citizen and, where legally required, to regulators. This ensures accountability, dispute resolution, and rapid detection of abuse.

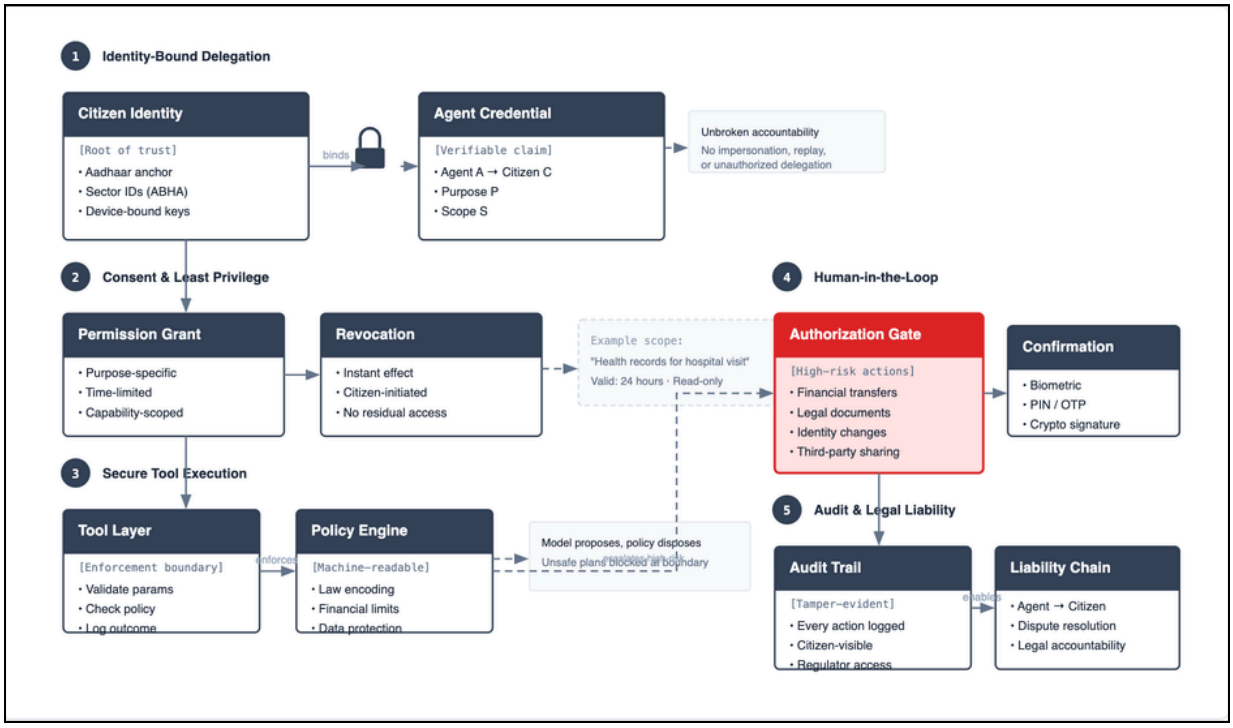

Figure 4: *Trust Infrastructure: Identity, Consent, Tool Execution, Authorization, and Audit*

5. Deployment at DPI Scale

Population scale requires architecture that assumes extreme heterogeneity of devices and connectivity, highly variable literacy and language capability, bursty demand (events, emergencies), and adversarial conditions (fraud, social engineering, and deepfakes).

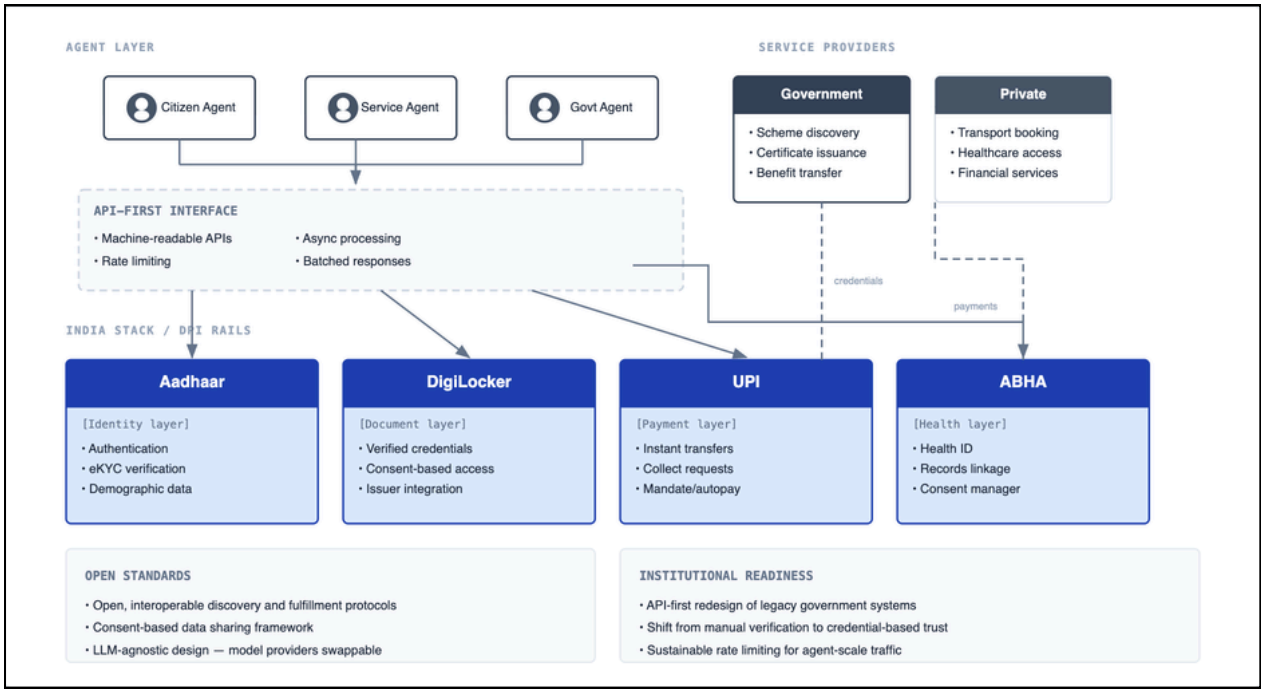

Figure 5: *Deployment Architecture at DPI Scale*

5.1 Deepfake Readiness

AI will unleash deepfakes in voice, video, documents, and real-time impersonation. This is not a future threat; it is a present reality that will intensify. Agent systems must embed:

- **Verification Rails:** A cryptographic attestation to verify human-initiated actions. PHC closes the gap that deepfakes exploit. For high-stakes interactions, PHC verification should be a mandatory precondition along with other verification methods like biometric, voice, and video.
- **Anomaly Detection:** Behavioral analysis to detect agents acting outside normal patterns.
- **Provenance Tracking:** Clear audit trails establishing the origin and chain of custody for all documents and credentials.

5.2 Equity

A key risk is that AI worsens inequality through asymmetrical access to information and capability. Citizens with better devices, connectivity, and literacy could gain a disproportionate advantage, widening rather than closing digital divides.

Citizen agents must reduce, not increase, these divides. This requires voice-first interfaces in regional languages, offline capability for low-connectivity areas, design that assumes minimal digital literacy, and graceful degradation that maintains core functionality even on low-end devices.

5.3 Institutional Capacity

Agents can make millions of requests around the clock. Government systems must be prepared for this shift:

- **API-first Redesign**: Legacy systems must expose machine-readable interfaces for agent interaction.
- **Cryptographic Trust:** Shift from manual verification to credential-based trust reduces institutional burden.
- **Asynchronous Processing:** Agents wait for batched responses rather than demanding real-time processing.
- **Rate Limiting:** Sustainable throttling that prevents abuse while maintaining service quality.

5.4 Open Standards and Indic Models

Interoperability requires open standards to prevent monopolistic capture and enable sovereign governance. Indic models must reflect language diversity (22 scheduled languages, hundreds of dialects), cultural context, and local data distributions. The architecture must be LLM-agnostic, allowing model providers to be switched without citizen disruption.

6. Infrastructure: Edge-first Compute

The country must choose between replicating frontier-style centralized infrastructure (expensive, ecological cost) or investing in hyper-efficient edge AI. Agent-scale computing is not about massive compute clusters but networks of micro-AIs with complex trust, consent, and coordination.

6.1 Dual Agent Framework

Each citizen has two synchronized agents:

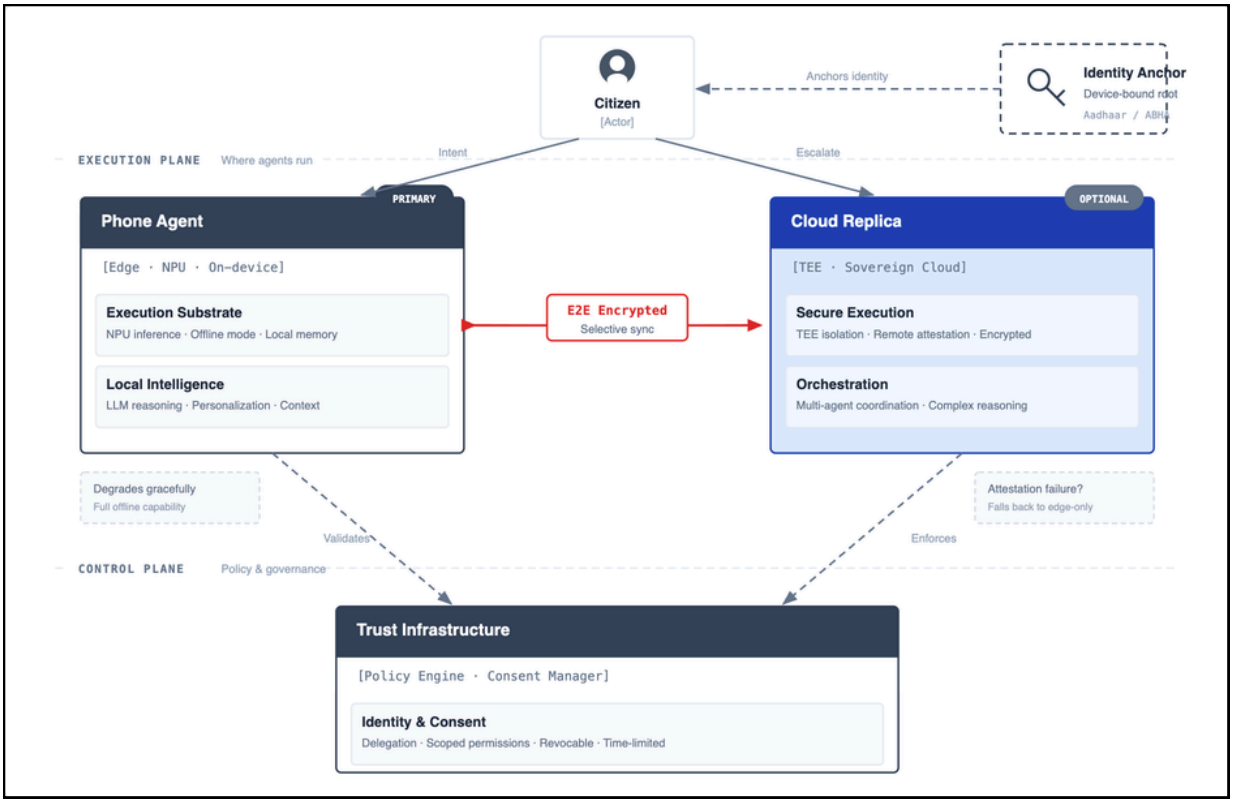

Figure 6: *Dual Agent Framework: Phone Agent (Edge) and Cloud Replica*

Phone Agent (Edge): Runs on-device chips. Handles local inference, personal data, conversation history, and voice interactions. Works offline for basic functions. This is the citizen's primary interface.

Cloud Replica Agent: Hosted on sovereign infrastructure (similar to DigiLocker architecture). Handles complex reasoning tasks, multi-agent coordination, and credential verification. Syncs with the phone agent through encrypted channels. Provides continuity when the citizen switches devices or loses connectivity with agent memory artifacts including conversation context, preference state, and task history stored in DigiLocker as citizen-owned documents.

Critically, the architecture allows multiple cloud replicas. A "reference cloud replica agent" operated by the government (sovereign), similar to BHIM for UPI, establishes the baseline, while private players are permitted to develop their own solutions subject to suitable legal and technological restrictions. This design mirrors the National Digital Health Mission (now ABDM) architecture, which allows more than one Health Information Exchange and Consent Manager.

This dual model ensures privacy (sensitive data stays local), resilience (works offline), and capability (cloud handles what edge cannot).

6.2 Edge AI Capabilities

Phones and wearables are entering a decisive shift where edge intelligence becomes the default:

- **Dedicated Edge-device**: Always-on edge devices process multi-modal sensor data at micro-watt power levels, enabling persistent inference without waking high-power processors.
- **Neuromorphic Architectures**: Computation placed closer to memory achieves orders-of-magnitude improvements in energy per decision, making always-on operation feasible in low-connectivity environments.
- **On-device Personalization**: Lightweight training and calibration occur locally using citizen-specific data without exporting to the cloud, preserving privacy while improving accuracy. Quantized Indic language models make this feasible across low-end devices, ensuring personalization in the linguistic and cultural context of Indian citizens.

6.3 Why Edge-first at Population Scale

These advances enable private local inference where sensitive signals are processed on-device, with only minimal artifacts shared when escalation is necessary. At population scale, edge-first is the only sustainable path—intelligence scales with devices already in citizens' hands rather than centralized infrastructure.

7. Agentic DPI Requires New Governance Structures

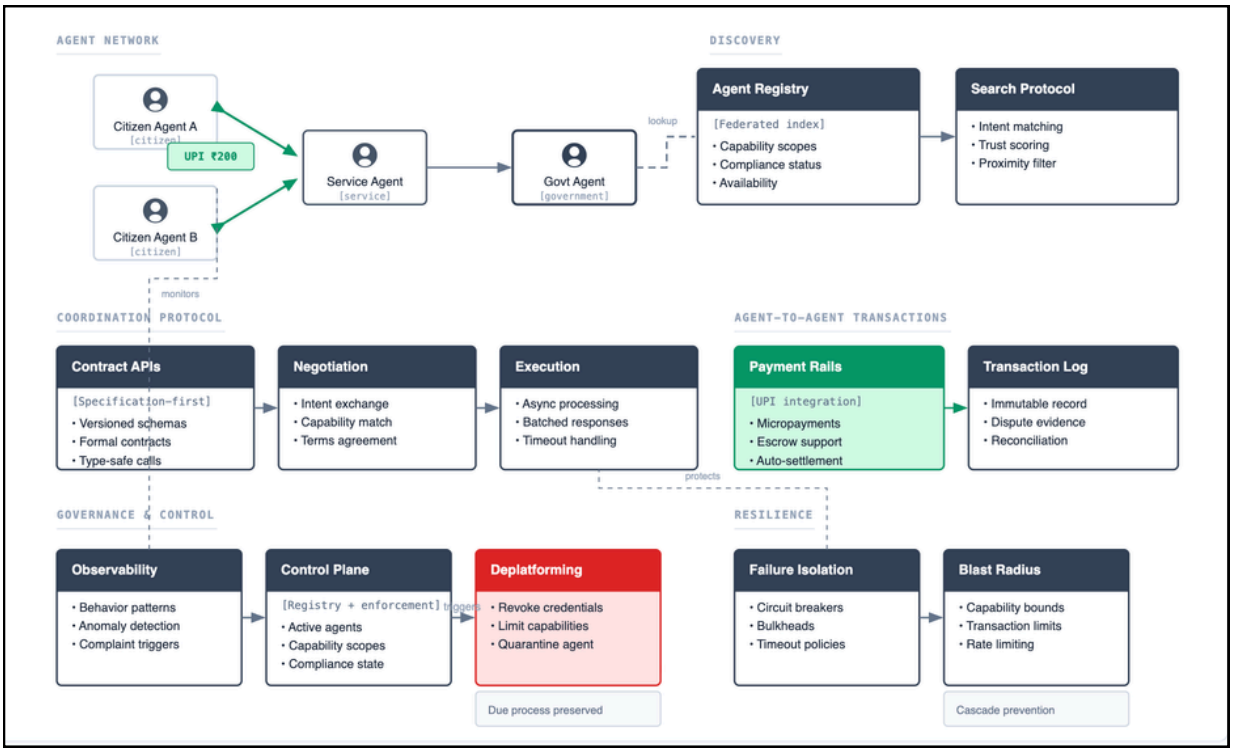

Figure 7: *Governance Framework for Agentic DPI*

7.1 Identity Anchoring Enables Accountability

Agentic systems linked to identity enable:

- Extension of human society law into agent society
- Liability: "if my agent misbehaves, I'm responsible"
- Incentive alignment and economic participation by citizens

This is not merely a technical choice. It is the foundation that makes agent governance tractable. Without identity anchoring, rogue agents become untraceable, liability becomes unassignable, and the system devolves into an ungovernable commons. The cryptographic binding defined in Section 4.1 makes every agent action attributable to a citizen, enabling legal and economic accountability.

7.2 Deplatforming Mechanics

Agents can be good or malicious. Thus society needs:

- Observability into agent behavior patterns and anomalies
- Detection of rogue agents through behavioral signatures and complaint triggers
- Mechanisms to remove compromised agents from circulation without disrupting legitimate citizens
- Governance that is auditable and participatory
- Agent control plane that balances rapid response with due process

The control plane operates as a registry and enforcement layer. It maintains the list of active agents, their capability scopes, and their compliance status. When an agent is flagged—through automated detection or citizen complaint—the control plane can revoke credentials, limit capabilities, or quarantine the agent pending investigation.

7.3 Regulatory Framework

Agent One requires a governing body analogous to National Payments Corporation of India (NPCI) for UPI or Unique Identification Authority of India (UIDAI) for Aadhaar. This body would:

- Define protocol standards for agent interoperability
- Certify agents and service providers for compliance
- Operate the dispute resolution and arbitration layer
- Coordinate with existing regulators such as Reserve Bank of India (RBI) for financial transactions, MeitY for data protection, sectoral regulators for domain-specific actions

The framework must avoid both over-centralization (which creates bottlenecks and single points of failure) and under-regulation (which enables abuse). A federated model—central standards with distributed enforcement—may be appropriate.

8. Engineering Reality: Resilience Over Throughput

8.1 Specification Before Intelligence

At a national scale, emergent behavior without contracts is chaos. The discipline must come before the intelligence:

- **Specification-driven APIs**: Formal schemas with versioned contracts, not ad-hoc integrations.
- **Contract-bound access**: Explicit capability grants governing what agents can request and execute.
- **Auditable actions**: Tamper-evident logging of all agent operations (Section 4.7).
- **Systematic observability**: End-to-end tracing across agent, tool layer, and service provider.

The tool execution layer (Section 4.3) enforces these contracts at runtime. Even if the model proposes unexpected behavior, the system boundary holds.

8.2 Non-determinism Changes the Game

Traditional DPI is rule-based and deterministic—UPI transactions succeed or fail by well-defined rules. Agentic systems break this assumption: probabilistic, non-deterministic, and capable of cascading failures when one agent's error propagates through interconnected services.

Core design pillars:

- **Blast radius control:** Capability bounds and transaction limits constrain any single agent's potential impact.
- **Failure isolation:** Circuit breakers and bulkheads prevent failures from propagating across the network.
- Incident containment: Rapid detection and quarantine of misbehaving agents before contagion affects other citizens or services.

9. Adoption: The Human Confidence Layer

9.1 Capability versus Confidence

The assumption that adoption follows capability is wrong. The gap is confidence—the citizen's confidence in their own judgment about when to trust the agent versus when to verify. A citizen who does not trust their judgment will either avoid the system or demand human verification for every action.

For populations like Sitabai, training programs are not the path to confidence. Her confidence will be built through interaction with the agent itself, or it will not be built at all. Confidence scaffolding means interaction design that builds citizen judgment as a byproduct of normal use.

9.2 People Must Not Feel Surveilled

Privacy is not only a technical property. It is a condition for psychological legitimacy. A citizen who believes their agent reports their behavior to the state will not use it.

The state has a legitimate interest in fraud prevention and system integrity. Agent One addresses this through Section 4's accountability mechanisms. But accountability must not become surveillance. The architecture ensures oversight is proportionate, governed by law, and limited to actions, not intent.
India's DPI success rests on this balance. Aadhaar works because citizens trust that biometrics authenticate, not track. Agent One extends this into natural language interaction and daily decision-making.

Design implications: Intent versus action separation (deliberation stays on-device; actions are logged). Visible data boundaries in the interface. Citizen-controlled conversation history. Adoption through value, not mandate.

9.3 Defaults as Governance

At a population scale, factory settings are normative. Most users never change defaults. What ships by default is what governs in practice.
Privacy by Design: No data-sharing permissions are default-on. The agent ingests, processes, and discards. Long-term memory and data sharing require explicit opt-in, purpose-bound, and time-delimited.
Protective Defaults: The most vulnerable citizens—those unlikely to customize settings—must be protected by the architecture itself. Defaults are the most powerful tool for democratization because they require no action from the citizen to take effect.

10. Open Questions

Edge-TEE Boundary Definition: What computation must remain on-device versus TEE, what triggers escalation between them, and what are the latency, privacy, and attestation constraints governing each boundary crossing?
Consent at Agentic Scale: How does consent work when intent becomes multi-step, data flows across multiple parties, and actions are continuously negotiated by agents?
Governance and Accountability: What are the enforceable mechanisms for auditing agent decisions, certifying safe agents, and deplatforming rogue agents?
Coordination Layer: What is the minimum agentic protocol layer needed for coordination between citizen agents, state agents, and private enterprises?
Interaction-scale Intensity: What are the sustainable compute cost models, caching strategies, and failure containment designs for billions of daily interactions?

11. Pilot Strategy

Phase 0 — Controlled Research Pilot. Population: AI researchers, government tech teams. Goal: Validate trust boundary, consent UX, observability, contract stack.

Phase 1 — Narrow Citizen Cohort. Population: Defined cohort (elderly travel, healthcare). Goal: Measure adoption, confidence scaffolding, deepfake resilience.
Phase 2 — City-scale Sandbox. Population: District or city environment. Goal: Test agent interactions with real services at scale.
Phase 3 — National Rollout. Goal: Multi-vendor ecosystem, policy alignment.

Kumbh Mela 2027 as a Stress Test: The Kumbh Mela is the world's largest peaceful gathering—20 million or more pilgrims over 3 months. It is the ultimate stress test: multi-jurisdiction, multi-lingual (22 official languages), multi-sector (health, transport, accommodation, safety, spiritual services), and time-bound.

Success Criteria:

- 90%+ task completion rate
- <10% human escalation
- Zero privacy breaches
- 95%+ user satisfaction
- 50%+ cost reduction versus manual coordination

12. Conclusion

Agent One is not merely an AI application. It is the beginning of a new national infrastructure layer: Agentic DPI. Success depends less on model intelligence and more on trust boundaries, resilience engineering, privacy legitimacy, and adoption design.
If India approaches Agent One for Doot with the same mindset that built Aadhaar and UPI—designing what works at scale from day one—it can build a citizen-first agent ecosystem that is safe, sovereign, equitable, and globally influential.
India has a unique opportunity to shape what the internet of agents looks like: open standards enabling innovation with citizens retaining sovereignty, rather than proprietary platforms where gatekeepers extract value. We have done it with payments, identity, and documents. Agent One is the natural next layer.

Acknowledgments

This white paper synthesizes insights from a scientific panel convened to explore the feasibility and design of citizen-scale AI agents for India. We thank the India AI Mission, Government of India, and K. Mohammed Y. Safirulla, IAS, Director, IndiaAI Mission, MeitY, for their support, engagement, and the foreword to this document. We thank Praveen Gedam, IAS, Divisional Commissioner of Nashik and Chairman of the Nashik–Trimbakeshwar Kumbh Mela Authority, for his contribution on multi-provider cloud replica architecture drawing from the ABDM precedent. We acknowledge the Project NANDA Consortium for foundational research on agentic coordination protocols. We are grateful to the panelists and reviewers who contributed their expertise across identity systems, distributed computing, edge AI, governance frameworks, and large-scale system design. The views expressed herein represent a synthesis of scientific panel discussions and do not necessarily reflect the official positions of any affiliated institution or government body.

Authors

Ramesh Raskar, MIT Media Lab
Srikanth Nadhamuni, Founding CTO Aadhaar
Nitin Saxena, IIT Kanpur
Kapil Vaswani, SPARC
Biplab Pal, University of Maryland
Umakant Soni, AI Foundry
Utkarsh, Xmplify.tech
Vijaya Kumar Ivaturi, Crayon Data
Manudev Jain, Indian Revenue Services
Sujith Nair, Beckn Protocol
Ujjwal Kumar, Quantum Alliance
Chris Pease, Project NANDA
Lipika Kapoor, NABU Sciences
Aditya Sharma, Project NANDA
Himanshu Tyagi, Indian Institute of Science
Praveen Gedam, IAS, Nashik–Trimbakeshwar Kumbh Mela Authority
Shekhar Singh, Commissioner of Kumbh Mela, Nashik
Santanu Bhattacharya, Researcher, MIT Media Lab

More Details

https://**DigiDoot**.in
Research, Standards and Partnerships