
Community-First: A Rights-Based Framework for AI Governance in India's Welfare Systems

Sanjnah Ananda Kumar
Independent

With
Howard University & Apart Research

Abstract

The Community-First AI Governance Framework addresses critical challenges in India's automated welfare systems through innovative, community-centered solutions. Our framework mandates 50% representation from beneficiary communities in AI system design, implements local language interfaces, and integrates traditional dispute resolution mechanisms. The implementation follows a strategic 24-month timeline with a total budget of INR 1.05 billion (USD 12.45 million). Phase 1 establishes Community AI Review Boards and develops culturally-sensitive algorithms. Phase 2 implements hybrid verification systems and trains community representatives. Phase 3 scales the system statewide while integrating continuous feedback mechanisms. Success metrics focus on both quantitative and qualitative outcomes. We target reduction in false positives from 10.1% to below 2%, complete implementation of local language interfaces, and 95% community satisfaction rates. The framework's emphasis on community participation and cultural integration sets a new standard for ethical AI governance in welfare systems, making it a viable model for other states and nations facing similar challenges in automated welfare delivery. Regular community audits and public dashboards ensure transparency and accountability, while the current monthly savings of INR 46 million demonstrates the potential for balancing efficiency with social justice.

Keywords: algorithmic accountability, welfare digitalization, community governance

1. Introduction

The integration of AI systems in India's public welfare delivery, particularly through Samagra Vedika, represents a fundamental shift in how social benefits are allocated.

This system affects the livelihood of millions of citizens by making automated decisions about their welfare eligibility through complex algorithmic assessments. With documented cases showing 10.1% of applications being flagged as ineligible and impacting welfare distributions worth Rs 16 Cr annually, the scale of potential algorithmic bias demands immediate policy intervention.

Samagra Vedika's implementation exposes critical gaps between technology and social responsibility in India's welfare delivery. The system analyzes multiple data sources to create citizen profiles, improving efficiency but introducing new vulnerabilities. Key challenges include data quality issues and name variations that disproportionately impact marginalized communities, risking systemic exclusion. The ecosystem involves four stakeholder groups with competing interests: welfare beneficiaries seeking essential services but lacking influence; government administrators focused on efficiency and fraud prevention; technology providers prioritizing system optimization over social impact; and civil society organizations advocating for transparency with limited access to information. This creates an imbalance where technical efficiency overshadows social equity in welfare delivery.

2. Methods

Our analysis combines Samagra Vedika's operational data (2016-2024) with human rights impact research, examining 380 million records across government databases. Primary sources include Al Jazeera's 2024 investigation and Amnesty International's technical audit. We employed a rights-based methodology focusing on both technical performance and human impact, particularly regarding entity resolution algorithms. Key limitations include restricted access to Posidex Technologies' system architecture and limited developer transparency.

Key Findings

Our research identifies three critical issues in Samagra Vedika's implementation: First, entity resolution algorithms struggle with name variations and demographic

inconsistencies, leading to wrongful exclusions. The 10.1% ineligibility flag rate (6,625 out of 65,693 applications) indicates significant algorithmic bias. Second, the system's proprietary nature creates a "black box" effect, preventing citizens from challenging decisions. Third, automated decision-making reduces complex human circumstances to numerical data points, disproportionately affecting marginalized communities with limited digital literacy.

We propose three solutions: First, a Rights-Based Hybrid System combining automated processing with mandatory human review for rejections, including appeal mechanisms and community engagement. Second, an Enhanced Manual System with Digital Support, prioritizing human decision-making while using technology as a tool. Third, a Modified Automated System with Enhanced Safeguards, maintaining automation but adding robust protections. The recommended hybrid approach offers the best balance of efficiency, accountability, and human rights protection while remaining technically feasible. `

3. Results

Policy Recommendation: Community-First AI Governance Framework

Our proposed solution introduces a revolutionary approach to AI governance in welfare systems by integrating community participation with technological innovation. The framework mandates the creation of Community AI Review Boards with 50% representation from beneficiary communities, ensuring decisions reflect local context and needs. The innovation lies in its three-tier structure: participatory algorithm design, cultural context integration, and hybrid rights protection mechanisms. Key components include mandatory local language interfaces, recognition of non-digital identity verification methods, and real-time community feedback loops. This approach fundamentally reimagines welfare delivery by combining traditional governance wisdom with advanced AI capabilities.

The implementation follows a structured 24-month timeline across three phases, with a total budget of INR 1.05 billion (USD 12.45 million). Phase 1 (6 months) establishes Community AI Review Boards and develops culturally-sensitive algorithms. Phase 2 (8 months) implements hybrid verification systems and trains community representatives. Phase 3 (10 months) scales the system statewide with continuous feedback integration.

Our success metrics combine quantitative and qualitative indicators. Key performance targets include:

- Reduction in false positives from 10.1% to below 2%
- 100% availability of local language interfaces
- 95% community satisfaction rates
- 90% reduction in wrongful exclusions within 18 months

The monitoring framework features monthly community review sessions and quarterly impact assessments, ensuring both system efficiency and social justice outcomes. This approach creates a replicable model for ethical AI governance in welfare systems while maintaining clear accountability to affected communities.

4. Discussion

The Community-First AI Governance Framework projects three-tiered impacts: Short-term (1-2 years) implementation of local language interfaces and reduction in false positives to 2%; Medium-term (3-5 years) integration of cultural context and adoption by five states; Long-term (5+ years) achievement of near-zero exclusions and establishment as a global benchmark.

The framework impacts multiple stakeholders: beneficiaries gain improved access but face adaptation challenges; government departments save INR 46 million monthly while investing in infrastructure; technology providers adapt to new guidelines; and civil society organizations require capacity building. The system demonstrates feasibility through political alignment with India's digital goals, economic justification via cost savings, and technical foundation built on existing infrastructure with 65,000+ application capacity.

5. Conclusion

The implementation of the Community-First AI Governance Framework for welfare systems represents an urgent imperative as algorithmic decisions increasingly determine access to essential services for millions of vulnerable citizens. With Samagra Vedika's current error rate affecting thousands of families and monthly welfare distributions worth Rs 4.6 Cr, immediate action is crucial to prevent systematic exclusion of eligible

beneficiaries while maintaining operational efficiency. The documented 10.1% ineligibility flag rate (6,625 applications) demands immediate policy intervention to ensure fair and transparent welfare delivery. The proposed framework offers three critical advantages for implementation. First, it establishes standardized protocols for algorithmic accountability through mandatory community oversight, ensuring welfare systems reflect local needs while maintaining efficiency. Second, the three-phase implementation over 24 months, supported by a detailed budget of Rs 105 Cr, provides a practical roadmap for systematic transformation. Third, by institutionalizing community participation in AI governance, this framework creates a replicable model for other states and nations facing similar challenges in automated welfare delivery. These measures will significantly reduce wrongful exclusions, enhance system transparency, and establish India as a leader in ethical AI governance for social justice.

6. References

- [1] Al Jazeera. (2024). Investigation report on Samagra Vedika's welfare exclusions [Special report].
- [2] Amnesty International. (2024). Technical explainer: Human rights risks in algorithmic welfare systems - The case of Samagra Vedika.
- [3] Government of Telangana. (2019). Samagra Vedika implementation report: Integration of welfare databases.
- [4] National Institute of Public Finance and Policy. (2024). Impact assessment of digital welfare systems in India.
- [5] Digital Rights Foundation. (2024). Algorithmic accountability in welfare delivery systems: A case study of Samagra Vedika.
- [6] OECD. (2024). AI Principles for public sector implementation.
- [7] European Union. (2024). EU AI Act: Guidelines for high-risk AI systems.

