International Journal of Leading Research Publication (IJLRP)

E-ISSN: 2582-8010 ● Website: www.ijlrp.com ● Email: editor@ijlrp.com

Inclusive by Design? Gender, Equity, and Ethical AI in Indian Public Service Delivery

Lt. Dr. Kongala Sukumar

Assistant Professor of Public Administration
MALD Government Degree College,
GADWAL, Pin-509126, Telangana

Abstract:
India's rapid embrace of artificial intelligence (AI) in public service delivery—spanning welfare targeting, biometric authentication, and administrative decision-making—has been framed by policymakers as a route to efficiency, transparency, and inclusive growth. This paper interrogates that framing through a gender lens, arguing that AI systems deployed within Indian e-governance infrastructures risk reproducing and even amplifying pre-existing gender inequities unless deliberately designed for inclusion. Drawing on secondary data from national and international sources, including the Oxfam India Inequality Report, the GSMA Mobile Gender Gap Report, and NITI Aayog's Responsible AI approach documents, the paper examines three interlinked dimensions: the gendered digital divide that conditions access to AI-mediated services; documented instances of algorithmic exclusion in Aadhaar-linked welfare schemes; and the institutional and regulatory apparatus currently governing AI in India. The analysis finds that women, particularly in rural and informal-sector contexts, face compounding barriers of connectivity, digital literacy, and biometric authentication failure that are largely invisible in existing AI governance frameworks. The paper proposes a gender-responsive design framework built around data disaggregation, participatory audits, human-in-the-loop grievance redress, and gender-diverse AI development teams, arguing that ethical AI in public administration cannot be achieved without explicit attention to gender equity at every stage of the design and deployment pipeline.

Keywords: ethical AI, gender equity, digital divide, e-governance, algorithmic bias, public service delivery, India.

1. Introduction

Over the past decade, "digital governance" has become one of the defining vocabularies of Indian public administration, and artificial intelligence its most recent and most consequential extension. What began with the digitisation of records and the introduction of online portals for grievance filing has, since the mid-2010s, evolved into a far more ambitious project: the layering of biometric identity verification, algorithmic eligibility screening, and increasingly automated decision-making onto the delivery of core welfare entitlements. Proponents of this transformation argue that it reduces corruption and leakage, speeds up service delivery, and extends the reach of the state into previously underserved geographies. Critics counter that a transformation conceived and measured primarily in terms of administrative efficiency risks treating the citizen as a data point to be verified rather than a rights-holder to be served, and that this risk falls unevenly across India's population.

Artificial intelligence has moved from the periphery to the centre of Indian public administration. Direct benefit transfer systems, biometric authentication for welfare disbursal, predictive policing tools, chatbot-based grievance redress, and algorithmic eligibility screening for schemes ranging from the Public Distribution System (PDS) to the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) increasingly rely on automated or semi-automated decision-making. The Government of India has articulated an ambitious vision for this transformation. NITI Aayog's National Strategy for Artificial Intelligence, published in 2018 under the banner "#AIforAll," positioned AI as a tool for inclusive growth across healthcare, agriculture, education, smart mobility, and infrastructure (NITI Aayog, 2018). This was followed by a two-part Responsible AI approach document, which laid out principles of safety, equality, inclusivity, and non-discrimination, and subsequently proposed operational mechanisms for realising them (NITI Aayog, 2021a, 2021b).

Yet the promise of AI-for-all sits uneasily alongside the lived reality of India's digital and administrative landscape, which remains deeply stratified by gender, caste, class, and geography. Feminist scholars of technology have long argued that technologies are not neutral instruments but artefacts shaped by, and reproductive of, the social relations in which they are built (Noble, 2018; Eubanks, 2018). When automated systems are layered onto an already unequal terrain of access and administrative capacity, the risk is not simply that women are left behind by digital transformation, but that AI systems actively misclassify, exclude, or misidentify them, converting a pre-existing social disadvantage into a computational error with material consequences—delayed pensions, denied rations, or lost employment guarantees.

The stakes of this question are not merely technical. Public service delivery in India mediates access to food security, income support, healthcare, and education for hundreds of millions of citizens, and women are disproportionately reliant on several of these entitlements—as primary caregivers accessing maternal and child health schemes, as MGNREGA workers in states with high female labour-force participation in public works, and as heads of household in a growing share of rural families affected by male out-migration. Any systematic degradation in the reliability of these services for women therefore carries consequences that extend well beyond individual inconvenience, touching on food security, health outcomes, and household economic stability. At the same time, the scale at which these systems operate means that even a modest per-transaction error rate translates into a very large absolute number of affected citizens: a biometric authentication system serving hundreds of millions of beneficiaries need only fail for a small percentage of transactions to generate exclusions numbering in the millions.

This paper asks a deceptively simple question: is Indian public service delivery, as it is being redesigned around AI and automated decision-making, inclusive by design—or only inclusive by rhetoric? To answer this, the paper examines three interlocking bodies of evidence. Section 3 reviews the gendered digital divide that determines who can meaningfully access AI-mediated public services in the first place. Section 4 documents recorded cases and empirical studies of algorithmic exclusion, with particular attention to Aadhaar-linked welfare systems. Section 5 maps the existing institutional and regulatory architecture for AI governance in India and assesses the extent to which it addresses gender as a distinct axis of harm. Section 6 proposes a framework for gender-responsive AI design in public administration, and Section 7 discusses the broader implications of the analysis before the paper concludes with directions for policy and further research.

2. Conceptual Framework: Gender, Equity, and Algorithmic Governance

The paper situates its analysis at the intersection of two literatures: feminist science and technology studies, and the emerging field of algorithmic accountability. Feminist STS scholarship has established that technological systems encode the assumptions, priorities, and blind spots of their designers, and that these encoded assumptions disproportionately disadvantage groups who are underrepresented in the design process (Noble, 2018). Applied to public administration, this suggests that AI systems built primarily by, and tested primarily on, data reflecting urban, literate, digitally-connected, and often male populations will perform less reliably for those outside this profile.

The algorithmic accountability literature contributes a complementary vocabulary: bias can be introduced at the level of data (non-representative training sets), design (the choice of variables and objective functions), or deployment (how a system's outputs are used by frontline bureaucrats) (O'Sullivan et al., 2019, as cited in recent AI governance scholarship). In the Indian welfare context, this distinction matters because failures are frequently attributed to "technical glitches" in biometric matching or connectivity, obscuring the extent to which these are foreseeable, patterned failures that fall disproportionately on women, elderly citizens, manual labourers whose fingerprints are harder to match, and residents of poorly networked rural areas (Khera, 2019).

This paper adopts a working definition of inclusive design for public-sector AI as design that (a) is informed by disaggregated data on how different social groups access and are affected by a system; (b) subjects automated decisions to meaningful human review and grievance redress; and (c) involves affected communities, including women, in the design and evaluation process rather than treating them solely as end-users or data subjects. Against this definition, the following sections assess the current state of AI-enabled public service delivery in India.

A related conceptual distinction, developed within algorithmic-fairness scholarship, separates formal from substantive equality in system design. A formally equal system treats every citizen identically—the same biometric threshold, the same verification workflow, the same digital interface—regardless of the circumstances under which different groups encounter that system. A substantively equal system, by contrast, calibrates its design to the differentiated starting points of different users: recognising, for instance, that a manual labourer's fingerprint may require a lower matching threshold, or that a woman without independent phone ownership may need an alternative channel for authentication that does not depend on a shared household device. Much of the criticism levelled at Aadhaar-linked systems in the literature reviewed for this paper is, at root, a criticism of formal rather than substantive equality: the system treats all citizens the same at the point of verification, while the antecedent conditions of connectivity, documentation, and bodily variation are anything but the same (Khera, 2019; Sambasivan et al., 2021). This distinction underpins the analytical structure of the remainder of the paper, which traces gendered inequality through three stages—access, algorithmic treatment, and governance response—corresponding to Sections 4 through 6.

It is also useful to distinguish between three sources of algorithmic harm that recur throughout the literature on AI and public administration: allocative harm, in which a system distributes a resource or opportunity unequally across groups; representational harm, in which a system's outputs reinforce demeaning or inaccurate stereotypes about a group; and what might be termed procedural harm, in which the process by which a decision is reached—its transparency, contestability, and the dignity afforded to the person subject to it—is itself unequal across groups even where the final allocation is formally identical (Noble, 2018). Much of the empirical literature on Aadhaar-linked welfare systems reviewed in Section 5 documents allocative harm most directly, in the form of denied or delayed benefits. However, the paper argues that procedural harm deserves equal attention: a woman who must travel further, wait longer, or navigate a more confusing appeals process than a similarly situated man to correct an identical authentication error experiences a form of inequality that aggregate benefit-disbursal statistics alone would not capture. A genuinely inclusive design framework, accordingly, must attend to the fairness of the process as well as the fairness of the outcome.

3. Methodology

This paper adopts a qualitative, document-based research design, synthesising secondary data from three categories of sources: (a) government and policy documents, principally NITI Aayog's National Strategy for Artificial Intelligence and its Responsible AI approach papers; (b) survey-based data from international bodies tracking digital and gender inclusion, notably the GSMA Mobile Gender Gap Report and the Oxfam India Inequality Report; and (c) peer-reviewed and grey-literature case studies documenting algorithmic exclusion in Aadhaar-linked welfare delivery. Data were thematically analysed and organised around the three research questions outlined in Section 1: access, exclusion, and governance. Sources were prioritised for inclusion where they met at least one of three criteria: publication or endorsement by a recognised national or international institution (such as NITI Aayog, GSMA, or Oxfam India); peer-reviewed or conference-reviewed status within the algorithmic fairness or science and technology studies literature; or documented, traceable case evidence of a specific instance of algorithmic exclusion, as opposed to anecdotal or unverified reporting. Given the descriptive and interpretive nature of the available secondary data, the paper does not claim causal inference but instead offers a structured evidentiary synthesis intended to inform policy design and to identify priorities for future primary research, including field-based audits of specific AI-enabled schemes.

4. The Gendered Digital Divide: Evidence and Implications

Access is the precondition for any discussion of algorithmic fairness: a citizen who cannot reach a digital service cannot be harmed or helped by how well that service is designed. India's digital divide is well documented, but its gendered dimension is particularly stark. According to Oxfam India's Inequality Report 2022, roughly 70% of India's population has poor or no connectivity to digital services, and this deficit is unevenly distributed: among the poorest 20% of households, only 2.7% have access to a computer and 8.9% to internet facilities (Oxfam India, 2022). Layered onto this general deficit is a distinct gender gap. The same report finds that Indian women are 15% less likely than men to own a mobile phone and 33% less likely to use mobile internet, with India recording the widest gender gap in mobile use—40.4%—of any country in the Asia-Pacific region for which comparable data exist (Oxfam India, 2022).

International survey data corroborate this picture. GSMA's Mobile Gender Gap Report 2022, based on face-to-face surveys across ten low- and middle-income countries, found that women in India remained roughly 7% less likely than men to own a mobile phone at the regional level, but that mobile internet use in India had stalled at 30% for women even as it rose to 50% for men over the same period—implying that the gender gap in actual internet use, as opposed to bare ownership, was widening rather than narrowing (GSMA, 2022). This distinction between ownership and use is analytically important for AI-enabled governance: a shared household phone, frequently controlled by a male family member, does not translate into independent access to an AI chatbot for grievance redress, a mobile banking application linked to a direct benefit transfer, or a self-service portal for scheme enrolment.

Table 1- *Selected Indicators of the Gendered Digital Divide in India*

Indicator	Women	Men	Gap
Mobile phone ownership (relative likelihood)	Reference group	15% more likely than women	15 percentage points
Mobile internet use, national average	30%	50%	20 percentage points
Mobile internet use gender gap, Asia-Pacific comparison	—	—	40.4% (widest in region)
Household computer access, poorest 20% of households	2.7% (households)	2.7% (households)	Not gender-disaggregated at this level
Household internet access, poorest 20% of households	8.9% (households)	8.9% (households)	Not gender-disaggregated at this level

Source: Compiled by the author from Oxfam India (2022) and GSMA (2022).

The consequences of this divide extend beyond simple exclusion. Where AI-enabled services are positioned as the default or sole channel for accessing entitlements—as has increasingly been the case with direct benefit transfers—citizens without reliable access are effectively pushed toward intermediaries: cyber cafés, common service centres, or male relatives who mediate the transaction on their behalf. Each point of mediation introduces the possibility of fees, delay, error, or loss of agency, and existing research on the digitisation of the PDS finds that these intermediation costs fall disproportionately on women and other marginalised groups (Oxfam India, 2022). A system billed as a bureaucratic simplification for the citizen can, in practice, function as a simplification for the state and a new set of frictions for the citizen it was meant to serve.

Digital literacy compounds the raw connectivity gap. Ownership of a device, or even an active mobile internet connection, does not guarantee the skills required to navigate a government portal, interpret an AI-generated eligibility notification, or contest an automated decision. Survey evidence on internet usage patterns among Indian women finds that a large share of women who do own or share a smartphone report low confidence in using it independently for tasks beyond voice calls and messaging—finding necessary information, using mobile money, or navigating unfamiliar applications remain skills concentrated among a minority of women users (Oxfam India, 2022). This has direct implications for AI-enabled grievance redress mechanisms, which are frequently designed as chatbot or app-based interfaces on the assumption that a citizen denied a benefit will be able to independently identify the denial, understand its stated reason, and navigate a digital appeal process. For a woman with limited independent digital access and low confidence in navigating unfamiliar interfaces, each of these steps represents an additional point of possible failure layered on top of the original algorithmic error.

The digital divide is also not static or uniform across India's states and social groups; it interacts with caste, religious minority status, disability, and rurality in ways that compound disadvantage for women who sit at the intersection of multiple marginalised identities. Research on the gendered dimensions of India's service-led economic growth notes that women's mobile internet adoption has been particularly resistant to national-level policy interventions such as the Digital India initiative, in part because such initiatives have historically lacked explicit gender-disaggregated targets and have not addressed the social and mobility constraints—rather than purely infrastructural constraints—that limit women's independent technology use in many households (Oxfam India, 2022). This suggests that infrastructure investment alone, absent gender-specific programming, is unlikely to close the access gap that conditions all subsequent stages of AI-enabled service delivery.

5. Algorithmic Exclusion in Practice: Case Evidence from Welfare Delivery

Beyond access, the design of AI and automated decision systems themselves introduces gender-differentiated risk. The most extensively documented case in the Indian context is the use of Aadhaar-linked biometric authentication to verify eligibility for welfare schemes, including the PDS, MGNREGA, and various pension schemes. Biometric authentication—typically fingerprint or iris matching—has been found to fail disproportionately for specific groups: manual labourers and elderly citizens, whose fingerprints are often worn down by physical work, and women in particular occupational and age categories, experience elevated rates of authentication failure (Khera, 2019). Because authentication failure at a point-of-sale device is frequently treated administratively as a presumption of ineligibility rather than a technical fault, the consequence is often outright denial of an entitlement rather than a referral for manual verification.

Investigations into starvation deaths linked to PDS exclusion in the years following Aadhaar's mandatory linkage to ration cards found a recurring pattern in which biometric or connectivity failure, rather than genuine ineligibility, was the proximate cause of exclusion; the Right to Food Campaign documented that a majority of a sample of investigated starvation deaths in this period had an identifiable connection to Aadhaar-related administrative failure (as cited in Bansal & Agrawal, 2022). More recent field research on migrant women workers has found that a substantial minority—approximately 36% of a sample of migrant women interviewed—reported biometric authentication failures specifically during pregnancy-related hospital visits, a period when reliable access to maternal health entitlements is most acute (Bansal & Agrawal, 2022). These findings illustrate a broader dynamic identified in the algorithmic fairness literature: a system designed around an implicit "default citizen" whose biometric and digital profile is stable and easily machine-readable systematically disadvantages those, including many women in physically demanding informal work, whose lived circumstances deviate from that default (Sambasivan et al., 2021).

A second, related concern is the near-total absence of gender-disaggregated design in the datasets underlying many public-sector AI applications. Internet access itself correlates strongly with gender: research on algorithmic fairness in the Indian context notes that internet users skew male, with women's underrepresentation in usage data translating directly into underrepresentation in the datasets used to train, validate, and calibrate AI models intended for national-scale deployment (Sambasivan et al., 2021). A model calibrated predominantly on the behaviour and profiles of digitally privileged, often male, users will tend to overfit to that population, treating deviations from it—including the usage patterns, language preferences, and documentation gaps more common among rural and low-income women—as anomalies rather than as a substantial share of the population it is meant to serve.

Table 2- *Documented Mechanisms of Gendered Algorithmic Exclusion in Indian Welfare Delivery*

Mechanism	Affected Group / Context	Documented Consequence
Biometric (fingerprint/iris) authentication failure	Manual labourers, elderly citizens, women in informal work	Denial or delay of PDS rations, pensions, MGNREGA wages
Authentication failure during pregnancy/maternal health visits	Migrant women workers (approx. 36% of sampled cases)	Interrupted access to maternal health entitlements
Non-representative training data (male-skewed internet usage)	Rural and low-income women, non-users of internet	Model overfitting to privileged user profiles; degraded accuracy for underrepresented groups
Treatment of authentication failure as ineligibility rather than technical fault	All Aadhaar-linked scheme beneficiaries, disproportionately women and marginalised castes	Administrative exclusion without manual review; documented linkage to starvation deaths

Source: Compiled by the author from Khera (2019), Bansal and Agrawal (2022), and Sambasivan et al. (2021).

It is worth pausing on why these failures are gendered rather than merely random. Three mechanisms recur across the case evidence reviewed for this paper. The first is physiological and occupational: repetitive manual labour, more common among women in agricultural and construction work, is associated with wear patterns on fingerprints that reduce biometric matching accuracy, so that the same authentication threshold calibrated on an office-working population performs measurably worse for a population of women agricultural labourers (Khera, 2019). The second is administrative: welfare entitlements most closely tied to reproductive and maternal health—registration for the Pradhan Mantri Matru Vandana Yojana, for instance, or hospital-linked benefit disbursal around childbirth—require authentication at precisely the life stages when women are most likely to be travelling, staying with natal family, or otherwise located away from the biometric or address records on file, elevating failure rates at exactly the moments when reliable access matters most (Bansal & Agrawal, 2022). The third is representational: because women are underrepresented among both the designers of these systems and the population reflected in usage data used for testing and calibration, failure modes specific to women's patterns of use are less likely to be identified and corrected before deployment (Sambasivan et al., 2021).

These mechanisms interact with, rather than substitute for, the access barriers discussed in Section 4. A woman who successfully authenticates for a scheme in one cycle may nonetheless be excluded in a subsequent cycle due to a biometric mismatch following a change in occupation or location, and reporting that failure through the prescribed grievance channel typically requires exactly the kind of independent digital access and confidence that Section 4 shows to be least available to her. The result is a compounding, rather than additive, pattern of disadvantage: each stage of the AI-enabled service pipeline—access, authentication, and redress—carries its own gendered failure risk, and a woman must clear all three successfully to receive an entitlement that a formally identical, but socially advantaged, citizen may receive without friction.

6. Institutional and Policy Responses: Is Gender on the Governance Agenda?

India's AI governance architecture has evolved considerably since 2018, but gender equity remains a largely implicit rather than operationalised commitment within it. NITI Aayog's National Strategy for Artificial Intelligence identified privacy, security, ethics, fairness, transparency, and accountability as core issues for AI deployment, and proposed healthcare, agriculture, education, smart cities, and mobility as priority sectors for AI-for-good applications, but did not articulate specific safeguards against gender-differentiated harm (NITI Aayog, 2018). The subsequent Responsible AI approach papers advanced the framework considerably, establishing seven principles—safety and reliability, equality, inclusivity and non-discrimination, privacy and security, transparency, accountability, and the protection of positive human values—of which "equality, inclusivity and non-discrimination" is the most directly relevant to the concerns raised in this paper (NITI Aayog, 2021a). Part 2 of the approach document went further, recommending regulatory and policy interventions, sector-specific capacity building, and the incentivisation of "ethics by design" among private-sector developers building AI systems for public use (NITI Aayog, 2021b).

Nonetheless, three gaps are apparent when this framework is read against the evidence in Sections 4 and 5. First, the principle of equality and non-discrimination is stated at a high level of generality, without a corresponding requirement for gender-disaggregated impact assessment prior to the deployment of a public-sector AI system, of the kind increasingly recommended in comparative regulatory frameworks such as the European Union's approach to high-risk AI systems (NITI Aayog, 2021b). Second, existing grievance redress mechanisms for Aadhaar-linked authentication failures remain largely manual, under-resourced, and poorly publicised, placing the burden of proving a technical failure on the citizen rather than the system operator. Third, none of the major strategy documents reviewed for this paper set out a target or mechanism for improving the representation of women in the AI research, design, and policy workforce that shapes these systems, despite the well-established link between design-team diversity and the identification of blind spots in system behaviour (Noble, 2018; Sambasivan et al., 2021).

Table 3- *Gender-Relevant Provisions in India's National AI Governance Documents*

Document	Year	Gender-Relevant Provision	Assessed Gap
National Strategy for Artificial Intelligence	2018	General commitment to fairness and inclusive growth across five priority sectors	No gender-specific metrics, targets, or safeguards
Responsible AI Approach Document, Part 1	2021	"Equality, inclusivity and non-discrimination" listed as core principle	Principle not operationalised into a disaggregated impact-assessment requirement

Document	Year	Gender-Relevant Provision	Assessed Gap
Responsible AI Approach Document, Part 2	2021	Recommends regulatory capacity-building and "ethics by design" for developers	No specific mechanism for gender-diverse design teams or participatory audits

Source: Compiled by the author from NITI Aayog (2018, 2021a, 2021b).

It is worth noting that these gaps are not unique to India, nor are they evidence of an absence of good-faith effort. Comparative AI governance scholarship observes that most national AI strategies published in the years following 2018, across both high-income and middle-income countries, articulate fairness and non-discrimination as general principles while leaving detailed operationalisation to sector-specific regulation that frequently lags years behind the strategy document itself (NITI Aayog, 2021b). What distinguishes the Indian case is less the presence of a gap between principle and practice than the scale and centrality of the systems already in deployment: unlike many jurisdictions still in a pilot or advisory phase of public-sector AI adoption, India's Aadhaar-linked authentication infrastructure already mediates benefit delivery for a population larger than the European Union, meaning that any gap between stated principle and operational safeguard translates into exclusion at a scale with few global precedents.

There are, nonetheless, emerging institutional resources on which a more operationalised gender-responsive framework could draw. India's sectoral regulators—including bodies responsible for telecommunications, banking, and digital identity—already maintain grievance redress mechanisms of varying maturity, and several state governments have begun piloting human-assisted service delivery models, such as women-staffed common service centres, that could serve as a template for the kind of accessible, low-cost redress channel proposed in Section 6. Similarly, the existence of a designated national AI oversight function within NITI Aayog's proposed governance architecture offers a plausible institutional home for a gender-disaggregated impact-assessment requirement, should such a requirement be adopted as policy (NITI Aayog, 2021b). The task ahead is therefore less one of building new institutions from scratch than of directing existing institutional capacity toward a set of measurable, gender-specific commitments.

7. Towards Inclusive Design: A Gender-Responsive Framework

Addressing the gaps identified above requires moving from a rhetoric of inclusion to an operational framework embedded within the design and deployment lifecycle of public-sector AI. Drawing on the evidence reviewed in this paper, four components appear central to such a framework.

First, mandatory gender-disaggregated data collection and impact assessment prior to the deployment of any AI system that determines eligibility for, or delivery of, a public entitlement. Without disaggregated data on who is authenticated successfully, who is flagged for manual review, and who is excluded outright, gendered patterns of failure of the kind documented in Section 5 remain statistically invisible to the agencies operating these systems (Sambasivan et al., 2021).

Second, a statutory right to prompt, low-cost human review whenever an automated system denies or delays an entitlement, shifting the burden of proof away from the citizen. Given that authentication failure is currently treated as a presumption of ineligibility in much of the existing PDS and MGNREGA architecture, this single change would directly address the mechanism identified as most consequential in Section 5 (Khera, 2019).

Third, participatory design and audit processes that involve women, including rural and informal-sector women, as active participants rather than passive data subjects. Comparative research on AI design processes finds that involving affected communities directly in system testing surfaces failure modes—such as language, dialect, or documentation barriers—that are otherwise invisible to design teams operating at a national or metropolitan remove from the populations a system is meant to serve (Sambasivan et al., 2021).

Fourth, explicit workforce and governance targets for the representation of women in AI research, product, and policy roles connected to public-sector deployment, paired with the institutionalisation of gender expertise within the oversight body that NITI Aayog's National Strategy proposes to establish (NITI Aayog, 2018, 2021b). None of these four components requires a wholesale redesign of India's existing AI governance architecture; each is compatible with, and could be incorporated into, the principles already articulated in the Responsible AI approach documents. What is currently missing is not the vocabulary of inclusion but the operational mechanisms that would make it enforceable.

Implementing these four components would also require attention to sequencing and cost. Gender-disaggregated impact assessment is the least costly of the four to introduce administratively, since it primarily requires a change in what data are collected and reported rather than a redesign of the underlying system, and could plausibly be mandated for all new public-sector AI deployments within a single budget cycle. A statutory right to human review carries a larger recurring administrative cost, since it requires trained personnel and accessible physical or telephonic channels, particularly in rural areas where digital-only redress would replicate the very access barriers the mechanism is meant to correct; phased implementation, beginning with the highest-volume schemes such as the PDS and MGNREGA, offers one route to managing this cost. Participatory design and audit processes require the least financial outlay but the most institutional change, since they depend on government agencies and their private-sector vendors treating community consultation as a required stage of system development rather than an optional add-on. Workforce and governance targets are, in a sense, a long-horizon investment: their benefits accrue gradually as a more diverse cohort of designers and policymakers shapes the next generation of systems, but they are also the component most likely to produce durable, self-reinforcing improvements in how blind spots are identified before deployment rather than after harm has occurred.

Taken together, these four components describe a shift from a compliance-oriented conception of fairness—in which a system is deemed acceptable if it does not explicitly encode a prohibited category such as gender—to a design-oriented conception, in which fairness is treated as an outcome that must be actively engineered through disaggregated measurement, accessible redress, participatory testing, and diverse teams. The evidence reviewed in Sections 4 and 5 suggests that the former, compliance-oriented conception is largely already satisfied by India's existing AI governance documents, which contain no provision that explicitly discriminates by gender; it is precisely because formal compliance of this kind coexists with the substantive exclusions documented in this paper that a design-oriented framework is necessary.

8. Discussion: Limitations and Implications

Several limitations of this analysis warrant acknowledgment. First, the paper relies on secondary data and documented case evidence rather than original field research; while the sources drawn upon—national inequality reports, international mobile-usage surveys, and peer-reviewed algorithmic fairness scholarship—are methodologically robust in their own right, they were not designed specifically to answer the question posed here, and the synthesis in this paper should be read as an evidentiary mapping rather than a primary empirical study. Second, the available data are unevenly disaggregated: national and international surveys typically report gender gaps at an aggregate level, and the paper's ability to speak to intersections with caste, religion, disability, and region is correspondingly constrained by what existing published sources report. Third, India's states vary considerably in administrative capacity, scheme design, and the maturity of grievance redress infrastructure, and a national-level synthesis of the kind offered here necessarily understates this variation; a Kerala or Tamil Nadu, with comparatively strong public health and social security administration, is unlikely to exhibit the same failure rates as states with weaker administrative capacity, even where the underlying AI system is nationally standardised.

These limitations point directly to priorities for future research. Sub-national comparative studies that examine how state-level administrative capacity mediates the gendered effects of nationally standardised AI systems would help clarify which failures are inherent to the technology and which are a function of local implementation quality. Field-based audits that trace individual beneficiaries through the full access-authentication-redress pipeline, rather than measuring each stage in isolation, would provide more direct evidence on the compounding dynamic identified in Section 5. Finally, comparative research examining whether the gender-responsive design principles proposed in Section 6 have been successfully implemented in other large digital-identity or welfare-delivery contexts—including, for instance, biometric systems in Kenya or Indonesia—could help identify which interventions are most cost-effective in reducing gendered exclusion at scale.

Notwithstanding these limitations, the weight of the evidence assembled across Sections 4 through 6 supports a consistent conclusion: gender inequity in AI-enabled public service delivery in India is not an incidental or transitional problem that will resolve itself as digital infrastructure matures, but a structural feature of how these systems have so far been designed, tested, and governed. Addressing it will require the kind of deliberate, resourced, and accountable intervention outlined in Section 6, rather than reliance on the general expansion of connectivity or the general articulation of fairness principles.

9. Conclusion

This paper set out to examine whether India's turn toward AI-enabled public service delivery is inclusive by design or only inclusive by rhetoric. The evidence reviewed suggests the latter remains closer to the current reality. A substantial and gendered digital divide continues to determine who can access AI-mediated services at all; documented cases of biometric authentication failure show that even successful access does not guarantee accurate or fair treatment by the systems citizens do reach; and India's national AI governance framework, while articulating strong principles of equality and non-discrimination, has yet to translate those principles into disaggregated impact assessment, accessible grievance redress, or workforce diversity requirements specific to gender.

None of this implies that AI has no role to play in improving public service delivery in India; used well, and audited rigorously, automated systems can reduce leakage, cut processing times, and extend administrative reach into underserved areas. The argument of this paper is narrower and more constructive: that these benefits will accrue unevenly, and in some documented cases have already produced serious harm, unless gender equity is built into the design, deployment, and oversight of these systems as a first-order requirement rather than an aspirational principle. Future research should prioritise primary, field-based audits of specific AI-enabled schemes—disaggregating outcomes by gender, caste, and location—to test empirically whether the mechanisms proposed in Section 7 measurably reduce the exclusion documented in this paper. As India continues to scale its AI ambitions toward the stated goal of "AI for All," the test of that ambition will lie less in the sophistication of the technology deployed than in whether the women currently least visible to these systems become visible within them.

For policymakers, the practical implication is that gender equity should be treated as a design specification to be engineered and measured from the outset of a system's development, in the same way that uptime, security, and processing speed are already treated as non-negotiable specifications for any public-facing digital system. For researchers, the implication is a need for sustained, disaggregated measurement of AI system performance across gender, caste, and geography, published with the same regularity and institutional weight as the connectivity and adoption statistics already tracked by bodies such as GSMA and Oxfam India. And for the broader project of "AI for All" articulated in India's National Strategy, the implication is that inclusion cannot be assumed as a by-product of scale; it must be a deliberate, resourced, and continually audited commitment, built into public-sector AI systems from their first line of code to their final grievance redress channel.

REFERENCES:

1. Bansal, P., & Agrawal, R. (2022). India's digital divide and the consequent welfare bias. Policy Circle. https://www.policycircle.org/opinion/digital-divide-public-services/
2. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
3. GSMA. (2022). The mobile gender gap report 2022. GSMA Connected Women Programme. https://www.gsma.com/mobilefordevelopment/programme/connected-women/the-mobile-gender-gap-report-2022/
4. Khera, R. (2019). Dissent on Aadhaar: Big data meets big brother. Orient BlackSwan.
5. NITI Aayog. (2018). National strategy for artificial intelligence: #AIforAll. Government of India. https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf
6. NITI Aayog. (2021a). Responsible AI: Approach document for India, part 1—Principles for responsible AI. Government of India. https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf
7. NITI Aayog. (2021b). Responsible AI: Approach document for India, part 2—Operationalizing principles for responsible AI. Government of India.
8. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.
9. Oxfam India. (2022). India inequality report 2022: Digital divide. Oxfam India.
10. Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021). Re-imagining algorithmic fairness in India and beyond. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 315-328). Association for Computing Machinery. https://arxiv.org/pdf/2101.09995