

Role of Artificial Intelligence in Governance: A Citizen-Centric Framework for Inclusive AI Governance in India

Author 1 - Abhay Thakur

MTech Scholar , Department of Computer Science and engineering , UIT , HPU , Shimla , India

Author 2 - Akshay Bhardwaj

HOD , Department of Information Technology , UIT , HPU , Shimla , India

Author 3 - Rajesh Chauhan

System Administrator , Department of Information Technology , UIT , HPU . Shimla , India

Abstract

Artificial Intelligence (AI) is now a core part of how public services are delivered in India. The clearest example is the Digital Public Infrastructure (DPI) stack: Aadhaar-based fingerprint or iris verification, Direct Benefit Transfer (DBT) of subsidies straight into bank accounts, and AI-based checks that decide who is eligible for a welfare scheme. For the majority of individuals, this method has improved the speed and transparency of government services. However, it has also led to a significant issue: if an individual fails an automatic identity check, the system typically considers them to be simply ineligible, and there is no way to correct the error. The main barrier to equitable and inclusive AI governance in India today, according to this research, is that very issue people being wrongfully excluded from benefits because an AI-based check fails, with no simple fallback. Using published research, government reports, court records, and real documented cases, the paper builds a solution called the Inclusive AI Governance (IAG) framework. It consists of four components: a voice-based, multi-language complaint system for individuals who are unable to read or use online forms; a straightforward score called the Digital Vulnerability Index (DVI) that indicates which citizens are more likely to be incorrectly excluded; a step-by-step fallback process that sends high-risk cases to a human reviewer instead of an automatic rejection; and a transparency and accountability layer that allows failures to be tracked and appealed. The paper then walks through real reported cases from Aadhaar-linked welfare delivery to show how this framework would have prevented wrongful denial. It also honestly discusses the framework's limitations, including the privacy risks of scoring citizens, and lays out a plan for testing it in the real world. Overall, this paper offers a practical, low-cost design that governments can use to make AI-based welfare systems fair by default instead of exclusionary by default.

Keywords: Artificial Intelligence, Digital Governance, Inclusive AI Governance, Algorithmic Exclusion, Digital Vulnerability Index, Aadhaar, Citizen-Centric Design, Public Service Delivery, e-Governance, India

1. Introduction

Artificial intelligence is being used by governments worldwide to automate choices in areas such as benefit targeting, tax collecting, identification verification, and complaint processing. One of the most notable instances of this is India. More than a billion individuals are covered by India's biometric identity system, Aadhaar, and access to food subsidies, pensions, employment guarantee pay, and even bank accounts now requires fingerprint or OTP-based authentication [2], [7]. AI is rapidly being utilised behind this system to match records, assess eligibility, and detect fraud—often without any human reviewing the decision [3], [4].

There are two main categories of research on this subject. One set of studies focuses on the advantages of AI, including improved service delivery, reduced corruption, and quicker complaint resolution [3], [4], and [6]. The other group concentrates on the harm: individuals who are most in need of welfare, such as elderly citizens, manual labourers, people with disabilities, and women working informal jobs, are also the most likely to fail a fingerprint or biometric check because their fingerprints change with age, hard physical labour, or illness [8], [9], and [10]. A system's design is flawed when it ends up excluding people when it was intended to include them.

This essay makes no claims about how much or how little AI should be used in India. Rather, it makes the case that governance systems require a citizen-centric safety layer that ensures AI judgements are equitable, reversible, and traceable in the event that they go wrong. In order to do this, the paper bases the Inclusive AI Governance (IAG) framework on a single, frequently disregarded issue: citizens now lack a suitable backup procedure in the event that an AI-based identity or eligibility check is unsuccessful.

2. Literature Review

2.1 AI and e-Governance in India

The majority of current studies concur that the use of AI in Indian government services is expanding rapidly due to national initiatives like the IndiaAI Mission and the widespread usage of Aadhaar, UPI, and DBT together [2], [7]. AI adoption clearly improved citizen-facing services and internal processes, according to a study of Chief Information Officers in Indian municipal bodies . However, only a small percentage of the organisations under study had the technical infrastructure necessary to support this, and improvements for citizens themselves were less significant than improvements in raw efficiency [4] . Broader review studies on AI and IoT in e-government tell a similar story: they focus mainly on efficiency, transparency, and accountability, while treating fair access as a secondary issue that is rarely solved in practice [6].

2.2 Citizen-Centric Platform Design

A second set of studies focuses on the design of these digital platforms. Single AI-assisted platforms that include eligibility checks, document verification, and complaint management are replacing department-by-department websites, according to research on citizen-centric digital projects throughout Indian states [1], [3]. One such platform, which was designed with multilingual support for individuals with low literacy, reportedly reduced the average complaint resolution time from fourteen days on previous government portals to five days [3]. These studies demonstrate that the field already intends to design for citizens, but they primarily gauge success by speed and integration rather than justice or the number of persons who are mistakenly excluded.

2.3 Algorithmic and Biometric Exclusion

Exclusion brought about directly by AI-based identity checks is documented in a third and, for the

purposes of this research, more significant set of studies. In her field research on Aadhaar-linked welfare, Reetika Khera tracks recurrent failure types, such as rejection during enrolment, failed authentication when attempting to collect a benefit, mismatched records, and an exception-handling process that is supposed to exist but frequently does not work in practice on the ground. She connects these failures to specific documented cases of denied pensions, denied rations, and, in the worst cases, starvation deaths [8]. The extent of this issue is demonstrated by independent analysis: out of approximately 312 million monthly Aadhaar biometric authentication attempts, about 20.3 million (6.5 percent) fail. This rate has remained nearly constant for more than ten years and is not random; it is concentrated among manual labourers whose fingerprints deteriorate from years of physical labour [10]. According to a working paper from ISEC, elderly individuals, those with disabilities, and those with low literacy rates bear a disproportionate share of these authentication failures, which impacts their independence and dignity in addition to the benefits they lose [7].

2.4 Research Gap

Put together, the research clearly shows two things: AI makes Indian governance more efficient overall, and it also causes exclusion that falls unevenly on the same vulnerable groups every time. What is missing is a practical, usable solution—something a real government system could actually plug in at the moment a check fails, rather than a general principle. Existing work either recommends broad ideas like "more transparency" or "human oversight" without explaining exactly when or how that oversight should be triggered [3], [6], or documents the harm in detail without turning it into something that can be built [8], [9], [10]. This paper is written to fill that specific gap.

3. Problem Statement

The majority of AI-based public service systems in India operate on a simplistic yes-or-no logic: either a person's fingerprint or other details match the record, or they don't. If they don't match, the system often handles this as a simple rejection with no appropriate backup. This means every failure is treated the same way, no matter what caused it or how vulnerable the person is. In practice, this means a healthy young adult whose finger is briefly unreadable because the scanner is dirty is treated exactly the same as an 82-year-old whose hands have permanently changed with age, or a construction worker whose fingerprints have worn down after decades of manual labour [8], [10]. The real-world results of this design include denied pensions and rations, delayed maternity and newborn care, and, in the most serious and repeatedly documented cases, deaths linked directly to failed authentication [8]. The specific, unique problem this paper focuses on is this: Indian public service systems have no graded, fair, and traceable backup process for when an AI-based identity or eligibility check fails — and because of this, the citizens who most need government support are also the ones most likely to be wrongly denied it.

4. Research Objectives

1. To investigate the current application of AI in Indian public service delivery and pinpoint the precise point at which exclusion occurs.
2. To gather and compile verifiable proof of AI-driven exclusion in India's welfare system, as well as to determine which demographics are most impacted.
3. To create the Inclusive AI Governance (IAG) framework, a workable, citizen-focused framework that incorporates a just backup procedure in the event that an AI-based check fails.

4. To evaluate the suggested framework on paper using actual documented exclusion scenarios and contrast it with the way the existing system responds to the same situations.
5. To determine the ethical, technical, and practical obstacles to using this framework in practice and to propose a strategy for future testing.

5. Research Methodology

This study employs a design-based, conceptual research strategy, which is appropriate for a study that suggests a new framework as opposed to evaluating an existing one. Three steps were taken to complete the task. First, published research and reports on AI in Indian governance, citizen-centric digital platforms, and Aadhaar or biometric authentication failures were analysed and arranged alongside publicly available government and legal materials, such as UIDAI circulars on managing biometric exceptions, CAG audit findings, and reported Supreme Court proceedings on Aadhaar-linked welfare denials. Second, the common failure patterns found in this material were turned into a small set of simple risk factors, which form the basis of the Digital Vulnerability Index described in Section 6. Third, the framework's behaviour is explained by walking through previously reported real exclusion cases [8], [10], and comparing what would happen under the current system versus what would happen under the proposed IAG framework.

It is crucial to understand the limitations of this approach: no fresh information was gathered directly from citizens for this study. Instead of being based on statistical analysis of actual data, the weights used in the Digital Vulnerability Index in Section 6.2 are illustrative, based on how frequently each risk factor appears in the examined research. Testing and adjusting these weights using real survey or government data is described as essential future work in Section 9. This paper should be read as a proposed design that is ready for review and discussion, not as a system that has already been tested and proven in the field.

6. Proposed Solution: The Inclusive AI Governance (IAG) Framework

The IAG framework is built around one simple idea: when an AI-based system denies a citizen access to a benefit, that denial should never be the final, unreviewable answer just because one automated check failed. The framework puts this idea into practice through four connected parts.

6.1 Design Principles

- Do not exclude by default: A failed automated check should lead to a proper backup process, not an automatic rejection.
- Match the response to the risk: The kind of backup used—simple retry, assisted retry, or human review—should depend on how vulnerable the person is, not be the same for everyone.
- Keep a human in the loop for risky cases: No citizen who is flagged as high-risk should be denied a benefit without a real person reviewing the case.
- Keep a clear record: Every automated decision and every backup step should be logged with a clear reason, so it can be checked and audited later.
- Make it accessible: Every step a citizen has to go through should be available by voice, in local

Indian languages, for people with low literacy or limited reading ability.

6.2 The Digital Vulnerability Index (DVI)

The DVI is a simple score, calculated only at the moment an authentication check fails (not stored permanently, to avoid creating an ongoing surveillance record), using information the welfare system already has on file. The DVI is calculated as:

$$\text{DVI}(i) = w1.A(i) + w2.O(i) + w3.L(i) + w4.C(i) + w5.D(i)$$

Here, A, O, L, C, and D are scores between 0 and 1 for age, occupation-related biometric wear, language or literacy access, connectivity or location, and disability status. The letters w1 to w5 are weights that add up to 1, showing how much each factor matters. Table 1 explains each part and why it was chosen, based on the research reviewed earlier.

Component	Illustrative Weight	Why It Matters	Evidence Basis
A - Age	0.25	Fingerprints become harder to read as people get older, so elderly citizens fail authentication more often.	[7], [8]
O - Occupation	0.25	Manual, farm, and construction work wears down fingerprint ridges over time.	[10]
L - Language/Literacy	0.20	People with low literacy or who don't use the dominant language struggle with text-based OTP or online forms.	[3], [5]
C - Connectivity	0.15	People in rural or low-network areas cannot rely on mobile OTP as a backup.	[5], [6]
D - Disability	0.15	Physical or visual disability directly makes it harder to give a fingerprint scan or use an online portal.	[7], [8]

Table 1. Digital Vulnerability Index (DVI) components (illustrative weights, to be tested and adjusted with real data later).

A citizen with a low score (below 0.35, as an example) is sent to a simple automatic retry. A citizen with a medium score (0.35 to 0.65) is sent to an assisted retry with help from a field officer. A citizen with a high score (above 0.65) is sent straight to a human reviewer, and importantly, the benefit is given to them right away while the review happens, instead of being held back. This is the opposite of how the system usually works today, where the benefit is withheld until the problem is resolved.

6.2.1 Weight Normalization and Formal Risk-Tier Function

For the DVI to stay bounded in [0, 1] and behave as an interpretable risk score, the weights must satisfy a normalization constraint:

$$\sum_{k=1 \text{ to } 5} w_k = 1, w_k \geq 0 \text{ for all } k$$

This holds for the illustrative weights in Table 1 ($0.25 + 0.25 + 0.20 + 0.15 + 0.15 = 1.00$). Given this constraint, $\text{DVI}(i)$ is a convex combination of the five component scores, so $\text{DVI}(i) \in [0, 1]$ whenever

each of $A(i)$, $O(i)$, $L(i)$, $C(i)$, $D(i) \in [0, 1]$. The routing rule described qualitatively above can then be written as a formal risk-tier function:

$R(i) = \text{Retry}$, if $DVI(i) < 0.35$

$R(i) = \text{Assisted}$, if $0.35 \leq DVI(i) \leq 0.65$

$R(i) = \text{HumanReview}$, if $DVI(i) > 0.65$

As a worked example, consider the elderly citizen discussed in Section 7, modelled with illustrative component scores $A = 0.95$ (advanced age), $O = 0.70$ (history of manual work), $L = 0.60$ (limited literacy), $C = 0.80$ (rural, low connectivity), and $D = 0.20$ (no registered disability). Applying the weights from Table 1:

$$DVI = (0.25)(0.95) + (0.25)(0.70) + (0.20)(0.60) + (0.15)(0.80) + (0.15)(0.20)$$

$$DVI = 0.2375 + 0.175 + 0.120 + 0.120 + 0.030 = 0.6825$$

Since $0.6825 > 0.65$, $R(i) = \text{HumanReview}$: the citizen is routed straight to a human reviewer and receives her benefit immediately, matching the outcome described qualitatively in Section 7.

6.3 Adaptive Multi-Modal Authentication State Machine

The second part of the framework is a step-by-step process that decides what happens after an authentication attempt fails, replacing today's simple "retry, then reject" approach. This process is shown below in a simplified, code-like format.

STATE MACHINE: AdaptiveAuthentication

START -> TRY_FINGERPRINT

TRY_FINGERPRINT:

result = check_fingerprint(citizen_id)

IF result == MATCH:

-> ACCESS GRANTED

ELSE:

attempts += 1

IF attempts < 2:

-> TRY_FINGERPRINT # try again, same device

ELSE:

-> CALCULATE_DVI

CALCULATE_DVI:

score = $w_1 \cdot \text{Age} + w_2 \cdot \text{Occupation} + w_3 \cdot \text{Language} + w_4 \cdot \text{Connectivity} + w_5 \cdot \text{Disability}$ IF score < 0.35:

-> TRY_OTP # low risk: simple backup ELSE IF score <= 0.65:

-> ASSISTED_CHECK # medium risk: help from field officer ELSE:

-> GIVE BENEFIT NOW + SEND TO HUMAN REVIEWER # high risk: don't withhold

TRY_OTP:

IF otp_correct: -> ACCESS GRANTED

ELSE: -> SEND TO HUMAN REVIEWER

ASSISTED_CHECK:

IF field_officer_confirms (photo + local ID + witness): -> ACCESS GRANTED ELSE: -> SEND TO HUMAN REVIEWER

HUMAN REVIEWER:

reviewer looks at the case and the reason it failed

record the decision and reason, with date and time # always logged IF reviewer says yes: -> ACCESS GRANTED

ELSE: -> DENIED, WITH RIGHT TO APPEAL

DENIED, WITH RIGHT TO APPEAL:

citizen can appeal to a district officer within 30 days

-> HUMAN REVIEWER (case goes to a second reviewer)

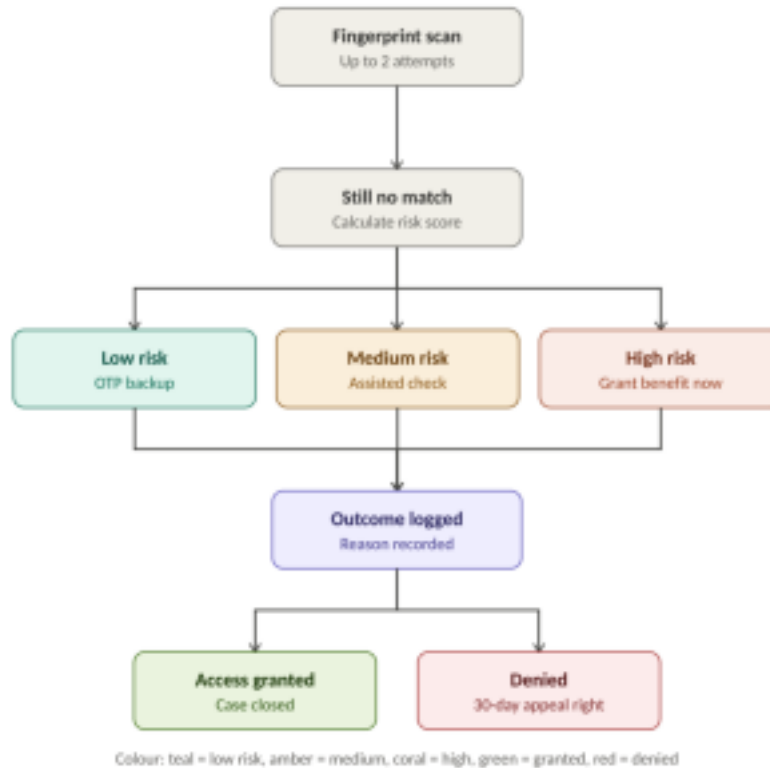


Figure 1. Adaptive multi-modal authentication state machine, shown as a flowchart.

The two biggest changes from how things work today are: high-risk citizens get their benefit right away instead of waiting for a review, and every single step is written down with a clear reason, which feeds into the transparency system explained in Section 6.5.

6.3.1 Estimated Human-Review Workload

A practical question for any state adopting the IAG framework is how many cases the human-review tier will actually generate. If N is the number of monthly authentication failures that reach CALCULATE_DVI, and pH is the share of those failures with $DVI(i) > 0.65$, the expected monthly human review load L is:

$$L = N \times pH$$

Using the national failure volume reported for Aadhaar-linked authentication, $N \approx 20.3$ million per month [10]. The evidence reviewed in Section 2.3 suggests that clearly high-risk failures — elderly citizens, people with disabilities, and severe cases of occupational biometric wear — are a minority of total failures rather than the majority [7], [10]. Taking an illustrative $pH = 0.10$ (to be replaced with a measured value once pilot data from Section 9 is available) gives:

$$L \approx 20,300,000 \times 0.10 = 2,030,000 \text{ cases per month, nationally}$$

Spread across India's roughly 766 districts, this works out to about 2,650 cases per district per month, or on the order of 100 cases per working day — a caseload a small, trained grievance-officer team per district (Section 6.5) could realistically absorb, though pH itself needs to be measured, not assumed, during the pilot described in Section 9.

6.4 Multilingual, Low-Literacy Grievance Redressal Pipeline

The framework includes a voice-first complaint system that is available in major Indian languages and uses a combination of an automated speech system and a human call-back option because many impacted persons find text-based complaint websites difficult to use. It operates in three straightforward steps: first, a citizen can check the status of their case or schedule another attempt via an automated voice line; second, if the case is not resolved within 48 hours, it is forwarded to a human operator who speaks the local language; and third, if it is still unresolved after seven days, it is forwarded to a district-level officer with the citizen's DVI score and complete case history automatically attached. This builds on the multi-language support already reported in citizen centric welfare platforms [3], but applies it specifically to fixing exclusion cases, not just general service use.

6.5 Institutional Accountability Layer

The framework requires three things that are often talked about but rarely put into practice: first, a public dashboard showing, by district, how many people at each risk level failed authentication and how their cases were resolved, addressing the fact that UIDAI does not currently publish this kind of data by age or demographic group [8]; second, an independent grievance officer role, kept separate from both the technology company and the department giving out benefits, who has the power to overturn an automatic rejection; and third, a guaranteed 30-day right to appeal on every denial, which is logged and can be checked independently, addressing the accountability gap that researchers have pointed out in Aadhaar-linked digital ID systems [7].

7. Results and Discussion

Since the IAG framework has not yet been tested in a real pilot, this section offers a conceptual evaluation: it walks through documented exclusion cases from the research reviewed earlier, and compares the framework to how the current system handles the same situations.

Aspect	Current System (Yes/No Authentication)	Proposed IAG Framework
What happens when a check fails	Retry, then denial; benefit is held back until resolved.	Case is scored by risk (DVI); high-risk cases get the benefit right away, while the case is reviewed.
Human review	Available in theory, but not consistent in practice.	Required by rule for every high-risk case.
Public data on failures	Not regularly published by age or demographic group.	Public dashboard showing failure and resolution rates by risk level and district.
Access for low-literacy users	Mostly text-based forms and online portals.	Voice-first system in local languages, with staged human escalation.
Right to appeal	Inconsistent, with no fixed timeline in most cases.	Guaranteed 30-day appeal, logged and independently checkable.

Table 2. Comparing the current system with the proposed IAG framework.

The framework's potential benefits are demonstrated by applying it to actual reported cases. Take the example of an elderly Aadhaar holder whose hands altered with age and her fingerprints ceased matching. As a result, she was cut off from support till she passed away. Under the IAG framework, her age alone (an elderly citizen without a registered mobile number) would have pushed her DVI score above the high-risk mark on the very first failed attempt, meaning she would have received her benefit immediately and been sent to a human reviewer, instead of facing repeated, unresolved denial [8]. In another reported case, one broken fingerprint scanner at a ration shop caused around 400 residents of a village to be denied their food rations. Under the IAG framework, this pattern — many people failing at the same time, from the same device — would be automatically flagged in the log as a device fault rather than treated as hundreds of separate individual failures, and would trigger an automatic escalation to the district officer instead of taking months of manual follow-up. Looking at the bigger picture, with about 20.3 million monthly authentication failures across India [10], the framework's risk-based approach means only the cases involving genuinely vulnerable citizens go to human review, keeping the workload manageable while still protecting the people most likely to be wrongly denied.

This qualitative comparison can be stated as a formal exclusion-reduction bound. Let F_{current} denote citizens wrongly denied under today's yes/no system, and F_{IAG} denote citizens wrongly denied under the IAG framework. Because every $DVI(i) > 0.65$ case receives the benefit immediately and is withdrawn only if a human reviewer later confirms genuine ineligibility (Section 6.3), the wrongful denial count under IAG is bounded by the reviewer error rate rather than by the raw authentication failure rate:

$$F_{\text{current}} \approx p_H \times N \text{ (today: every high-risk failure is denied outright)}$$

$$F_{\text{IAG}} \leq p_H \times N \times e_{\text{reviewer}} \text{ (under IAG: denied only if the reviewer also upholds it)}$$

Here e_{reviewer} is the human reviewer's own error rate, expected to be well below the roughly 6.5 percent automated failure rate [10], since a reviewer can weigh evidence — photographs, local identification, witness statements — that an automated check cannot. Even a conservative $e_{\text{reviewer}} = 0.20$ (a reviewer wrongly upholding a rejection one time in five) implies an 80 percent reduction in wrongful denials among the highest-risk group relative to the current system. This figure is illustrative, not measured, and is exactly the quantity the pilot in Section 9 is designed to estimate.

Two main points come out of this discussion. First, the exclusion problem is not really a technology accuracy problem — UIDAI's own reported lab accuracy for fingerprint and iris matching is a high 98 to 99 percent [10] — it is a governance design problem: nothing is currently in place to properly handle the small remaining 1 to 6.5 percent of cases that fail. Second, adding a simple, well-recorded risk score into the existing authentication process is a relatively low-cost step compared to the money already spent building AI-based authentication itself, since it does not need any new hardware and simply reuses information the welfare system already has.

8. Implementation Challenges

- Privacy risk: The DVI itself generates a new piece of information about a citizen that, if not strictly regulated, could be abused for surveillance or unfair targeting; the framework should adhere to the protections in the Digital Personal Data Protection (DPDP) Act, with the score being computed momentarily and not kept indefinitely.
- Choosing the proper weights: Before being used, the sample weights used in Section 6.2 must be tested and modified using actual survey or government data; otherwise, new injustice may unintentionally be created.
- Staffing and budget: Every district must hire qualified grievance officers in order to handle high-risk complaints, which is expensive, particularly in places with little funding.
- Risk of misuse: Giving benefits to high-risk situations immediately before review could theoretically be abused by some individuals; the framework attempts to mitigate this by requiring record-keeping and after-the-fact review rather than preventing access up front, but further research is needed to fully understand this trade-off.
- Political will: As suggested in Section 6.5, publishing a public dashboard that breaks down failures by demographic group necessitates a degree of transparency that has historically been lacking [8].
- Language support costs: Maintaining a voice-based system in multiple Indian languages requires ongoing funding, not just a one-time setup fee.

9. Future Scope

In order to compare exclusion rates with the current system, a small, controlled pilot of the IAG framework will be conducted in a small number of ration distribution or pension-payment locations. This will be done using the same type of randomised study approach that has previously been employed in Aadhaar welfare research [9]. The data from this pilot would allow for precise testing and modification of the DVI weights using real statistics instead of approximations. A machine-learning layer that automatically differentiates between a single person's biometric failure and a malfunctioning equipment causing several failures at once (like the ration-shop example) could be incorporated into future studies. It could also consider integrating this framework with the government's larger IndiaAI Mission and its Responsible AI guidelines so that comparable safeguards could be applied to other AI-driven public services, such as AI-based healthcare triage or automatic tax scrutiny.

9.1 Sample-Size Calculation for the Proposed Pilot

For the pilot in a limited number of ration-distribution or pension-payment areas to reliably detect a real difference in wrongful-exclusion rates between the current system and the IAG framework, it needs a large enough sample in each arm. Using the standard two-proportion power calculation for a randomised comparison, in line with the methodology used in past Aadhaar welfare research [9], the required sample size per arm is:

$$n = [z(\alpha/2)\sqrt{2\bar{p}(1-\bar{p})} + z(\beta)\sqrt{p_1(1-p_1) + p_2(1-p_2)}]^2 / (p_1 - p_2)^2$$

Here p_1 is the current wrongful-exclusion rate among failed authentications, p_2 is the target rate under IAG, $\bar{p} = (p_1 + p_2)/2$, and $z(\alpha/2)$ and $z(\beta)$ are the standard normal critical values for the chosen significance level and statistical power. Taking illustrative values consistent with the evidence in Section 2.3 — $p_1 = 0.065$ as an upper-bound proxy for the current system, using the reported 6.5 percent overall authentication failure rate [10]; $p_2 = 0.02$ as an illustrative target under IAG; a two sided significance level $\alpha = 0.05$ ($z(\alpha/2) = 1.96$); and 80 percent power ($z(\beta) = 0.84$) — gives:

$$n \approx [1.96\sqrt{2 \times 0.0425 \times 0.9575} + 0.84\sqrt{(0.065 \times 0.935 + 0.02 \times 0.98)}]^2 / (0.045)^2 \approx 315 \text{ cases per arm}$$

That is roughly 630 failed-authentication cases in total – a scale consistent with the small, controlled pilot proposed above, and achievable within a limited number of ration-distribution or pension payment areas over a single evaluation cycle. As with the DVI weights in Section 6.2, p_1 and p_2 here are illustrative planning inputs, not measured results, and should be revisited once baseline exclusion data from the pilot sites is available.

10. Conclusion

AI is now a permanent part of how public services work in India, and its benefits for efficiency are well documented. This paper has argued that the real unanswered question is not whether AI should be used in governance, but what happens to a citizen the moment an AI-based decision goes wrong. The Inclusive AI Governance (IAG) framework proposed here – built around a Digital Vulnerability Index, a step-by-step backup authentication process, a multi-language complaint system, and a public accountability layer – offers a practical, low-cost, and traceable way to solve this problem, based directly on real, documented failures in Aadhaar-linked welfare delivery. This paper presents the framework as a proposal ready for academic review, not as a finished, tested system, and its main contribution is turning the widely repeated idea of "citizen-centric AI governance with a human in the loop" into something a real government system could actually build, use, and check.

References

- [1] "Transforming Governance through Citizen-Centric Digital Initiatives: A Theoretical and Exploratory Analysis of Innovations in Public Service Delivery in Jammu and Kashmir (UT), India," ResearchGate, 2025.
- [2] "Artificial Intelligence in E-Governance in India," International Journal of Creative Research Thoughts (IJCRT), vol. 12, 2024.
- [3] "AI-Enabled Citizen-Centric Platform for Integrating Government and Private Welfare Schemes," International Journal of Environmental Sciences, vol. 11, no. 23s, 2025.
- [4] "Enhancing Public Service Delivery Efficiency: Exploring the Impact of AI," Digital Government: Research and Practice, ScienceDirect, 2024.
- [5] "Digital Governance in India: Opportunities and Challenges for Effective Public Service Delivery," Journal of Emerging Technologies and Novel Research (JETNR), 2026.
- [6] "Elevating E-Government: Unleashing the Power of AI and IoT for Enhanced Public Services," PMC (National Library of Medicine), 2024.
- [7] Grace, "Biometric Governance and Human Rights: The Social and Ethical Implications of Aadhaar-Based Exclusion," Institute for Social and Economic Change (ISEC) Working Paper No. 610, 2026.
- [8] R. Khera, "Aadhaar Failures: A Tragedy of Errors," Economic and Political Weekly, vol. 54, no. 14, Apr. 2019.

[9] NCAER, "Integrating Biometric Authentication in India's Welfare Programmes," India Policy Forum Working Paper, National Council of Applied Economic Research, 2021.

[10] "Aadhaar Authentication Failures Trigger an Invisible Exclusion Crisis," Policy Circle, 2026.