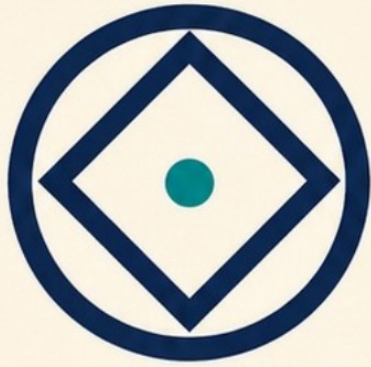




Technology Stewardship

INSTITUTE • GOVERNING AUTONOMOUS TECHNOLOGY

INSTITUTE FOR TECHNOLOGY STEWARDSHIP

RESEARCH PAPER No. 2

From Promise to Protection

The Institutional Reach of AI Governance
Oversight, Contestability and Redress for
People Affected by AI Systems

July 2026

Gaurav Arora

Founder, Institute for Technology Stewardship



*Ensuring technology remains in service of humanity —
not the other way around.*

RESEARCH PAPER

From Promise to Protection

The Institutional Reach of AI Governance

Oversight, Contestability and Redress for People Affected by AI Systems

AUTHOR	Gaurav Arora
INSTITUTION	Institute for Technology Stewardship
VERSION	1.4 - SECOND PEER-REVIEW REVISION
RESEARCH CUT-OFF	18 July 2026
DOCUMENT TYPE	Independent research paper
LICENCE	CC BY-NC 4.0
SUGGESTED CITATION	Arora, G. (2026). From Promise to Protection: The Institutional Reach of AI Governance - Oversight, Contestability and Redress for People Affected by AI Systems. Institute for Technology Stewardship, Research Paper No. 2, Version 1.5.

CENTRAL FINDING

Protection Reach is a weakest-link institutional chain: notice, coverage, evidence, contestation, authoritative review and timely remedy must connect for an affected person or representative. Across the cases, established constitutional, administrative and labour institutions often provide stronger practical protection than newer AI-specific instruments when they already connect those links.

Abstract

Protection Reach is the operational capacity of a governance arrangement to place an affected person or representative inside an actionable pathway from notice and legal coverage to usable evidence, authoritative review and timely remedy. This paper develops the concept to evaluate the distance between visible AI-governance commitment and lived protection. Its strongest comparative finding is that established constitutional, administrative and labour institutions can sometimes provide more practical protection than dedicated AI instruments when they already connect standing, evidence, authority and remedy. Instrument novelty is therefore less important than whether the institutional chain completes.

The diagnostic uses six dimensions - visibility; legal and institutional coverage; evidence accessibility; contestability; institutional authority and independence; and remedy effectiveness - while treating participation timing and responsibility allocation as cross-cutting conditions. The dimensions are partly sequential: a missing record, inaccessible forum or delayed remedy can nullify strength elsewhere. This chain logic distinguishes Protection Reach from additive framework measurement, compliance activity, transparency alone and general access to justice.

A structured comparison applies the diagnostic to six algorithmic or automated-decision cases: Australia's Robodebt scheme; the Netherlands' SyRI litigation; Kenya's NIIMS/Huduma Namba judgment; India's Aadhaar authentication and welfare-access architecture; Spain's platform-work regime; and Canada's federal automated-decision framework. South Africa's Black Sash social-grant litigation remains a bounded, unscored institutional analogue. The cases show that collective representation is conditional on evidence and authority, that digital exclusion can be both the harm and a barrier to remedy, and that imported or distributed systems create cross-border evidence problems. The paper concludes with a policy architecture centred on dual notice, exception logging, evidence preservation and export, proportionate suspensive review, representative standing, independent authority, regional cooperation and sustained funding for public-interest intermediaries.

Keywords: AI governance; algorithmic accountability; administrative justice; contestability; access to remedy; public-sector AI; biometric identity; digital exclusion; algorithmic management; Global South; institutional capacity; Protection Reach.

Executive findings

No.	Proposition	Implication
1	Institutional inheritance can outperform instrument novelty.	Constitutional, administrative and labour institutions may provide stronger practical protection than dedicated AI instruments when they already connect standing, evidence, authority and remedy.
2	Framework presence is an unreliable proxy for protection.	The relevant test is whether an affected person can activate a pathway from notice to timely remedy, not whether a policy document exists.
3	Protection Reach is a chain with bottlenecks.	Visibility without evidence, appeal without suspensive effect, or oversight without corrective authority can each make a formally complete regime operationally weak.
4	Collective contestability is conditional, not self-executing.	SyRI shows the value of representation linked to evidence and an authoritative court; Kenya, Aadhaar and Spain show that mobilisation alone can leave implementation, proof and routine remedy incomplete.
5	Public-sector systems create accountability and exclusion asymmetries.	The state may demand identity, data and proof from people while withholding system records; authentication and documentation failures may also exclude a person before an appealable decision is recorded.
6	Timeliness is constitutive of remedy.	A review that succeeds after benefits are lost, work is terminated or an identity system becomes entrenched may produce legal correction without full lived protection.

Contents

1. Introduction
2. From Commitment to Protection
3. Literature and Conceptual Foundations
4. Defining Protection Reach
5. Theory and Propositions
6. Methodology and Case Selection
7. Comparative Cases
8. Comparative Findings
9. Protection Reach Diagnostic
10. Global Policy Implications
11. Policy Recommendations

11.1 Institutional implementation package

12. Limitations and Research Agenda
13. Conclusion
- Appendices
- References

Tables and figures

Table 1. Regulatory Sovereignty and Protection Reach

Figure 1. Six dimensions of Protection Reach

Table 2. Six dimensions and observable indicators

Figure 2. Institutional pathway from commitment to protection

Figure 3. Feedback loops that suppress evidence and pressure

Table 3. Propositions and observable implications

Table 4. Comparative diagnostic matrix for six algorithmic cases

Table 5. Recurrent failure points and consequences

Table 6. Protection Reach diagnostic scale

Table 7. Policy intervention matrix

1. Introduction

1.1 The puzzle

The global AI-governance landscape is increasingly document-rich and institution-poor. The Global Index on Responsible AI 2026 records some form of responsible-AI commitment in 128 of 135 governments assessed, yet reports an average score of roughly 35 out of 100 and evidence of implementation in only 55 per cent of cases where a framework is active, falling to 45 per cent in the Global South. Its implementation data identify training courses and workshops in 70 countries, compared with independent oversight bodies in 28, monitoring and evaluation systems in 22, standards or certification mechanisms in 23, and multistakeholder advisory processes in 16. [1] A separate OECD assessment reaches the puzzle from the public-sector side: only 11 of 36 OECD countries had a formal algorithmic-transparency law or standard, six had an open algorithm register, and 25 had neither. [2] Together, the sources show that formal commitment and visible activity can expand faster than the institutional pathways through which people obtain protection.

These figures do not establish that implementation activity is ineffective, nor do they measure whether a particular person obtained justice. They identify a governance puzzle: visible commitment can expand while the institutions that connect rules to people remain incomplete. The Index's foreword frames the danger directly - counting frameworks can allow recognition to stand in for remedy - and calls for resources for oversight, audits, redress mechanisms and public-interest groups. It further argues that rules protect affected workers only when they can reach them, systems can be contested, and governmental uses of AI are visible and accountable. [1] This paper treats that passage as an intellectual trigger, not as proof. Its task is to define and test what it means for a rule to “reach” a person.

1.2 The affected-person perspective

Most AI-governance measurement begins from the state, the organisation or the system: Is there a strategy? Is the rule binding? Has an impact assessment been completed? Is a human nominally in the loop? Those questions matter, but they do not reveal the institutional experience of the person whose benefit, job, migration status, identity, credit, health service or public entitlement is shaped by the system. From that position, governance appears as a sequence of practical questions. Was I told that an automated system was used? Can I identify the legal entity that deployed it? Do I know whether the human decision-maker relied on a score, recommendation or flag? Can I correct the data? Will the organisation preserve the relevant version, log and decision record? Can a union, lawyer or civil-society body act for me? Does the reviewing institution have power to compel evidence and pause the decision? Will the remedy arrive while it still matters?

The affected-person perspective therefore changes the unit of analysis. The object is not only the quality of a framework but the institutional pathway between a governance commitment and an actionable remedy. A system can comply with documentation requirements while remaining effectively unchallengeable. A legally binding right can be nominal when notice is absent, evidence is proprietary, procedures are unaffordable or review has no suspensive effect. Conversely, an older constitutional or sectoral regime may provide strong practical protection even without an AI-specific statute because courts, unions, ombuds institutions or regulators already have standing, evidence powers and remedial authority.

1.3 Research question, contribution and boundary

The central research question is: What institutional conditions determine whether AI-governance commitments translate into accessible, enforceable and effective protection for people affected by AI systems? The sharper policy question is how global AI governance should be evaluated when governments can demonstrate strategies, laws and implementation activities but affected people still cannot identify, challenge or obtain remedy from AI-mediated exercises of power.

The paper contributes the concept of Protection Reach and an operational diagnostic. It examines the downstream relationship between deploying institutions and affected individuals, workers and communities. An earlier ITS paper analysed the upstream state-provider relationship - dependence, substitutability, procurement leverage and provider-facing enforcement - and is cited once here to establish the boundary between the two projects. [34] The question in this paper is whether the resulting governance arrangement can be activated by a person whose life is affected.

The scope is bounded in four ways. First, the paper addresses high-impact administrative and labour settings rather than all uses of AI. Second, “AI” is used functionally to include algorithmic and automated decision systems where classification disputes would otherwise distract from the exercise of mediated power. Third, the cases are theory-building probes, not a country ranking. Fourth, public evidence is uneven; diagnostic scores are qualified judgments with confidence statements, not measures of national virtue.

Table 1. Regulatory Sovereignty and Protection Reach

Dimension	Regulatory Sovereignty	Protection Reach
Primary relationship	State -> technology provider	Deploying institution / AI system -> affected person or community
Core question	Can the state set, verify and enforce rules against an indispensable provider?	Can the affected person identify, challenge and remedy AI-mediated power?
Main constraints	Dependence, substitutability, procurement leverage, infrastructure and evidence control	Notice, coverage, evidence access, standing, authority, affordability and remedy timing
Unit of analysis	State-provider bargaining relationship	Defined system or governance arrangement around an affected population
Characteristic failure	Formal jurisdiction without credible enforcement leverage	Formal right or activity without an actionable institutional pathway
Policy response	Provider competition, exit, shared assurance, procurement leverage, regional coalitions	Registers and notice, evidence preservation, collective standing, independent authority, suspensive review and timely remedy

2. From Commitment to Protection

2.1 Framework proliferation and measurement limits

Framework-based measurement remains useful because it reveals political attention, normative coverage and the formal allocation of responsibility. The Global Index reports that Global South frameworks substantially broadened the responsible-AI topics they covered between editions, even though 78 per cent of framework cases in those countries were non-binding. [1] The OECD findings introduced in Section 1.1 add a separate measure of how rarely public-sector transparency standards and registers have become routine. [2] These indicators expose gaps that a simple count of national AI strategies would conceal, but neither source measures whether a particular affected person completed a pathway to remedy.

Yet an implementation indicator can still be too remote from protection. A workshop is an activity; a published impact assessment is a document; a register entry is visibility infrastructure. None establishes that an affected person was notified, obtained evidence, secured review or received correction. Implementation is therefore best treated as a middle category between commitment and outcome. It shows that a rule has generated organisational activity, but not necessarily that the activity has altered the distribution of decision authority, evidence or remedy.

2.2 Protection as an institutional pathway

International norms increasingly recognise the pathway elements. The Council of Europe Framework Convention requires documentation relevant to AI systems and their use to be available to affected persons in a form sufficient to challenge decisions and the use of the system itself; it also calls for effective complaints and procedural safeguards. [3] The EU AI Act creates a right to complain to market-surveillance authorities and, for specified Annex III high-risk systems, a right to a clear and meaningful explanation of the role of the system and the main elements of an adverse individual decision. [4] The UN Guiding Principles on Business and Human Rights place access to effective judicial and non-judicial remedy alongside state duties to protect and corporate responsibilities to respect rights. [5]

These instruments help identify the components of protection, but they do not guarantee their connection in practice. A complaint right may be broad while the evidence needed to substantiate the complaint remains with the deployer. An explanation right may apply only to specified systems or decisions. A regulator may receive complaints but lack technical staff, compulsory access or authority to suspend a deployment. Effective protection is therefore relational and sequential: each stage depends on institutions performing in time and on information moving across organisational boundaries.

2.3 Three distinctions

First, legal bindingness is not Protection Reach. Bindingness concerns the force of a norm; reach concerns whether the norm can be activated and completed in an operational setting. Second, transparency is not Protection Reach. Transparency may expose a system, but visibility that does not generate evidence, standing or action can become documentation without consequence - a limitation long identified in scholarship on algorithmic transparency. [10] Third, human oversight is not Protection Reach. A human reviewer located inside the deploying organisation may be under time pressure, lack the authority or information to depart from the output, or simply ratify it. Studies of public-sector human oversight warn that nominal review can become an empty procedural shell and may reproduce automation bias. [12-13]

The movement from commitment to protection is thus not a linear compliance checklist. It is an institutional conversion process in which rules acquire practical force through notice, evidence, representation, authority and remedy.

3. Literature and Conceptual Foundations

3.1 Administrative justice and technological due process

Administrative justice supplies the oldest foundation for Protection Reach. Due process is not satisfied by the abstract availability of a rule; it depends on notice, an opportunity to be heard, reasons and a decision-maker capable of reviewing the case. Citron's account of "technological due process" showed how automated administration can combine rulemaking and adjudication while weakening the procedural safeguards of both: code translates policy into operational rules, individual hearings may presume computer accuracy, and the person affected cannot see where policy, data and implementation diverge. [6] This insight shifts attention from model explainability alone to the institutional production of decisions.

Accountability scholarship similarly treats answerability and consequences as a relationship among an actor, a forum and a duty to explain and justify conduct. Algorithmic systems complicate that relationship because design, data, integration, deployment and review are distributed across providers, contractors and public bodies. Wieringa's review emphasises that algorithmic accountability is networked and socio-technical; the relevant account must specify who owes what information to which forum, at what stage and with what consequences. [11] Protection Reach adopts this relational structure but centres the ability of the affected person or representative to enter it.

3.2 Contestability, reviewability and evidence

Contestability is more demanding than explanation. Kaminski and Urban argue for a right to contest AI decisions that adapts due-process principles to digital systems and complements ex ante governance. [8] The right is meaningful only when it permits a person to question the factual basis, the rule, the procedure and, where appropriate, the legitimacy of using the system at all. An explanation that merely describes influential factors can leave the decision institutionally insulated. Malgieri and Comandé's account of legibility likewise treats intelligibility as a legal and procedural condition rather than a demand for unrestricted source-code disclosure. [32]

Reviewability provides the evidentiary bridge. Cobbe, Lee and Singh conceptualise automated decision-making as a socio-technical process extending before and after the decision, and argue for context-specific records across technical and organisational stages so that individual decisions and the process as a whole can be reviewed. [9] Kroll and colleagues likewise show that accountability may require technical mechanisms that verify compliance rather than disclosure of source code alone. [7] Raji and co-authors place audit artefacts across the lifecycle to preserve organisational decisions and responsibilities. [33] Together these approaches support a central claim of this paper: evidence accessibility concerns not only whether a model can be explained but whether the institution has preserved the record needed to test the decision, identify responsibility and reconstruct the system version in use.

Evidence has at least four layers. The first is personal evidence: the input data, decision record, reasons and correction history relevant to an individual. The second is operational evidence: logs, thresholds, model versions, overrides, incidents and quality monitoring. The third is governance evidence: impact assessments, procurement records, peer reviews, audits and approvals. The fourth is structural evidence: patterns of disparate impact, repeated complaints and population-level outcomes. A system may be contestable at one layer and opaque at another. Individual explanations can therefore coexist with an inability to challenge the deployment itself.

3.3 Transparency, registers and the limits of seeing

Transparency is necessary for many challenges because an undisclosed system rarely produces a complaint directed at the system. Public registers can identify the deploying body, purpose, vendor, risk assessment and contact point. The OECD reports that the Netherlands' register contains more than 1,300 algorithms and the United Kingdom's Algorithmic Transparency Recording Standard listed 131 records at the time of its 2026 assessment. [2] Canada's first federal AI Register, launched as a minimum viable product in November 2025, drew from impact assessments, access-to-information requests, parliamentary responses and service inventories; the government cautions that entries are incomplete and the register is an early version. [16]

Registers nevertheless illustrate the difference between visibility and protection. Ananny and Crawford warn that seeing a system is not equivalent to understanding or governing it. [10] A register entry can omit the decision threshold, error distribution, contractual limits, system version or route to challenge. It may also be published at an organisational level that is too general for an individual to connect it to a decision. Protection Reach therefore asks whether visibility is

actionable: does disclosure identify an accountable entity, link to the relevant evidence and forum, and occur before or at the time the system affects the person?

3.4 Labour rights and collective contestability

Labour law demonstrates why collective institutions matter. Algorithmic management distributes decisions across rating, allocation, pricing, monitoring and deactivation systems. Individual workers often lack technical information, bargaining power and protection from retaliation. Collective bargaining can aggregate incidents, obtain recurring information, negotiate system changes and convert a private complaint into a workplace governance issue. The ILO's analysis of platform work identifies algorithmic information rights as a prerequisite for effective negotiation and notes that Spain's statutory right was operationalised in the Just Eat collective agreement through an Algorithm Commission. [28]

The ILO's Convention No. 193 on decent work in the platform economy, adopted on 12 June 2026, includes information duties concerning automated systems used to monitor, evaluate or make work-related decisions. As of the research cut-off it had no ratifications and was not in force. [27] It is therefore used once here as evidence of an emerging normative baseline, not as evidence of existing Protection Reach or of operational outcomes in Spain.

3.5 Participation, authority and timing

Participation literature distinguishes information, consultation and shared authority. Arnstein's classic ladder remains useful because it asks whether participation redistributes power or merely communicates a decision already made. [36] In AI governance, timing is decisive. Consultation after procurement may identify concerns but leave the system, contract, data architecture and performance target locked in. Participation during design can change features but not the earlier decision to automate. Participation when the public problem is first defined can influence whether AI is an appropriate response.

This paper therefore treats participation timing as cross-cutting rather than a seventh score. The relevant question is not whether a meeting occurred, but what decision remained open and whether participants had evidence, representation and influence. Participation can improve legitimacy without changing authority. It can also become a source of Protection Reach when it creates continuing rights: access to monitoring data, representation in oversight, veto or escalation procedures, and review of system changes.

3.6 Institutional capacity, street-level discretion and remedy

Administrative systems do not implement themselves. Street-level bureaucracy research shows that frontline officials translate policy under constraints of workload, resources and organisational incentives. [37] Automation can relocate discretion rather than remove it: discretion moves into data selection, threshold setting, exception handling, system procurement and the decision to override. Koulu's procedural account of human oversight and Green's critique of nominal review both show why a human reviewer with a formal power to disagree may still lack the time, information or organisational authority to exercise it. [12][13]

Remedy closes the accountability relationship. It includes reversal, restoration of a service or entitlement, correction of data, reinstatement, compensation, suspension of the system, sanctions, public disclosure and systemic reform. Its quality depends on timeliness and scope. Counterfactual explanations may help a person alter future inputs, but they do not necessarily correct an unlawful rule or a false factual premise. [31] Protection Reach therefore distinguishes recourse - changing one's circumstances to obtain a different outcome - from contestation - asserting that the decision or system is wrong - and from remedy - an authoritative response that corrects or prevents harm.

4. Defining Protection Reach

4.1 Definition

Protection Reach is the operational capacity of a governance arrangement to place an affected person or representative inside an actionable institutional pathway: from timely notice and legal coverage, through access to contestable evidence and an authorised forum, to correction, suspension, compensation or systemic change within the timeframe of the harm.

The definition has five properties. It is relational because reach exists between a person and a set of institutions, not inside a policy document. It is operational because nominal rights are insufficient unless they can be used. It is time-sensitive because delayed protection may not protect. It is representative because unions, civil-society bodies, lawyers, ombuds institutions and class actions can act where individual challenge is unrealistic. It is both individual and systemic because a remedy may need to correct one decision, a recurring process or the deployment itself.

4.2 Boundaries from related concepts

Framework presence records whether a policy or legal instrument exists. Bindingness records its legal force. Regulatory capacity concerns a state's administrative ability to design and administer rules, while regulatory sovereignty concerns provider-facing enforceability under conditions of technological and economic interdependence. Implementation activity records actions taken under a framework. Compliance records whether regulated actors conform to obligations. Transparency concerns the availability of information. Algorithmic accountability is a wider field of answerability, assessment and consequences. Human oversight is one mechanism within a decision process. General access to justice concerns the broader legal environment in which claims can be made.

Protection Reach intersects with each concept but asks a narrower empirical question: can the person affected activate the relevant arrangement and obtain an authoritative, timely result? A binding law may have weak reach if standing is narrow, evidence cannot be compelled or the regulator is advisory. A voluntary sector code may have some reach if a well-resourced ombuds institution can investigate and order compensation under another legal authority. Protection Reach is therefore not a substitute for binding law; it is a test of how formal and informal elements combine in practice.

4.2.1 Relationship to the Global Index on Responsible AI

GIRAI and Protection Reach are complementary but not interchangeable. GIRAI assesses countries across responsible-AI themes, records the presence of frameworks and government-led implementation activities, and aggregates indicators for national comparison. Protection Reach uses a different unit: a defined deployment or governance domain around an affected population. GIRAI asks what a state has committed to and built; Protection Reach asks whether a person or representative can activate what exists and complete a pathway to remedy. [1]

The difference is also structural. An additive index can reveal broad institutional scarcity, but aggregation can conceal a binding constraint: strong policy coverage, training and transparency cannot compensate for absent evidence or a forum without corrective power. Protection Reach therefore disaggregates to locate the bottleneck. GIRAI can identify countries or themes where pathway assessment is urgent; Protection Reach can explain why a respectable national profile may coexist with weak protection in a specific welfare, identity or workplace system.

4.3 Six dimensions

Visibility asks whether affected people can know that an automated or algorithmic system is used and identify its role, purpose, deployer and responsible contact. Legal and institutional coverage asks whether relevant people, harms and organisational relationships fall within applicable law and mandates. Evidence accessibility asks whether decision records, reasons, data, versions, logs, audits and incident information are preserved and obtainable by the person, representative, court or regulator. Contestability asks whether the person can challenge the output, process or deployment through affordable and accessible procedures, with protection from retaliation and, where necessary, suspensive effect. Institutional authority and independence asks whether a forum can compel evidence, investigate, coordinate and order correction or suspension without dependence on the deploying body. Remedy effectiveness asks whether the response prevents, stops, corrects or compensates harm in time and whether it can generate systemic reform.

These dimensions are partly complementary. A weakness in one can block the chain: no visibility prevents challenge; no evidence prevents proof; no authority makes the complaint symbolic; no timely remedy turns success into retrospective recognition. The dimensions are not perfectly multiplicative, however. Collective actors can sometimes overcome low individual visibility through investigation. Courts can order disclosure when administrative access is weak. Political scrutiny can produce systemic correction without individual litigation. The diagnostic therefore presents a disaggregated profile rather than a single global score.

4.4 Unit of analysis and levels of reach

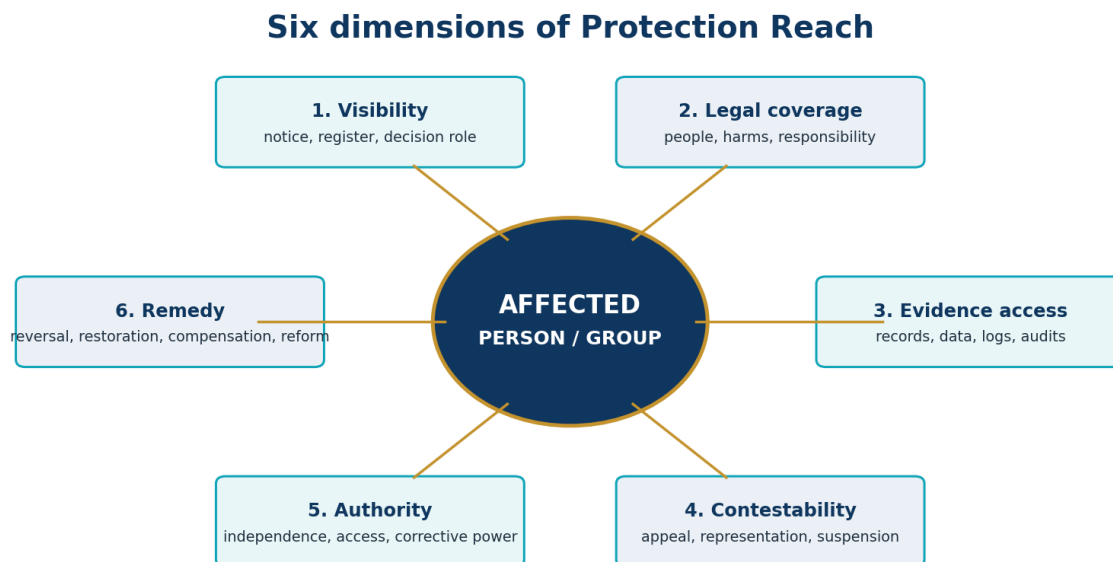
The unit of analysis is a governance arrangement around a defined system or domain, not an entire country. Reach can also differ across three levels. Decision-level reach concerns correction of an individual output. System-level reach concerns the ability to challenge the design, deployment or rule embedded in the system. Population-level reach concerns whether people who are most affected - including informal workers, migrants, people without stable digital access or communities exposed to group harms - can use the mechanism.

A regime can be strong at one level and weak at another. An individual may obtain a new human review while the system continues unchanged. A court may invalidate a legal basis without compensating everyone affected. A register may provide population-level visibility while offering no individual pathway. This layered approach prevents the concept from being reduced to an appeal right.

4.5 Participation timing as a cross-cutting condition

Participation is assessed at six moments: problem definition; pre-procurement; design; implementation; post-deployment; and continuing oversight. At each moment, the analysis asks who participated, what information they possessed, what decision remained open and what authority followed. A consultation held after architecture and contracts are fixed may produce information but little decision influence. By contrast, continuing representation with access to monitoring data and escalation rights can strengthen several dimensions at once.

Figure 1. Six dimensions of Protection Reach



Participation timing cuts across all six: problem definition -> procurement -> design -> implementation -> deployment -> continuing oversight

Table 2. Six dimensions and observable indicators

Dimension	Observable indicators
Visibility	Notice; public register; system purpose; deployer and provider; automated/assisted role; impact assessment; incident disclosure; service-entry and exception notice where a person is excluded before a decision.
Legal and institutional coverage	Coverage of employees, platform and informal workers, service users, migrants, children and communities; people lacking documents or connectivity; indirect and collective harm; responsibility across providers, deployers, contractors, donors and public bodies.
Evidence accessibility	Decision and exception records; reasons; relevant data; system/model version; logs; audits; incidents; evidence preservation and export; access through representatives; local or reachable evidence holder.
Contestability	Human reconsideration; administrative and judicial review; collective and community representation; challenge to system use; suspension; anti-retaliation; local-language, offline and affordable procedures.
Authority and independence	Independent budget and appointments; technical competence; inspection and compulsory access; authority over public-sector AI; suspension and coordination powers.
Remedy effectiveness	Prevention; reversal; restoration; data correction; reinstatement; compensation; system restriction; sanctions; disclosure; prevention of recurrence; timeliness.

5. Theory and Propositions

5.1 Institutional pathway

The proposed pathway is:

Framework or legal commitment -> institutional mandate -> system visibility -> evidence availability -> accessible contestation -> authoritative investigation -> timely remedy -> actual protection.

The pathway is not a claim that every case follows a fixed sequence. It identifies the minimum conversions through which a general commitment becomes usable. The chain can break at mandate ambiguity, undisclosed deployment, missing records, narrow standing, unaffordable procedures, weak authority or delayed remedy. Responsibility fragmentation creates additional breaks where a provider controls model information, an integrator controls logs, a contractor operates the service and a public authority formally owns the decision.

5.2 Feedback loops

Three feedback loops can stabilise weak reach. Low visibility produces few complaints, which creates little documented evidence of harm, which reduces political pressure to improve visibility. Weak remedies reduce trust in complaint mechanisms, producing under-reporting that can be misread as absence of harm. Individualised contestation imposes high costs on affected people, leading to few precedents and allowing the same system to continue without systemic correction. These loops explain why low complaint numbers are not reliable evidence of safety.

5.3 Propositions and rival explanations

The comparison examines four propositions rather than formal hypotheses. P1 - Framework-protection divergence: the existence or comprehensiveness of an AI framework will not reliably predict Protection Reach where notice, evidence, authority and remedy institutions are weak. P2 - Visibility, evidence and responsibility as gateway conditions: undisclosed systems, unpreserved records and fragmented ownership of evidence suppress ordinary challenges and make correction harder. P3 - Conditional collective contestability: representative organisations reduce individual burdens, but stronger outcomes occur only when they can obtain usable evidence and reach an authoritative forum. P4 - Authority and remedy timing: independent bodies with compulsory access, corrective powers and proportionate interim relief provide stronger protection than advisory or non-suspensive review. Participation timing and responsibility allocation are treated as cross-cutting scope conditions within P2 and P4 rather than separate propositions.

Rival explanations include general rule-of-law quality, judicial independence, media attention, political salience, sectoral rights, litigation funding, organised civil society and administrative competence. Cases may also produce apparently strong remedies because a crisis becomes exceptional, not because the ordinary pathway works. The propositions are therefore theory-building expectations, not claims tested through statistical inference. The comparison asks which mechanisms are observed, where counterevidence appears and what the design cannot establish.

Figure 2. Institutional pathway from commitment to protection

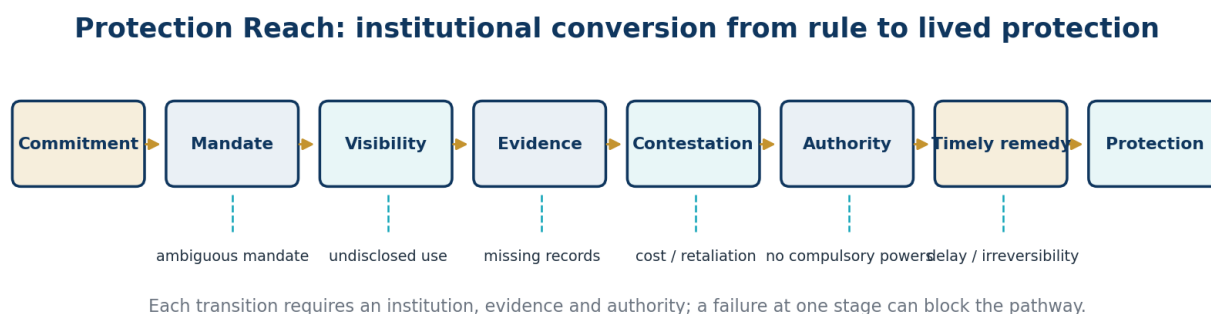
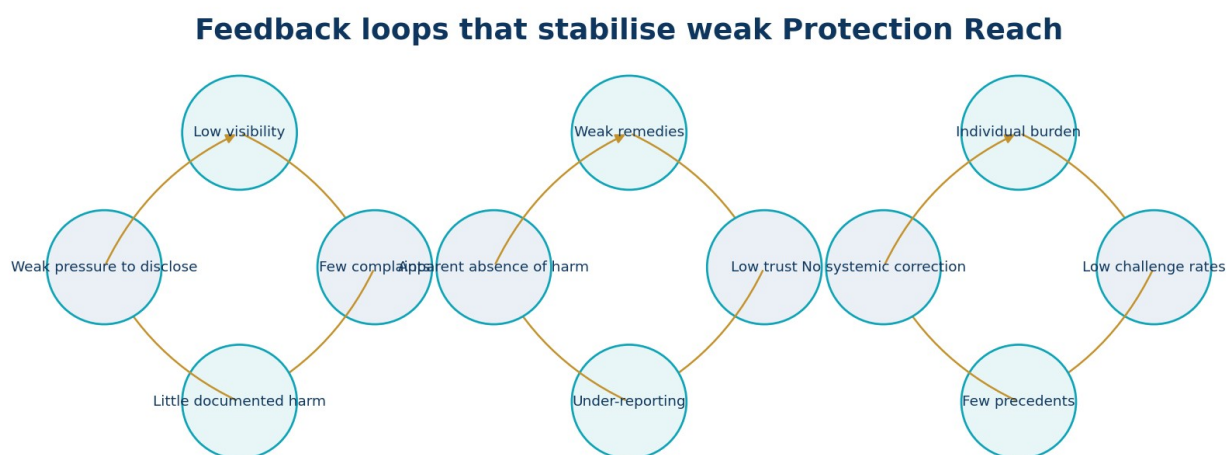


Figure 3. Feedback loops that suppress evidence and pressure

Low complaint volumes can be an outcome of weak reach, not evidence that a system is safe.

Table 3. Propositions and observable implications

Proposition	Observable implication
P1 Framework-protection divergence	Framework coverage or implementation activity rises without corresponding notice, evidence access, authoritative review or remedy outcomes.
P2 Visibility, evidence and responsibility	Evidence strand: undisclosed systems and missing or non-reconstructable records generate fewer system-specific challenges and weaken proof of harm. Responsibility strand: fragmented provider-integrator-deployer or agency-to-agency responsibility weakens remedial ownership unless evidence, escalation and correction duties are allocated in advance.
P3 Conditional collective contestability	Representative bodies reduce cost and retaliation burdens, but broader remedies occur only where they can obtain usable evidence and reach an authoritative forum.
P4 Authority and remedy timing	Independent bodies with compulsory access, corrective powers and proportionate interim relief prevent or stop harm more effectively than advisory or delayed review; early participation is a scope condition.

6. Methodology and Case Selection

6.1 Design and evidence hierarchy

The study uses structured, focused comparison for theory building. Each case is analysed through the same questions: system and affected population; visibility; coverage; evidence; contestability; authority; remedy; participation timing; outcome; confidence; and rival interpretation. Primary evidence is prioritised: legislation, judgments, government directives, royal-commission findings and official registers. Peer-reviewed research and credible civil-society assessments are used to interpret implementation where official sources do not establish lived outcomes. The coding template was also informed by the Council of Europe's HUDERIA methodology and its context-based risk-analysis model, used here as a source of assessment domains rather than as evidence that a remedy pathway exists in practice. [35]

The Lexicon method is applied as a critical-reading discipline. Facts are separated from institutional interpretations, inferences and normative claims. For each case, the analysis identifies decision authority, absent actors, governance timing, evidence ownership, operational constraints, incentives, contradictions and feedback loops. Confidence is rated high where the legal outcome and institutional powers are documented in primary sources; moderate where implementation outcomes rely on secondary evidence; and low where public information does not reveal routine practice.

6.2 Case selection

The six core cases vary across sector, region, governance form and outcome. Australia's Robodebt is a negative public-sector case with eventual large-scale remedy but severe delay. The Netherlands' SyRI case shows collective system-level contestation and judicial invalidation. Kenya's NIIMS/Huduma Namba judgment provides a negative test of collective contestability because authoritative litigation did not itself create a routine correction pathway. India's Aadhaar architecture adds a large-scale Asian identity and welfare case in which constitutional review, statutory administration and grievance channels coexist with fragmented evidence and service-delivery responsibility. Spain's platform-work regime tests collective information rights against fragmented employment structures. Canada's federal regime is an ex ante governance case combining a directive, impact assessments, peer review and an emerging register, but with limited public evidence of routine remedial outcomes. South Africa's Black Sash litigation remains only a bounded analogue of standing, cross-actor responsibility, supervision and time-sensitive remedy; it is not included in the scored algorithmic comparison.

The sample is purposive and outcome-rich: litigation, commissions, audits and official registers make institutional pathways visible. It cannot estimate how frequently organised civil society succeeds or whether quiet deployments have stronger or weaker reach. Selection bias is addressed analytically, not eliminated, by treating Kenya and Aadhaar as hard negative tests of the assumption that mobilisation plus access to a supreme or constitutional court necessarily produces routine Protection Reach; Spain as a limiting test of statutory collective information; Canada as strong process without demonstrated remedial completion; and Black Sash as an unscored analogue. Claims remain conditional mechanisms rather than average effects.

6.3 Coding

Each dimension is coded with a single integer on a four-point scale: 0 absent; 1 nominal; 2 partial; 3 credible. Ranges, arrows and hybrid labels are not permitted in the score cells. "Credible" does not mean perfect or universal; it means that affected people or representatives can normally use the relevant mechanism in the defined domain and period. Temporal changes, decision-level versus system-level reach and later correction are recorded in narrative notes, not blended into the score. Where evidence is missing, the assessment records the evidenced level and lowers confidence; it does not infer a midpoint or convert absence of evidence into moderate confidence.

7. Comparative Cases

7.1 Australia: Robodebt - eventual remedy after institutional failure

The Income Compliance Program used annual income data from the Australian Taxation Office, averaged across fortnights, to identify and raise social-security debts between July 2015 and November 2019. Parliamentary reviews found that automation enabled roughly 20,000 compliance interventions a week, replacing a process that had previously required manual inquiry; the burden shifted to recipients to produce historical records and disprove asserted debts. [23] Services Australia states that eligible Robodebts were based on averaged ATO income information and that most refunds were processed in 2020. [24]

Protection Reach was weak during operation. Recipients received debt communications, but the system's evidentiary method and legal defect were not meaningfully visible to most individuals. The state possessed the data and procedural knowledge; recipients were required to reconstruct old payslips or bank records. Formal review routes existed, yet debt recovery could proceed and the individual burden was high. The Royal Commission later recommended that recovery be ethical, proportionate and transparent, that it normally pause while a debt is disputed or reviewed, and that recipients have ample opportunity to challenge proposed debts before recovery. [22] The recommendation itself demonstrates that the earlier pathway lacked an effective suspensive safeguard.

Collective institutions eventually transformed the outcome. Parliamentary inquiries, legal-aid and advocacy work, individual litigation, a class action and a Royal Commission created visibility and evidence that ordinary administration had not supplied. The scheme ended in November 2019; debts were refunded or reduced to zero, and the first class-action settlement distributed about A\$112 million in 2022. In June 2026, the Federal Court approved an additional A\$475 million settlement, with separate amounts for legal and administration costs. [25] This is a substantial remedy, but it arrived years after debts, distress and loss of trust had occurred.

Outcome: strong retrospective accountability, weak contemporaneous protection. The case supports P4 on remedy timing and provides evidence for P3 because exceptional collective mobilisation overcame an ordinary process that burdened individuals. It also illustrates the public-sector accountability gap embedded in P1 and P2. Rival interpretation: Australia ultimately demonstrated strong rule-of-law capacity. That is correct, but the capacity was activated after routine safeguards failed. Confidence: high for legal and institutional chronology; the public record does not support a quantified or complete account of individual lived harms.

7.2 Netherlands: SyRI - collective challenge to the system itself

SyRI was a statutory risk-indication system that enabled public bodies to combine data for fraud detection. In February 2020, the District Court of The Hague held the SyRI legislation incompatible with Article 8 of the European Convention on Human Rights because it did not strike a fair balance and was insufficiently transparent and verifiable. The plaintiffs included human-rights and privacy organisations, a national clients' council and individuals; the Dutch trade-union federation FNV intervened. [18]

The case illustrates the value of collective standing. No single claimant had to prove that a particular risk score caused an adverse decision. Representative bodies challenged the legal architecture and use of the system. The court examined the design of the statutory safeguards, the opacity of risk models and indicators, the scope of data processing and the concentration of oversight in the responsible minister. It declared the provisions specific to SyRI inoperative, producing a system-level remedy rather than a new individual calculation.

Visibility was still limited before litigation. People in targeted neighbourhoods could not readily know the risk model or whether they were included, and the government defended secrecy on enforcement grounds. Evidence access was therefore partial, but litigation and judicial scrutiny converted public-law documentation into an authoritative assessment. The remedy was timely in the sense that it halted the legal basis for future use, although it did not provide individual compensation or reconstruct all prior effects.

Outcome: credible system-level contestability and authority, enabled by civil society, a clients' council, a trade union and an independent court. The case supports P3 and P4 and qualifies P2: low individual visibility did not prevent challenge because collective organisations created system-level visibility and standing. Rival interpretation: the judgment is primarily a privacy case rather than an AI-governance case. That is also its analytical value; existing human-rights and administrative-law institutions reached an algorithmic system without waiting for dedicated AI legislation. Confidence: high.

7.3 Kenya: NIIMS/Huduma Namba - conditional protection and unresolved exclusion

Kenya's National Integrated Identity Management System (NIIMS), associated with Huduma Namba, was designed as a single source of identity information for citizens and resident foreigners. Civil-society and human-rights organisations challenged the legislative basis, biometric collection, privacy safeguards, exclusion risks and public participation. In January 2020, the High Court declared the collection of DNA and GPS coordinates intrusive and unconstitutional, and permitted NIIMS implementation only after enactment of an appropriate and comprehensive regulatory framework compliant with constitutional requirements. [19]

The judgment contains an unusually explicit institutional analysis. It observed that effective data protection requires legal, regulatory and institutional safeguards, that data subjects had limited insight into and control over how information was used, and that implementation of the new Data Protection Act required a commissioner, registration of controllers and processors, and operational regulations. It also recognised exclusion risks for people without identity documents or with poor biometrics and called for a framework addressing enrolment. [19]

Protection Reach was mixed. The system was publicly salient and the litigation created strong collective contestability. The court had authority to invalidate elements and condition implementation. Yet the remedy stopped short of invalidating NIIMS as a whole; it allowed the government to proceed once a framework was enacted. Individuals at risk of future exclusion did not receive a standing, routine mechanism for correcting biometric failure or contesting service denial in the judgment itself. Participation was litigated, but the court's broader discussion shows the difference between an opportunity to submit views and influence over the architecture of a nationwide identity system.

Outcome: meaningful constitutional constraints and a strong judicial forum, but partial operational reach for future users. The case supports P4 on authoritative institutions and illustrates participation timing as a scope condition, while providing counterevidence to any unconditional version of P3: organised civil society and collective litigation did not by themselves create a routine correction mechanism or halt the architecture as a whole. Rival interpretation: the conditional approach may have been proportionate because digital identity can support access to services. The diagnostic does not reject that goal; it asks whether exclusion, data-use and correction safeguards were operational before dependence deepened. Confidence: high for the judgment. Subsequent lived protection is outside the case record and is treated as unobserved.

7.4 India: Aadhaar - constitutional review without a complete everyday remedy chain

Aadhaar is a nationwide biometric identity and authentication architecture rather than a machine-learning model, but it falls within this paper's functional scope because automated biometric matching can become a gateway to welfare and public services. In September 2018, the Supreme Court of India upheld the core statutory scheme and Section 7 welfare use by majority, while striking down or reading down several provisions. The Court recognised that authentication becomes effectively compulsory for people seeking Section 7 benefits, rejected the constitutional challenge based on what

it treated as an unproven plea of exclusion, and required alternative identity routes for children unable to produce Aadhaar for covered benefits. It also criticised the then complaint rule that prevented an individual victim from initiating criminal proceedings and indicated that amendment was appropriate. [38]

The case provides a harder test of collective contestability than a simple win-or-loss account. A large constitutional challenge reached the highest court and produced limits on commercial use, disclosure and some forms of mandatory identification, but the majority left the central welfare-authentication architecture intact. A system-level forum was available; the result did not create a single routine procedure through which a person denied a ration, payment or service could obtain the authentication record, compel the relevant requesting entity and secure immediate restoration.

Operational responsibility is distributed. UIDAI currently provides phone, portal, email, walk-in, post, social-media and government grievance channels, while its authentication guidance requires exception handling and states that technological or biometric limitations should not cause denial of service. [40a-40c] Yet the service-delivery agency, authentication user agency, device operator and UIDAI may each possess different parts of the record and different corrective powers. The Comptroller and Auditor General found that complaints received through regional channels were not centrally captured, unresolved complaints were not escalated, complaint histories could not be tracked, and large pendencies compromised the effectiveness of grievance redress. [39] The audit predates current system changes, so it is evidence of a documented institutional failure point rather than proof of present performance.

Outcome: high public visibility and authoritative constitutional review, but partial everyday Protection Reach. Aadhaar shows that digital exclusion can occur before a conventional adverse-decision record exists: poor biometrics, network failure or documentation problems may prevent entry into the service process itself. It supports P3 conditionally because collective litigation achieved important safeguards without guaranteeing routine individual correction, and it supports P2 and P4 because evidence fragmentation and distributed authority remain central. Confidence: high for the judgment and CAG audit; moderate for the design of current grievance and exception channels; routine service-denial outcomes are not established by the public sources used here. The evidence score remains 1 because the sources show multiple records and channels, but not a single reconstructable case file accessible to the affected person or a reviewer across UIDAI, the authentication user agency, the device operator and the service-delivery body.

7.5 South Africa: Black Sash - a bounded institutional analogue, not an AI case

South Africa's 2017 Black Sash litigation concerned the impending expiry of an invalid contract through which Cash Paymaster Services paid social grants on behalf of the South African Social Security Agency. It was not an AI or algorithmic-decision case and is not used to validate AI-specific claims about model opacity, statistical error or automated causation. It is retained as a bounded analogue because a technology-mediated essential service faced interruption, responsibility was split between a public authority and a private contractor, and the court had to connect rights to evidence, supervision and time-sensitive relief.

The transferable institutional mechanisms are specific. The Constitutional Court granted direct access to a public-interest organisation; bound the public authority and contractor to a common constitutional outcome; required audited financial information and quarterly reporting; used continuing supervision; and designed relief around uninterrupted grant payment. [20] In April 2026, the Court ordered the contractor to refund adjusted certified profit of R81.3 million to SASSA, demonstrating the long tail of supervised financial accountability. [21] These features inform the paper's analysis of representative standing, cross-actor responsibility, evidence duties and remedy timing.

The analogy has clear limits. It does not show that affected people received notice of an algorithmic role, that model or decision logs were accessible, or that a court could reconstruct an evolving automated system. Contractual and financial evidence is not a substitute for model-version, threshold, override or disparate-impact evidence. The case is therefore excluded from the comparative scorecard and does not carry the paper's AI-specific findings. Its narrower lesson is that standing, compulsory reporting, cross-actor responsibility and continuing remedial supervision can be transferred into algorithmic settings when law supplies the necessary technical evidence powers.

Outcome: a strong institutional analogue for population-level continuity and accountability, but not evidence of AI-specific Protection Reach. Exceptional Constitutional Court litigation is also resource-intensive and cannot substitute for routine administrative pathways. The case informs Recommendations 4 and 5 and the design of time-sensitive relief; it is not treated as confirmation of P2 or as the strongest empirical case in the comparison. Confidence: high for the orders and institutional mechanism; transfer to algorithmic settings is a bounded analytical inference.

7.6 Spain: platform work - binding information rights with uneven operational reach

Spain's Law 12/2021 amended the Workers' Statute in two material ways. It gave workers' representatives a right to information about the parameters, rules and instructions underpinning algorithms or AI systems that affect working conditions, access to employment and job retention, including profiling. It also created a presumption of employment for

paid delivery work where the platform exercises organisational, directional and control powers through algorithmic management. The law emerged from social dialogue and drew on Supreme Court jurisprudence. [26]

The regime strengthens legal coverage and collective contestability. The information right is held by representatives, not left solely to individual workers, and labour status opens access to wages, leave, social protection, inspection and collective bargaining. The Just Eat collective agreement operationalised the statutory right through a joint Algorithm Commission. [28] This is a concrete example of participation becoming continuing institutional access rather than a one-off consultation.

Operational reach remains uneven. A 2025 International Labour Review study found that reclassification improved fixed wages, paid leave and social protection in parts of the sector, while outsourcing, involuntary part-time work, intensified surveillance and ineffective access to platform algorithms persisted. [29] Fairwork's 2025 assessment similarly reported progress toward employment models but continuing employer fragmentation, low wages and algorithmic opacity. [30] These are secondary assessments and do not prove uniform non-compliance, but they indicate that statutory information can be diluted by subcontracting, commercial secrecy, incomplete representation and the technical form in which information is provided.

Outcome: stronger coverage and collective institutions than an individual-transparency model, but partial evidence accessibility and uneven enforcement. The case provides both support and counterevidence for P3: representation creates an institutional route, yet incomplete technical information, subcontracting and employer fragmentation prevent that route from becoming uniformly effective. It also illustrates responsibility fragmentation as a scope condition under P2. A legal entitlement to “parameters, rules and instructions” does not automatically produce a usable account of an evolving system. Confidence: high for legal design; moderate for sector-wide operational outcomes based on convergent secondary evidence.

7.7 Canada: federal automated decisions - strong ex ante process, limited remedy evidence

Canada's Directive on Automated Decision-Making applies to federal departments using systems that fully or partially automate administrative decisions or related assessments in production. It covers systems that assist judgment, including scores, summaries and recommendations presented to human officers, and aims to support transparency, accountability, legality and procedural fairness. It requires impact assessment, transparency, quality controls, recourse and public reporting. [14]

The Algorithmic Impact Assessment (AIA) is mandatory and contains 65 risk questions and 41 mitigation questions. Departments must publish final results in accessible form and both official languages on the Open Government Portal before production, review them when system scope or functionality changes, and apply scaled requirements. [15] Projects at impact level two or above require peer review, with the complete review or a plain-language summary published before production. [17] Canada's AI Register adds a cross-government visibility layer, although the government describes the current version as an incomplete minimum viable product assembled from multiple sources. [16]

This architecture performs strongly on ex ante visibility and governance evidence. It recognises partial automation, reducing the risk that a nominal human decision-maker removes a system from scope. Publication before production improves governance timing. The weaknesses concern coverage and enforcement. The directive applies to defined federal administrative decisions, not all government AI. Compliance responsibility sits within departments under central policy oversight rather than an independent AI regulator with general suspension and compensation powers. Public sources document requirements more clearly than they document whether affected people routinely obtain timely correction or whether non-compliant deployments are stopped.

Outcome: a credible preventive architecture with partial Protection Reach. The directive documents recourse requirements and mandatory governance artefacts sufficiently to support a remedy score of 2 for the formal pathway, but public sources do not establish routine completion, compensation or suspension outcomes. That uncertainty is carried in a low remedy-outcome confidence rating rather than converted into a nominal score. The case supports P1 by showing that substantial implementation activity does not establish lived protection, and P2 because the register and impact assessments improve visibility and governance evidence. Confidence: high for formal requirements; low for routine remedy outcomes.

Table 4. Comparative diagnostic matrix for six algorithmic cases (operative period)

Case	Vis.	Coverage	Evidence	Contest.	Authority	Remedy	Confidence
Australia - Robodebt	1	2	1	1	1	1	H*
Netherlands - SyRI	1	3	2	3	3	3	H
Kenya - NIIMS	2	2	2	3	3	2	H*
India - Aadhaar	2	2	1	2	2	2	H/M*
Spain - platform work	2	2	2	2	2	2	H/M
Canada - federal ADM	3	2	3	2	2	2	H/L*

Confidence key: H = high confidence in legal or institutional evidence; M = moderate confidence in implementation evidence; L = low confidence in routine outcome evidence; * = relevant lived or post-period outcomes remain unobserved.

Scale: 0 absent; 1 nominal; 2 partial; 3 credible. Integer scores describe the defined arrangement during the operative period. Later correction and level-specific qualifications appear in the case notes. Black Sash is an unscored institutional analogue.

8. Comparative Findings

8.1 Recurrent failure points

Across the six algorithmic cases, the most persistent bottleneck is not the absence of a high-level principle but the failure to connect evidence, representation and authority. Robodebt recipients carried an unrealistic evidentiary burden. Spanish representatives possess an information right whose usefulness depends on technical specificity and employment structure. Canada publishes risk assessments but offers limited public evidence of completed individual remedies. Kenya's court required a regulatory framework, while Aadhaar's constitutional review left routine service-denial correction distributed across multiple actors. In each case, a formal element existed while an adjacent element constrained reach.

A second failure point is responsibility fragmentation. Public bodies can attribute opacity to vendors, vendors to contractual limits, integrators to provider documentation, and frontline officials to the system's recommendation. Spain illustrates labour-market fragmentation. Aadhaar provides the clearer public-sector example: UIDAI, an authentication user agency, a device operator and the service-delivery body may each hold a different part of the transaction, exception and grievance record, while no single actor necessarily owns end-to-end correction. The Black Sash analogue shows a potential institutional response - binding public and private actors to a shared outcome and common reporting duties - but does not itself establish that model-specific evidence would be available in an AI case. Protection Reach improves when law assigns joint or clearly sequenced responsibility and prevents contracts or agency boundaries from displacing public duties.

A third failure point is the individualisation of challenge. Robodebt's ordinary process expected recipients to reconstruct records. SyRI and NIIMS produced broader system-level scrutiny because representative organisations aggregated affected interests and challenged the architecture. Spain shows the limit: collective representation can create a route to information without guaranteeing that the information is complete, technically usable or enforceable across subcontracting chains. Collective contestability changes the object of review, but it does not replace evidence and authority.

8.2 Strong and weak forms of reach

Strong reach combines four conditions: an identifiable forum, access to usable evidence, authority over responsible actors, and a remedy designed around the operational harm. Among the algorithmic cases, SyRI provides the strongest evidence of system-level contestability and authority, although not individual compensation. Robodebt demonstrates strong retrospective accountability but weak contemporaneous reach. Canada demonstrates preventive visibility and governance records without equivalent evidence of routine remedial completion. Spain demonstrates collective information rights with incomplete evidence usability. Kenya and Aadhaar demonstrate authoritative constitutional review whose everyday corrective pathways remain implementation-dependent. Black Sash illustrates how standing, supervision and continuity-focused relief can be designed, but it remains an analogue outside the scored comparison.

The evidence is strongest for P1 and P4: framework or implementation presence does not predict completed protection, while authority and remedy timing repeatedly distinguish nominal from usable pathways. P2 is supported by Robodebt, Aadhaar, Spain and Canada, but collective organisations can sometimes create visibility where individual notice is weak. P3 receives conditional support, not confirmation: SyRI shows the value of representation linked to an authoritative court, while Kenya, Aadhaar and Spain show that representation alone can leave evidence, implementation or remedial gaps. Participation timing and responsibility fragmentation operate as cross-cutting scope conditions rather than separate propositions. These are theory-building propositions, not statistically tested hypotheses.

8.3 Negative tests and selection effects

Kenya and Aadhaar are the principal negative tests of collective contestability. In both, organised or multi-party constitutional challenges reached authoritative courts and produced meaningful constraints, yet neither judgment itself created a complete routine mechanism for correcting biometric failure, reconstructing a service denial or compelling all operational actors. Spain provides a third limiting case: workers' representatives possess a statutory information right, but fragmented employment and incomplete technical disclosure can still weaken reach. The finding is therefore conditional: collective intermediaries reduce individual burdens, but they require evidence access, remedial ownership and an authorised forum.

This reframing reduces, but does not remove, selection bias. The sample remains weighted toward cases that generated litigation, commissions or published governance records. It cannot estimate how often organised civil society fails, how many systems remain unchallenged because they are invisible, or whether quieter administrative pathways produce better remedies. Those questions require unsuccessful-challenge cases, random deployment samples and longitudinal observation.

8.4 Public-sector implications

Public-sector deployments deserve a higher, not lower, standard because they exercise legal authority and often control essential services. GIRAI reports that only 18 per cent of countries assessed had frameworks requiring public disclosure of government algorithmic systems, and only 26 per cent had frameworks addressing fair and accountable public procurement of AI. [1] The OECD's 2026 findings likewise show that most surveyed governments lack both transparency standards and open registers. [2] These gaps matter because people cannot choose another welfare agency, border authority or identity system. Public power also creates a responsibility asymmetry: the state can compel information from citizens while withholding the operational basis of its own automated scrutiny.

8.5 Global South conditions that alter the pathway

The Global South cases show that Protection Reach may be structured differently, not simply scored lower. Kenya, India and the Black Sash analogue demonstrate legal imagination and authoritative review, while also revealing bottlenecks created by informality, identity and connectivity gaps, distributed service delivery and limited routine evidence access. This moves the argument beyond a deficit narrative: the relevant question is which institutional assumptions built into contestability do not hold in the deployment context.

Informality changes legal coverage. ILOSTAT estimated that 57.9 per cent of the world's employed population was in informal employment in 2025; the corresponding rate for Central and Southern Asia was 83.9 per cent. [41] A contestability model built around a recognised employer, contract, personnel file and trade-union channel may not reach informal workers, self-employed people or workers whose status is disputed. Coverage assessment must therefore ask who can represent a person without a formal legal relationship, which entity can be compelled, and whether community or sector bodies have standing.

Digital exclusion is both a harm-generation mechanism and a remedy-access problem. Aadhaar and NIIMS show how poor biometrics, weak connectivity or missing documents can prevent a person from entering the service pathway at all. In such cases there may be no conventional adverse-decision notice, reason or appealable record. The six dimensions should therefore record alternative verification, assisted and offline access, exception logs and a route to restore the service without requiring the person to solve the same digital failure that caused the exclusion.

Imported and distributed systems create a second pattern. A public agency may deploy a vendor-controlled system under a contract that places model documentation, logs or technical staff outside domestic reach. This is where Protection Reach and regulatory sovereignty meet: provider-facing leverage is upstream, but its absence becomes a downstream evidence and remedy failure for the affected person. The concepts remain distinct, yet procurement terms, local representation, evidence export, version preservation and cross-border assistance are necessary connecting mechanisms. [34]

Regional institutions can supply forms of institutional inheritance where domestic courts or regulators are weak. The African Commission has placed AI and emerging technologies within its human-rights mandate and called for

compatibility with dignity, privacy, equality, inclusion, transparency and accountability. [42] The African Union Convention on Cyber Security and Personal Data Protection entered into force on 8 June 2023 after reaching the 15 ratifications required by Article 36; its practical reach remains limited to states that have ratified and implemented it. [43] These institutions do not automatically deliver individual remedies, but they can support model laws, interpretive standards, complaints, peer pressure and shared technical capacity. Local-language and community-level dispute mechanisms should likewise be assessed for actual authority, inclusion and safeguards rather than assumed to be either inherently accessible or inherently inadequate.

International assistance should consequently finance the full pathway rather than the strategy alone: multi-year regulatory staff, legal aid, union and community representation, evidence infrastructure, local-language and offline complaints, exception handling, regional referral and contract enforcement after a donor- or vendor-supported system is deployed.

8.6 A qualified central proposition

The comparative evidence supports a qualified proposition: governance should be assessed not only by whether a rule exists, but by whether an affected person or representative can activate it. Activation requires notice, coverage, evidence, an authorised forum and timely remedy. These conditions are not sufficient for substantive justice - a lawful process can still produce contested policy outcomes - but their absence reliably weakens the practical force of governance.

Table 5. Recurrent failure points and consequences

Failure point	Consequence	Leverage point
System not disclosed	People challenge an outcome without knowing the automated role; low complaint volume appears to signal low harm.	Mandatory register plus decision-specific notice.
Population or harm outside mandate	Platform, informal, migrant or indirect harms fall between labour, data and sector regulators.	Coverage map and joint responsibility rules.
Evidence not preserved or controlled by vendor	Review cannot reconstruct the operative system, threshold, override or data.	Retention protocol and compulsory tiered access.
Individualised complaint burden	High cost, fear and technical complexity suppress challenges and precedent.	Collective standing, legal aid and public-interest technology.
Oversight body advisory only	Findings do not change deployment or provide relief.	Independent budget, inspection and corrective powers.
Non-suspensive or delayed review	Legal success follows irreversible loss of benefit, work, status or trust.	Interim relief and statutory remedy timelines.
Fragmented responsibility	Provider, integrator, contractor and authority displace blame and records.	Named lead deployer and joint/sequenced liability.
Participation after lock-in	Consultation improves legitimacy but not requirements, architecture or contracts.	Participation at problem definition and pre-procurement stages.
Authentication or documentation failure prevents entry	No notice or decision record is created; exclusion appears as non-use or non-eligibility.	Alternative verification, assisted/offline access, exception logging and immediate service restoration.
Evidence or technical responsibility held offshore	Domestic forum cannot compel logs, preserve the operative version or bind the vendor to a remedy.	Named local representative, evidence-export clauses, escrow/mirroring where proportionate and cross-border cooperation.

9. Protection Reach Diagnostic

9.1 Scoring logic

The diagnostic uses one integer score from 0 to 3 for each dimension and records participation timing separately. It is designed for case assessment, procurement review, regulatory audits and comparative research. Users should cite evidence for every score, identify the relevant period and state confidence. Ranges, arrows and hybrid labels are not allowed in score cells. Temporal change, decision-level versus system-level reach and later correction belong in accompanying notes. A score of 3 requires credible operational use, not only a legal text; where public evidence is absent, confidence is reduced

rather than a midpoint inferred. The dimensions should not be summed into a country league table because a total can conceal a decisive bottleneck.

9.2 Minimum evidence pack

A credible assessment should include: system identity and responsible entities; individual notice and service-entry or exception records; the applicable legal basis and affected-population map; impact assessments and procurement records; data and decision-record retention rules; system/version logs; complaint and appeal procedures, including offline and representative routes; powers and resources of the reviewing body; remedy timelines and outcomes; cross-border evidence and local-representation arrangements; and participation records showing what decisions remained open. Where evidence is unavailable, the absence should be coded as uncertainty rather than assumed compliance.

9.3 Interpretation

The matrix captures the six algorithmic cases during the period most relevant to the alleged harm, while case notes identify later correction. It shows why disaggregation matters. Robodebt's later settlement does not erase weak visibility and evidence access during operation. Canada's strong visibility architecture does not prove strong compensation or suspension outcomes; its remedy score reflects documented recourse provisions and carries low outcome confidence. Aadhaar's high visibility and constitutional review do not establish that every authentication-related denial produces an obtainable record and timely restoration. Black Sash remains excluded because its institutional lessons are transferable but its evidence problem was not algorithmic.

Table 6. Protection Reach diagnostic scale

Score	Interpretation
0 - Absent	No meaningful right, institution or operational mechanism is evidenced.
1 - Nominal	A formal commitment or right exists, but affected people cannot realistically activate it.
2 - Partial	Usable mechanisms exist for selected actors or cases, but serious accessibility, authority or remedy gaps remain.
3 - Credible	Affected people or representatives can normally identify the system, access challenge, trigger authoritative investigation and obtain meaningful remedy.

10. Global Policy Implications

10.1 Move from framework counts to pathway indicators

GIRAI and other global indexes should retain framework and implementation indicators while treating Protection Reach as a complementary deployment-level diagnostic. Country comparison can identify where institutional scarcity is concentrated; pathway assessment can then locate the binding constraint in a particular system. This is not a competing league table. It is a second-stage method for testing whether the commitments and institutions counted at national level complete for affected people. [1]

At minimum, measurement should ask whether public bodies maintain complete system inventories; whether individual notice or an exclusion record links to an accountable entity; whether records are preserved and exportable; whether representative organisations have standing; whether a forum can compel domestic and offshore evidence and suspend use; and whether remedy outcomes and timelines are reported. The unit should be a documented deployment or domain, not a national claim of principle.

10.2 Establish minimum contestability standards

The Council of Europe Convention offers a useful baseline: relevant documentation sufficient to challenge both decisions and system use, effective complaints and procedural safeguards. [3] Minimum standards should add practical elements: accessible local-language notice; correction of input data; reasoned human reconsideration by a person able to depart from the system; preservation of the operative model and logs; protection against retaliation; and suspensive effect where continued operation risks irreversible harm. The EU AI Act's complaint and explanation rights provide important components but should be read alongside sectoral administrative, labour and equality law rather than treated as a complete remedy architecture. [4]

10.3 Treat collective rights as core infrastructure

Collective rights are not an optional civil-society add-on. They respond to information and power asymmetry. Indicators should record whether unions, consumer bodies, equality institutions and public-interest organisations can obtain information, represent affected groups, initiate system-level challenges and participate in ongoing monitoring. SyRI, NIIMS and Spain show different forms of collective reach and its limits: representation is most effective when it is connected to usable evidence and an authoritative escalation route.

10.4 Apply higher transparency to public authority

Public-sector algorithm registers should be mandatory, updated and linked to specific services and challenge routes. Registers should identify the deployer, provider, purpose, decision role, affected population, legal basis, impact assessment, version/change history, incident history and responsible contact. Security and fraud-prevention exceptions may justify withholding defined details, but not the existence of the system, the accountable authority or the route to challenge. GIRAI's 18 per cent public-disclosure finding indicates that this remains a global gap. [1]

10.5 Finance institutions, not only strategies

International development finance and technical assistance should include line items for independent oversight, evidence infrastructure, legal aid, worker organisations, public-interest technology and multilingual complaint systems. Projects should specify who funds these functions after the pilot phase. A strategy without recurring institutional finance can increase the appearance of governance while leaving the pathway incomplete.

11. Policy Recommendations

Recommendation 1 - Create a dual visibility duty: public register plus individual notice.

Responsible actors: legislatures, central digital-government offices and deploying bodies. Authority: a binding disclosure duty covering public and contracted systems. Mechanism: a machine-readable register linked to service-specific notice at or before the first consequential decision. Evidence: system purpose, decision role, responsible entities, provider, legal basis, impact assessment, change history and challenge route. Resources: a central registry team and departmental data stewards. Timeline: register existing high-impact systems within 12 months; notice duties for new systems before production. Symbolic-compliance risk: generic entries that do not identify the service or decision. Success indicator: proportion of deployed systems with complete entries and notices that affected people can connect to a real complaint path.

Recommendation 2 - Establish an evidence-preservation and disclosure protocol.

Responsible actors: sector regulators, archives authorities, courts and public procurement bodies. Authority: power to require retention and compel production from deployers, contractors and reachable provider representatives. Mechanism: preserve input and exception data, decision records, reasons, model or rule version, logs, overrides, audit findings and incidents; for imported or hosted systems, require evidence export in open formats, a named local representative, regulator access and proportionate escrow or mirrored records. Evidence: a complete record sufficient to reconstruct the operative decision and system. Resources: secure repositories, technical investigators and cross-border legal support. Timeline: standard protocol within 18 months and mandatory clauses for new procurements immediately. Symbolic-compliance risk: generic logging language without tested retrieval, or contractual rights that cannot be enforced across borders. Success indicator: proportion of sampled decisions reconstructed end to end within a fixed time.

Recommendation 3 - Give contestation suspensive effect where harm is difficult to reverse.

Responsible actors: legislatures and sector agencies. Authority: a statutory power to pause debt recovery, benefit termination, deactivation, deportation, identity denial or another difficult-to-reverse action while a credible challenge is reviewed. Mechanism: automatic or rapid interim relief triggered by harm severity, reversibility and an initial showing of error, with narrowly defined emergency and public-safety exceptions. Evidence: notice of the automated role, grounds for challenge and immediate record preservation. Resources: triage teams, expedited review capacity and service-continuity budgets. Timeline: adopt rules within 12 months in priority sectors. Symbolic-compliance risk: exceptions so broad, thresholds so demanding or review so slow that suspension is rarely granted. Success indicator: median time to interim decision and proportion of successful appeals in which harm was prevented rather than compensated later.

Counterargument and design constraint. Suspensive effect can delay fraud control, create fiscal exposure, prolong unsafe conduct or encourage strategic appeals. A universal automatic stay is therefore neither necessary nor proportionate. The appropriate design is risk-tiered: preserve essential subsistence or legal status while review occurs; allow recovery or restriction where an independent decision-maker records a concrete and urgent risk; require expedited determinations; and

monitor false-positive, fiscal and recurrence costs. The relevant comparison is not between suspension and cost-free enforcement, but between the cost of interim protection and the often irreversible cost of acting on an erroneous automated decision.

Recommendation 4 - Build collective contestability and representative standing.

Responsible actors: labour ministries, consumer and equality regulators, courts and legislatures. Authority: standing for unions, civil-society organisations, ombuds institutions, legal-aid bodies and recognised community or sector representatives to obtain information and challenge systems. Mechanism: representative complaints, aggregate evidence requests, anti-retaliation protection and system-level remedies, including for informal workers and people excluded before a formal decision record exists. Evidence: population patterns, complaint clusters, exception logs and distributional outcomes. Resources: legal aid, technical experts and sustained public-interest funding. Timeline: legislative or procedural reform within 18 months. Symbolic-compliance risk: formal standing without compulsory evidence or an authorised forum. Success indicator: representative cases producing system-level correction rather than isolated reconsideration.

Recommendation 5 - Separate oversight from deployment and give it corrective authority.

Responsible actors: governments and parliaments, supported where appropriate by regional human-rights and data-protection institutions. Authority: independent budgets, protected appointments, inspection and compulsory evidence powers, authority over public-sector systems, and powers to order correction, restriction or suspension. Mechanism: separate deployment from review, designate a lead authority for fragmented systems and create regional referral or pooled-investigation arrangements where national technical reach is limited. Evidence: technical, contractual and operational records. Resources: permanent multidisciplinary teams and shared specialist capacity. Timeline: mandate and budget within 24 months. Symbolic-compliance risk: coordination bodies that can request but not compel, or regional commitments without domestic execution. Success indicator: time to secure evidence, implementation rate of orders and number of cross-actor failures resolved without burdening the affected person.

Recommendation 6 - Convert participation into decision influence before lock-in.

Responsible actors: procuring authorities and deploying organisations. Authority: procurement and impact-assessment rules requiring affected-group participation at problem definition, pre-procurement and design stages. Mechanism: publish the problem statement, alternatives to automation, draft requirements and non-negotiable constraints; record which proposals changed the decision; establish continuing representation after deployment. Evidence: participation log, response matrix, conflict-of-interest disclosure and decision rationale. Resources: compensation, accessibility, translation and independent facilitation. Timeline: mandatory for new high-impact procurement immediately; retrospective review at major contract renewal. Symbolic-compliance risk: consultation after vendor selection. Success indicator: proportion of engagements held before procurement and documented changes attributable to participants.

Recommendation 7 - Publish remedy performance, not only compliance artefacts.

Responsible actors: oversight bodies and deployers. Authority: reporting duties covering complaints, review outcomes and systemic corrections. Mechanism: publish anonymised statistics on notice, complaints, interim relief, evidence access, reversals, compensation, system changes and timelines, disaggregated where lawful and safe. Evidence: case-management records and independent sampling. Resources: standardised reporting systems and audit capacity. Timeline: annual reporting within 12 months. Symbolic-compliance risk: reporting complaint counts without denominator, outcome or delay. Success indicator: median remedy time, recurrence rates and percentage of successful cases producing systemic correction.

Recommendation 8 - Attach a Protection Reach budget to AI investment.

Responsible actors: finance ministries, development banks, donors and programme owners. Authority: funding and procurement conditions for high-impact public AI and automated systems. Mechanism: require a costed plan for oversight staff, records, public registers, exception handling, offline and local-language complaints, independent evaluation, legal aid, representative organisations, regional referral and post-contract evidence access. Evidence: staffing and recurring budgets, vendor obligations, local representation, exit and evidence-export clauses, and remedy service levels. Resources: lifecycle funding rather than one-off technical assistance. Timeline: one budget cycle and immediate application to new donor-funded deployments. Symbolic-compliance risk: financing the model and training while leaving oversight and remedies unfunded. Success indicator: essential Protection Reach functions sustained three years after deployment and tested during an incident or complaint.

Table 7. Policy intervention matrix

Intervention	Responsible actor	Mechanism	Evidence	Timeline	Success indicator
Dual visibility	Legislature / central digital office	Register + individual notice	System, role, entities, impact assessment, challenge route	12 months	Complete entries linked to notices
Evidence protocol	Regulators / procurement / archives	Retention + compulsory tiered access + cross-border evidence/export clauses	Data, reasons, version, logs, exceptions, overrides, audits and reachable evidence holder	18 months	End-to-end reconstruction rate
Suspensive challenge	Legislature / sector agency	Risk-tiered pause for difficult-to-reverse action; expedited review; narrow emergency exceptions	Initial grounds, harm severity and preserved operative record	12 months	Harm prevented, interim-decision time and fiscal/error monitoring
Collective standing	Legislature / courts / labour bodies	Representative complaints, including informal/community and exclusion cases	Aggregate patterns, exception logs and distributional outcomes	18 months	System-level corrective actions
Independent authority	Parliament / government	Inspection, access, restriction, suspension and regional referral	Technical and contractual evidence	24 months	Time to evidence and implementation of orders
Early participation	Procuring authority	Problem-definition and pre-procurement participation	Alternatives, requirements, response matrix	Immediate for new systems	Documented changes before contract
Remedy reporting	Deployers / oversight bodies	Outcome and timeline publication	Complaints, interim relief, reversals, recurrence	Annual	Median remedy time and recurrence
Reach budget	Finance ministries / development funders	Lifecycle funding and donor/vendor governance conditions	Staffing, exception handling, local access, evidence export and remedy service levels	One budget cycle	Functions sustained after three years

12. Limitations and Research Agenda

The study relies on public evidence and uses purposive, outcome-rich case selection. Litigation, commissions, judgments, audits and registers make institutional pathways visible; quiet deployments and unsuccessful challenges are harder to observe. This creates transparency bias and selection on the dependent variable. Kenya and Aadhaar provide hard negative tests of collective constitutional mobilisation, and Spain limits the collective-information claim, but the correction remains incomplete. Low complaint counts may reflect low harm, low visibility, fear, digital exclusion, informality or distrust. The diagnostic cannot resolve that ambiguity without interviews, administrative data, unsuccessful-case records and longitudinal observation.

The cases also vary by sector and legal tradition. A constitutional court's authority over digital identity is not directly comparable to a workplace information right or a central-government directive. Black Sash is a bounded analogue rather than an algorithmic case and remains excluded from scoring. Adding India improves Asian and Global South representation, but Latin America and the Middle East remain absent, and no six-case comparison can represent the institutional diversity within the Global South. Scores describe defined governance arrangements, not general national capacity. The analysis identifies plausible mechanisms but does not establish causal effects through statistical comparison.

Further research should test Protection Reach through practitioner and affected-person interviews; audit random samples from service entry and authentication through notice, evidence, appeal and remedy; compare systems before and after registers or suspensive review; and deliberately select cases in which unions, community organisations or civil-society groups mobilised but failed to obtain evidence, standing or correction. Priority comparisons include Latin American data-protection and public-sector systems, Middle Eastern digital-government deployments, donor- and vendor-imported

systems, and regional complaint pathways. Research should examine customary and community dispute institutions without presuming that they are either accessible substitutes or rights-compatible forums, and should track whether successful cases produce durable institutional change.

13. Conclusion

AI governance becomes meaningful at the point where a person can use it. Frameworks, laws and impact assessments remain necessary, but they are promises until institutions connect them to notice, evidence, contestation, authority and remedy. Protection Reach names and measures that connection.

The comparative cases show that the strongest practical protection does not necessarily come from the newest AI instrument. It often comes from institutional inheritance: constitutional standing, administrative review, labour representation, judicial authority, ombuds functions and public-interest litigation. SyRI shows these institutions reaching an algorithmic system directly; Kenya and Aadhaar show that access to authoritative courts can still leave everyday exclusion and correction dependent on implementation; Spain shows that collective information can be weakened by fragmented responsibility. Black Sash supplies a bounded analogue for standing, supervision and time-sensitive relief, not proof of algorithmic evidence access. Dedicated AI governance adds value when it strengthens these pathways rather than creating parallel documentation without an accessible forum.

The central policy implication is therefore precise. Governments and global institutions should evaluate not only whether a rule exists or an activity has occurred, but whether an affected person or representative can activate a complete pathway. In many Global South settings that pathway must also reach informal relationships, record exclusion before a conventional decision exists, compel evidence held across borders, work in local languages and connect domestic institutions to regional authority. GIRAI identifies the global gap between commitment and durable institutions; Protection Reach is the deployment-level instrument for finding where that gap becomes a bottleneck for a person. [1]

Appendix A. Protection Reach coding guide

Visibility

- 0 - No evidence of notice, inventory, accountable contact or record that a person was excluded at system entry.
- 1 - General disclosure or policy statement, but affected people cannot connect it to a decision or failed authentication.
- 2 - System-level register, notice or exception record exists, with gaps in timing, completeness or linkage.
- 3 - Timely, accessible notice or exception record identifies the system, role, responsible entities and challenge route.

Legal and institutional coverage

- 0 - Relevant people, informal relationships, exclusion harms or responsible actors are outside applicable law or mandate.
- 1 - Formal right exists but status, documentation, jurisdiction or responsibility gaps make it difficult to invoke.
- 2 - Core populations and actors are covered, but indirect, community, informal or cross-border gaps remain.
- 3 - Coverage includes affected groups and allocates responsibility across deployers, providers, contractors and public bodies.

Evidence accessibility

- 0 - Relevant decision, authentication or exception records are not preserved or obtainable.
- 1 - Reasons or data can theoretically be requested, but evidence is incomplete, delayed, distributed or beyond domestic reach.
- 2 - Key decision and governance records are obtainable by selected actors, with gaps in technical depth, timeliness or export.
- 3 - Affected people, representatives and authorised institutions can obtain usable personal, operational, governance and structural evidence, including from hosted or imported systems.

Contestability

- 0 - No meaningful complaint, appeal, review or route for a person excluded before a conventional decision.
- 1 - Formal process exists but is unaffordable, digital-only, non-suspensive, retaliatory or limited to output correction.
- 2 - Usable individual or collective processes exist for selected actors, but offline, local-language, informal-worker or system-level gaps remain.
- 3 - Accessible individual and representative routes can challenge the decision, process and deployment with appropriate interim protection.

Institutional authority and independence

- 0 - No identifiable forum.
- 1 - Advisory or internally dependent body without compulsory access or corrective power.
- 2 - Forum can investigate and recommend or order limited correction, but lacks independence, resources or authority over all actors.
- 3 - Independent, resourced forum can compel evidence, coordinate, suspend or restrict systems and publish reasoned outcomes.

Remedy effectiveness

- 0 - No demonstrated correction or relief.
- 1 - Recognition, explanation or review without timely correction.
- 2 - Individual reversal, correction or compensation is available but delayed, incomplete or weakly systemic.
- 3 - Harm can be prevented or stopped, decisions and data corrected, compensation delivered where appropriate, and recurrence addressed within defined timelines.

Appendix B. Lexicon-based case review template

1. Verified facts: What is established by primary evidence?
2. Institutional interpretation: How do the court, regulator, government or author characterise the issue?
3. Inference: What mechanism is inferred, and with what confidence?
4. Decision authority: Who selected, approved, operated, reviewed and could stop the system?
5. Affected but absent actors: Which workers, communities or service users lacked representation?
6. Governance timing: When did oversight appear relative to problem definition, procurement, design and deployment?
7. Evidence ownership: Who held personal, operational, governance and structural evidence?
8. Operational constraints: Workload, budgets, contracts, technical dependencies, languages and access barriers.
9. Contradictions: Transparency/security; speed/due process; innovation/protection; individual/systemic remedy.
10. Feedback loops: Which mechanisms suppressed complaints, evidence or institutional investment?
11. Leverage points: What intervention could change more than one Protection Reach dimension?
12. Confidence and missing evidence: What cannot be established from the public record?

References

1. Adams, R., Adeleke, F., Alayande, A., Abdella, S.E., Florido, A., Junck, L. and Grossman, N. (2026). Global Index on Responsible AI 2026, 2nd ed. Global Center on AI Governance. Data period: 1 November 2023–30 September 2025.
2. OECD (2026). Digital Government Outlook 2026: Adopting and Governing AI in Government. Paris: OECD Publishing, June 2026.
3. Council of Europe (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, CETS No. 225, opened for signature 5 September 2024; Council of Europe explanatory materials on remedies and procedural safeguards.
4. European Union (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, especially Articles 85–86.
5. United Nations Human Rights Council (2011). Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework, UN Doc. A/HRC/17/31.
6. Citron, D.K. (2008). Technological Due Process. *Washington University Law Review*, 85(6), 1249–1313.
7. Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G. and Yu, H. (2017). Accountable Algorithms. *University of Pennsylvania Law Review*, 165(3), 633–705.
8. Kaminski, M.E. and Urban, J.M. (2021). The Right to Contest AI. *Columbia Law Review*, 121(7), 1957–2048.
9. Cobbe, J., Lee, M.S.A. and Singh, J. (2021). Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems. *Proceedings of ACM FAccT 2021*.
10. Ananny, M. and Crawford, K. (2018). Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society*, 20(3), 973–989. doi:10.1177/1461444816676645.
11. Wieringa, M. (2020). What to Account for When Accounting for Algorithms: A Systematic Literature Review on Algorithmic Accountability. *Proceedings of FAT* 2020*, 1–18.
12. Koulu, R. (2020). Proceduralizing Control and Discretion: Human Oversight in Artificial Intelligence Policy. *Maastricht Journal of European and Comparative Law*, 27(6), 720–735. doi:10.1177/1023263X20978649.
13. Green, B. (2022). The Flaws of Policies Requiring Human Oversight of Government Algorithms. *Computer Law & Security Review*, 45, 105681.
14. Treasury Board of Canada Secretariat (2025–2026). Directive on Automated Decision-Making and Guide on the Scope of the Directive on Automated Decision-Making. Current official versions accessed 18 July 2026.
15. Treasury Board of Canada Secretariat (2026). Algorithmic Impact Assessment Tool: Instructions, Scoring and Publication Requirements. Updated 28 May 2026.
16. Government of Canada (2025–2026). Government of Canada AI Register (Minimum Viable Product), released 28 November 2025; dataset modified 28 April 2026.
17. Treasury Board of Canada Secretariat (2025). Guide to Peer Review of Automated Decision Systems.
18. District Court of The Hague (2020). NJCM et al. v State of the Netherlands (SyRI), ECLI:NL:RBDHA:2020:865, 5 February 2020 (English version ECLI:NL:RBDHA:2020:1878).
19. High Court of Kenya (2020). Nubian Rights Forum & 2 others v Attorney General & 6 others; consolidated Petitions 56, 58 and 59 of 2019, [2020] KEHC 8772 (KLR), judgment of 30 January 2020.
20. Constitutional Court of South Africa (2017). Black Sash Trust v Minister of Social Development and Others, [2017] ZACC 8, judgment of 17 March 2017.
21. Constitutional Court of South Africa (2026). Black Sash Trust and Another v Minister of Social Development and Others, [2026] ZACC 12, judgment of 8 April 2026.
22. Royal Commission into the Robodebt Scheme (2023). Report of the Royal Commission into the Robodebt Scheme, 7 July 2023, especially Recommendation 18.1.
23. Parliament of Australia, Senate Community Affairs References Committee (2022). Accountability and Justice: Why We Need a Royal Commission into Robodebt; and final/interim reports on Centrelink’s compliance programme.
24. Services Australia (2025). Robodebt Refunds, official guidance, updated 6 January 2025.
25. Services Australia (2026). New Robodebt Class Action Settlement, Federal Court approval 23 June 2026; page updated 26 June 2026.
26. Spain (2021). Ley 12/2021, de 28 de septiembre, amending the Workers’ Statute to guarantee labour rights of persons engaged in delivery through digital platforms, BOE-A-2021-15767.
27. International Labour Organization (2026). Decent Work in the Platform Economy Convention, 2026 (No. 193), adopted 12 June 2026; NORMLEX instrument profile reporting 0 ratifications as of the 18 July 2026 research cut-off.
28. International Labour Organization (2023). Realizing the Opportunities of the Platform Economy through Freedom of Association and Collective Bargaining, ILO Working Paper 80.
29. Vieira, T. (2025). An Incomplete Double Movement: Spain’s Legislative Strategy for Platform Courier Reclassification. *International Labour Review*, 164(4).
30. Fairwork (2025). Fairwork Spain Ratings 2025: Insufficient Progress in Protecting Work on Digital Platforms.
31. Wachter, S., Mittelstadt, B. and Russell, C. (2018). Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887.
32. Malgieri, G. and Comandé, G. (2017). Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation. *International Data Privacy Law*, 7(4), 243–265.
33. Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D. and Barnes, P. (2020). Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. *Proceedings of FAT* 2020*.
34. Arora, G. (2026). The Regulatory Sovereignty Gap: Technological Dependence, Bargaining Power and the Bindingness of AI Governance in the Global South, Version 2.0. Institute for Technology Stewardship.

35. Council of Europe (2025-2026). HUDERIA Methodology and Model: Context-Based Risk Analysis (COBRA), approved 26 February 2025 and 25 February 2026 respectively.
36. Arnstein, S.R. (1969). A Ladder of Citizen Participation. *Journal of the American Institute of Planners*, 35(4), 216-224.
37. Lipsky, M. (1980). *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.
38. Supreme Court of India (2018). Justice K.S. Puttaswamy (Retd.) & Anr. v. Union of India & Ors., W.P. (C) No. 494 of 2012, judgment of 26 September 2018, [2018] 8 S.C.R. 1.
39. Comptroller and Auditor General of India (2022). Performance Audit on Functioning of Unique Identification Authority of India, Report No. 24 of 2021, especially Chapter 6: Redressal of Customer Grievances; tabled 6 April 2022.
- 40a. Unique Identification Authority of India (2026). Grievance Redressal Mechanism. Official webpage, accessed 18 July 2026: <https://www.uidai.gov.in/en/contact-support/have-any-question/313-english-uk/faqs/about-uidai/grievance-redressal-mechanism.html>
- 40b. Unique Identification Authority of India (2026). Authentication Devices & Documents - exception handling and no-denial-of-service provisions. Official webpage, accessed 18 July 2026: <https://uidai.gov.in/en/ecosystem/authentication-devices-documents/>
- 40c. Unique Identification Authority of India (2026). Authentication FAQ - alternative processes where authentication fails. Official webpage, accessed 18 July 2026: <https://uidai.gov.in/en/contact-support/have-any-question/303-faqs/authentication.html>
41. International Labour Organization, ILOSTAT (2026). The world of work and the 2030 Agenda: a ten-year review, reporting 2025 informal-employment estimates; see also the ILOSTAT informal-employment database. Accessed 18 July 2026.
42. African Commission on Human and Peoples' Rights (2021). Resolution on the Need to Undertake a Study on Human and Peoples' Rights and Artificial Intelligence, Robotics and Other New and Emerging Technologies in Africa, ACHPR/Res. 473 (EXT.OS/XXXI) 2021, adopted 25 February 2021.
43. African Union (2014/2023). African Union Convention on Cyber Security and Personal Data Protection (Malabo Convention), adopted 27 June 2014; Article 36 required 15 ratifications for entry into force; entered into force 8 June 2023. African Union treaty text and status materials.

Declarations

Funding: No external funding was received. Competing interests: The author declares no competing interests. Data availability: No original dataset was generated; the analysis is based on publicly available sources cited in the paper. Ethical approval: Not applicable.

Suggested citation

Arora, G. (2026). From Promise to Protection: The Institutional Reach of AI Governance - Oversight, Contestability and Redress for People Affected by AI Systems. Institute for Technology Stewardship, Research Paper No. 2, Version 1.5.

Author note

Gaurav Arora is the founder of the Institute for Technology Stewardship. His research focuses on operational AI governance, critical infrastructure, regulatory institutions and the conditions under which governance commitments become enforceable protection.

Institute note

The Institute for Technology Stewardship is an independent initiative focused on governing autonomous and AI-enabled technologies in ways that preserve human judgment, public accountability and institutional resilience.

Revision note

Version 1.5 | Research cut-off: 18 July 2026. Final peer-review revision. Changes include a two-sentence Central Finding that combines chain logic with the institutional-inheritance result; corrected recommendation heading hierarchy; a manual list of tables and figures; separated UIDAI references; explicit rationale for the Aadhaar evidence score; a distinct responsibility-fragmentation implication under P2; use of Aadhaar in the fragmentation analysis; updated ILOSTAT sourcing; and clarification of the Malabo Convention ratification threshold.