Operationalizing Reparative Algorithmic Impact Assessments: Four Technical Artifacts for Healthcare AI Governance

Elise Eby Racine [1]

Abstract

Algorithmic impact assessments (AIAs) and fairness audits, as currently practiced, miss dimensions that prove central to AI deployments in low-resource healthcare contexts, particularly across the Global Majority. These include power asymmetries between developers, institutions, and affected communities; the political economy and sovereignty conditions of training data; the epistemic exclusion of non-Western knowledge systems; and the distribution of technical capacity between deployment partners as systems persist over time. This paper specifies a Reparative Algorithmic Impact Assessment (R-AIA) protocol that augments standard AIA tooling along these dimensions. Building on prior work (Racine, 2024; 2025), the protocol pairs a six-step evaluation framework with four classes of technical artifact: provenance audit trails, co-constructed impact taxonomies, community-defined evaluation rubrics, and capacity-transfer metrics. For the provenance artifact we give a concrete relational schema, example records, and an audit query. For the others, we provided structured specifications and worked examples. We then illustrate the protocol through a hypothetical maternal health system in Sub-Saharan Africa, analyzing its implementation burden, failure modes, and capture risks in the process. Taken together, we strive to offer a minimum viable instantiation for resource-constrained settings. The R-AIA gives technical AI governance a concrete instrument for assessing the structural conditions that determine whether such deployments serve or harm the populations they target.

[1]Ethox Centre and Institute for Ethics in AI, University of Oxford, Oxford, United Kingdom. Correspondence to: Elise Eby Racine <elise.racine@ndph.ox.ac.uk>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

1. The Assessment Gap

AI governance increasingly employs algorithmic impact assessments (AIAs) and related ex ante mechanisms to identify, evaluate, and mitigate risks throughout the AI lifecycle (Moss et al., 2021; Stahl et al., 2023; Reuel et al., 2024). Although implementation varies across jurisdictions, impact-assessment approaches feature in regulatory frameworks, including the EU AI Act's conformity assessment and fundamental-rights assessment requirements, the US NIST AI Risk Management Framework, and emerging AI legislation in jurisdictions such as South Korea and Brazil, as well as in private-sector governance. In healthcare specifically, the stakes for getting assessment right are immediate and life-critical, particularly as AI deployment in low-resource contexts accelerates despite persistent implementation challenges. Several recent deployments, surveyed in more detail in Racine (2025), illustrate the scale and the stakes:[1]

- **Clinical Decision Support:** A clinician copilot built by Penda Health and OpenAI on top of a Kenyan electronic-health-records system was associated with 16% fewer diagnostic errors and 13% fewer treatment errors across roughly 40,000 primary-care encounters (Korom et al., 2025).
- **Diagnostic Image Recognition:** In South Sudan, a tool trained on some 380,000 snake images has helped Médecins Sans Frontières staff identify species and choose antivenom, reportedly matching or beating specialist accuracy, in a setting where a wrong antivenom can cost a patient their yearly income (Ahmed, 2024).
- **Fetal Monitoring:** A National Health Ministry and Texas Children's Hospital partnership at Malawi's Area 25 Health Centre reported an 82% fall in stillbirths and newborn deaths over three years alongside use of PeriGen's AI-augmented fetal monitoring system (Kimeu, 2024).
- **Patient Retention:** An ARMMAN and Google DeepMind model flagged pregnant women at risk of drop-

[1]These examples illustrate a range of deployment settings rather than constituting a comprehensive survey.

ping out of an mHealth messaging program so that limited outreach can be aimed at them; a 100,000-woman pilot saw retention rise by 30% (Hu, 2025; Verma et al., 2024).

Yet current AIA practice, together with the technical fairness audit literature it draws on, systematically misses dimensions that determine whether such deployments harm or benefit the populations they target, particularly across the Global Majority. These include: (1) the power asymmetries that structure relationships between developers, institutions, and affected communities (Birhane, 2022; Davis et al., 2021); (2) the political economy and sovereignty conditions of the data on which these systems are trained and operated (Mohamed et al., 2020); (3) the epistemic exclusion of non-Western knowledge systems from the framing of what counts as legitimate evidence and meaningful outcome (Benjamin, 2019); and (4) the distribution of technical capacity between deployment partners as systems persist over time. The fourth is largely invisible to conventional AIA practice, where assessments are typically conducted once before or around deployment rather than as longitudinal evaluations of who can maintain, modify, and govern the system over time.

Standard fairness audits, such as disparate impact testing, subgroup accuracy gaps, and calibration analyses, are necessary but insufficient. They detect statistical disparities but, by design, do not evaluate the structural conditions that produce those disparities (Ashar et al., 2024). For instance, a fairness audit can certify that a maternal health AI system performs equivalently across patient demographics while remaining silent on whether traditional birth attendants retain decision authority during deployment, the consent terms governing the training data, or whether retraining capacity is concentrated in an external partner that may withdraw at any point. Each of these factors can determine whether the deployment ultimately strengthens or undermines the health system it enters.

The R-AIA extends a growing body of work on participatory and community-centred AI design, which argues both for involving affected communities in shaping systems and against treating participation as a procedural fix that launders decisions made elsewhere (Sloane et al., 2022; Delgado et al., 2023; Birhane et al., 2022; Costanza-Chock, 2020). Our contribution is to convert those commitments into auditable instrumentation: artifacts a third party can inspect, rather than engagement processes whose quality must simply be taken on trust. This situates the R-AIA within technical AI governance, which has mapped open problems in assessment, access, and verification (Reuel et al., 2024) but paid less attention to the structural conditions of deployment this protocol targets.

Building on prior work formalizing reparative impact assessment principles (Racine, 2024; 2025), this paper specifies a Reparative Algorithmic Impact Assessment (R-AIA) protocol that augments standard AIA workflows along these missing dimensions. We make three contributions:

1. A six-step assessment protocol with defined inputs, evaluation criteria, and indicators of misalignment for each step.

2. Four classes of technical artifact, namely provenance audit trails, co-constructed impact taxonomies, community-defined evaluation rubrics, and capacity-transfer metrics, that together form a layered assessment stack. For each we give a structured specification, and for the provenance artifact a concrete relational schema, example records, and an audit query.

3. An illustrative walk-through of a hypothetical integrated AI-powered maternal health system in Sub-Saharan Africa, together with an analysis of the protocol's implementation burden, failure modes, capture risks, and a minimum viable instantiation.

The protocol augments rather than replaces; standard fairness, safety, and performance audits remain necessary. What the R-AIA adds is the instrumentation required to make structurally-mediated harms visible, auditable, and addressable, and to make the redistribution of technical authority over the system a measurable outcome.

2. The R-AIA Protocol

The R-AIA protocol comprises six assessment steps. Each step has (i) a function, that is, what it is doing in the assessment, (ii) the activities that produce that function, and (iii) misalignment indicators that flag concrete failure modes. The protocol is sequential in logical structure but iterative in practice: findings surfaced in later steps routinely revise earlier ones, and assessors should expect to revisit earlier steps (for example, the socio-historical context audit and power analysis) as deployment unfolds.

Step 1, Socio-historical context audit. Healthcare AI systems do not enter neutral terrain. Prior technologies, including colonial medical infrastructures, pharmaceutical trial regimes, and mHealth pilots that arrived and departed, have shaped, and in many contexts disrupted, local healthcare practices, knowledge systems, and trust relationships. Step 1 documents that context before the system arrives. A collaborative research team including local information specialists, frontline practitioners, and, where relevant, medical anthropologists or historians, synthesises peer-reviewed literature alongside local knowledge absent from indexed sources.

Documented through: a research synthesis with named contributors and engagement methodology. Misalignment indicators: the synthesis cites only indexed, English-language sources; no named local contributor appears on the research team; local practices are logged as design constraints rather than knowledge inputs.

Step 2, Power analysis. AI deployments enter and reshape existing distributions of authority over data, clinical decisions, budgets, and the framing of what "success" looks like. Step 2 makes those dynamics explicit and establishes a governance mechanism that gives affected communities decision authority over the system. The analysis spans institutional, data, economic, technological, and narrative power; the governance mechanism is, where feasible, a community advisory body with consent and veto rights rather than consultation alone. Documented through: a power map and a governance instrument that establishes decision authority. Misalignment indicators: the power map omits informal and community authority; the governance instrument grants consultation but not veto rights; the map is produced once and never revised as authority shifts.

Step 3, Reparative data sovereignty. Training and operational data carry the conditions of their collection, including consent terms, sovereignty claims, benefit-sharing expectations, and jurisdictional restrictions. Standard data documentation captures composition; Step 3 captures these governance conditions and audits them against deployment practice. Where appropriate, this step may also establish a community-controlled data governance entity (a data trust or equivalent), and audits data representation across otherwise underrepresented populations. Documented through: a provenance audit trail (artifact a, Section 3) and the data governance instrument. Misalignment indicators: provenance records break at one or more hops; training assets draw overwhelmingly from urban or non-local sources; consent or benefit-sharing fields sit empty for sensitive assets; evaluation targets reference only biomedical metrics.

Step 4, Community-led impact co-construction. What counts as impact, that is, what counts as a benefit or a harm, is not given. It is defined by whoever is in the room when the assessment scope is set. Step 4 puts patients, frontline workers, traditional practitioners, and other relevant community members in that room with authority over the scope. Engagement protocols cover translation, fair compensation, accessibility, and explicit measures to mitigate consultation fatigue (Partnership on AI, 2024). Documented through: a co-developed impact taxonomy (artifact b, Section 3) and the engagement protocol that produced it. Misalignment indicators: engagement begins only after scope is fixed; participant lists skew to institutional experts; community-defined dimensions are dropped wherever they conflict with the existing technical plan.

Step 5, Ongoing monitoring with community-defined metrics. Deployment is not a single event but an ongoing co-evolution: systems shift behavior as they interact with their context, and contexts shift in response to the system's presence. Step 5 builds an evaluation rubric combining technical performance metrics with community-defined success criteria, alongside an adaptation protocol that specifies what triggers system changes and who has authority to make them. Documented through: the evaluation rubric (artifact c, Section 3) and an adaptation log. Misalignment indicators: the rubric scores only technical performance; its dimensions are frozen and cannot absorb new community-defined criteria; the adaptation log shows system changes with no corresponding community sign-off.

Step 6, Active redress and capacity transfer. Identifying impacts is not the same as redressing them. Step 6 tracks measurable shifts in local technical authority over time, through training programs, infrastructure investment, and explicit knowledge-transfer commitments, and treats the absence of such shifts as a finding the partnership must respond to rather than as background context. Documented through: capacity-transfer metrics (artifact d, Section 3) and the funding and capacity commitments they are measured against. Misalignment indicators: dependency indicators flat or rising across review cycles; training counted by headcount rather than transferred capability; local teams lack administrative access to development tools and documentation; contractual transfer milestones passed without being met.

The six steps draw on a body of normative argument, grounded in decolonial and intersectional theory, that has been developed elsewhere (Racine, 2024; 2025). The purpose here is more constrained: to specify what the protocol requires as instrumentation, that is, the artifacts deployment partners must produce and the dimensions assessors examine. The next section turns to that specification.

3. Four Classes of Required Technical Artifact

The protocol calls for four classes of artifact not currently standard in AIA practice. Throughout, we refer to the four *artifact classes*, and to the arrangement they form together as the *assessment stack*; the two terms are not interchangeable, one names the components and the other names their layered composition. The stack runs data → definition → measurement → governance. Provenance trails record what data came in and on what terms. Impact taxonomies name what counts as impact. Evaluation rubrics specify how impact gets scored. Capacity-transfer metrics track whether authority over the first three is being redistributed over time. Each layer answers a question the previous one cannot.

Moreover, each artifact is designed to extend existing tech-

nical infrastructure rather than to introduce new categories of tooling. That design choice is deliberate. It makes the protocol implementable by deployment teams and assessable by reviewers regardless of whether either party holds the protocol's full normative grounding.

3.1. Provenance Audit Trails

Definition. A queryable, structured record that fuses technical data lineage with the normative attributes governing each data asset's use across the system lifecycle.

The gap in existing tooling. Datasheets for Datasets (Gebru et al., 2021) document a dataset once at release, capturing composition, collection process, and recommended uses; they are static rather than queryable, and consent terms appear as narrative rather than enforceable structured attributes. Model cards (Mitchell et al., 2019) document the trained model, its intended uses, performance, and ethical considerations, but not the operational data flow. Industry lineage tooling (Apache Atlas, OpenLineage, DataHub) tracks which jobs produce which tables but treats data as a technical commodity rather than a normatively governed resource. Recent work on data provenance (Longpre et al., 2023) extends lineage to licensing but not yet to consent, sovereignty, or benefit-sharing at the depth healthcare deployment across jurisdictions requires.

Specification. For each data asset (training, fine-tuning, evaluation, operational input), the provenance trail records: (i) source provenance, that is, who collected the data, when, in what jurisdiction, and under what authorization; (ii) consent terms, the specific terms data subjects agreed to, with linkage to underlying consent records, including scope, duration, opt-out provisions, and secondary-use authorization; (iii) sovereignty assertion, which entity (state, community, council, or data trust) asserts governance rights over this data; (iv) benefit-sharing terms, what proportion of value derived from the data flows back, to whom, and in what form (revenue share, attribution, or derived-model access); (v) downstream restrictions, including jurisdictional limits, prohibited use cases, fine-tuning permissions, and third-party sharing terms; and (vi) a usage log, recording which models and systems have used this data, in what configurations, and with what justifications. The trail is queryable through a structured schema rather than presented as a free-text appendix.

Figure 1 gives a simplified relational schema for the trail, one example record per core table, and an audit query that operationalizes the consistency criterion below. Fixing field types and cross-table constraints turns misalignment indicators into queries rather than manual reading tasks. The design is deliberately minimal and maps onto attributes that consent-management and lineage systems already record; the contribution is to bind them in one queryable structure with the normative fields those systems omit.

Construction. Normative attributes are embedded at the point of data collection through consent management infrastructure and data trust governance frameworks; lineage tooling propagates these attributes through pipelines alongside technical metadata.

Audit criteria. Coverage, do all data assets carry complete records? Consistency, do operational uses align with the consent terms recorded? Right to erasure (where applicable), can data subjects' contributions be identified and removed? Sovereignty, are the governance entities recognized in the trail still in control of the relevant data assets? The artifact maps to protocol Step 3.

3.2. Co-Constructed Impact Taxonomies

Definition. A structured, co-developed document that names the dimensions along which deployment impact will be evaluated, the indicators for each, the stakeholders who participated in defining them, and the conditions under which the taxonomy gets revised.

The gap. Model cards typically include free-text "ethical considerations" sections rather than structured taxonomies. Risk taxonomies in AI safety (Weidinger et al., 2022; Shelby et al., 2023) are researcher-produced and top-down rather than co-constructed with affected communities. Work on the participatory turn in AI documents that stakeholder engagement, where it occurs, tends to produce one-off qualitative outputs rather than structured artifacts that downstream evaluation can reference, and cautions that participation without authority over outcomes can become a legitimating exercise (Delgado et al., 2023; Sloane et al., 2022). The gap this artifact addresses is therefore not the absence of participation but the absence of a durable, auditable record that binds participation to the evaluation that follows.

Specification. The artifact contains: (i) named impact dimensions specific to the deployment context; (ii) for each dimension, an explicit definition, indicators of success and harm, and the rationale for inclusion; (iii) provenance metadata, recording who participated in co-construction, when, under what conditions, and how disagreements were resolved; (iv) update triggers, the conditions under which the taxonomy gets revised; and (v) versioning, with changes to the taxonomy tracked over time. For healthcare deployment in a Global Majority context, dimensions might include: preservation of patient agency, respect for traditional practitioner authority, integration with non-biomedical care, avoidance of pathologizing language around traditional practices, and alignment with collective conceptions of wellbeing such as those articulated in non-Western (for example, Ubuntu) philosophy. Each dimension is stored as a structured record: a definition, paired success and harm indicators, a named

Figure 1. Simplified provenance-trail schema (core tables), one example record, and the consent-consistency audit query. The query returns the cross-border training violation of Section 4.3: an operational use whose `purpose` is not covered by the recorded consent scope. Types and cross-table constraints turn the Step 3 misalignment indicators into queries. Schema is illustrative and simplified, not a production data model.

```
-- Core tables (abbreviated) --
CREATE TABLE data_asset (
  asset_id      TEXT PRIMARY KEY,
  description   TEXT,
  asset_type    TEXT CHECK (asset_type IN
                  ('training', 'fine_tuning', 'evaluation', 'operational')),
  collected_by  TEXT, collected_when DATE, jurisdiction TEXT);
CREATE TABLE consent_record (
  consent_id    TEXT PRIMARY KEY,
  asset_id      TEXT REFERENCES data_asset(asset_id),
  scope         TEXT,                        -- terms subjects agreed to
  duration      TEXT, optout_provision BOOLEAN,
  secondary_use_authorized BOOLEAN);         -- e.g. external cross-border use
CREATE TABLE usage_log (
  use_id        TEXT PRIMARY KEY,
  asset_id      TEXT REFERENCES data_asset(asset_id),
  model_id      TEXT, system_id TEXT, used_when DATE,
  purpose       TEXT,                        -- what the asset was used for
  deployment_jurisdiction TEXT);              -- ISO country code
-- (sovereignty_assertion, benefit_sharing, downstream_restriction omitted)
-- Example records --
-- data_asset:  A17 | 'urban referral hospital records, country X' | training
-- consent_rec: C09 | A17 | 'improvement of clinical care' | secondary_use=FALSE
-- usage_log:   U42 | A17 | cross_border_model_training | jurisdiction=REGIONAL
-- Consent-consistency audit (Step 3): flag uses not covered by consent --
SELECT da.asset_id, da.description,
       ul.purpose, ul.deployment_jurisdiction,
       cr.scope, cr.secondary_use_authorized
FROM   data_asset     da
JOIN   usage_log      ul ON ul.asset_id = da.asset_id
JOIN   consent_record cr ON cr.asset_id = da.asset_id
WHERE  ul.purpose = 'cross_border_model_training'
  AND (cr.secondary_use_authorized = FALSE
       OR  cr.scope NOT LIKE '%cross-border%');
-- Returns A17: consent scope did not authorize the cross-border use.
```

participant set, and a link to the rubric row that scores it (Section 3.3).

Construction. Structured engagement workshops with patients, community health workers, traditional practitioners, frontline clinicians, and other relevant parties; iterative drafting that surfaces dimensions, validates them, and refines indicators; an explicit adjudication mechanism for unresolved disagreement; and publication as part of pre-deployment documentation.

Audit criteria. Provenance, who participated, were affected groups represented, and were participation conditions adequate (compensation, accessibility, language)? Coverage, do dimensions span the relevant impact surface, or are key concerns absent? Update, has the taxonomy evolved as deployment proceeded? Linkage, does the taxonomy actually drive the evaluation rubric (Section 3.3)? The artifact maps to protocol Step 4.

3.3. Community-Defined Evaluation Rubrics

Definition. The measurement instrument that operationalizes the taxonomy. Where the taxonomy names what counts as impact, the rubric specifies how that impact gets scored, who scores it, and what scores trigger which actions.

The gap. Standard ML evaluation suites (HELM, BIG-bench, MMLU) measure model capabilities in a deployment-agnostic way. Formal fairness metrics (demographic parity, equalized odds, predictive parity) can coincide with a community-defined concern for a given group, and where they do they should be used; what they do not provide is a systematic way to derive evaluation criteria from community-defined impact dimensions, being fixed statistical definitions over pre-specified groups rather than instruments a taxonomy of locally-defined harms can populate. Algorithmic auditing methodologies typically produce technical reports rather than stakeholder-derived scoring instruments tied to specific decisions.

Specification. For each dimension named in the taxonomy: (i) measurement methodology (interviews, structured observation, surveys, system log analysis, or mixed methods); (ii) scoring scale (qualitative bands, ordinal levels, or numerical with explicit definitions); (iii) sample size and coverage requirements; (iv) who applies the rubric (community evaluators, mixed teams, or third-party auditors); (v) decision rules, that is, what scores trigger redesign, escalation, rollback, or routine reporting; and (vi) calibration

Table 1. One populated rubric row for the taxonomy dimension *preservation of TBA decision authority*. A rubric contains one such row per taxonomy dimension.

Field	Value
Method	Structured birth-event observation, cross-referenced with system logs of alerts resulting in clinician override of TBA judgment
Scale	4-point ordinal, explicit anchors (1: routine override; 4: TBA judgment retained and documented)
Sampling	Stratified across health centres; minimum n per centre per quarter
Applied by	Mixed teams of community evaluators and clinical observers
Decision rule	Score below threshold for two consecutive quarters triggers protocol review
Calibration	Inter-rater reliability checked each cycle

provisions, including inter-rater reliability assessment and periodic validity checks. Table 1 shows a single populated rubric row for the TBA-authority dimension. Keeping (b) and (c) as separate artifacts matters: the same taxonomy can be applied through different rubrics (a quick proxy rubric versus a deep ethnographic one), and rubric-validity work is methodologically distinct from taxonomy-construction work.

Construction. Co-developed with the same stakeholders who built the taxonomy (continuity matters); pilot-tested on existing system data; and iterated based on the first round of measurement.

Audit criteria. Validity, does the rubric actually measure what the corresponding taxonomy dimension claims? Reliability, do different evaluators applying the same rubric produce consistent results? Action linkage, when scores trigger decision rules, are those decisions actually taken? The artifact maps to protocol Step 5.

3.4. Capacity-Transfer Metrics

Definition. A small set of longitudinal indicators that track the redistribution of technical authority over the system over time.

State artifacts versus the trajectory artifact. The first three artifacts are *state* artifacts: each records a current state (what data is present, what counts as impact, how it scores) revised when a trigger fires. What each measures is a state at a point in time. Capacity-transfer metrics are the *trajectory* artifact: change over time is not an update to the measurement, it is the thing measured. Sustainability and capacity-building reporting in development economics is typically narrative, financial, or focused on training-as-output rather than authority-as-outcome; MLOps dashboards track performance and pipeline health but not who holds authority over the system; and capacity-building claims in tech-for-development contexts are typically not measured longitudinally with named accountability.

Table 2. Worked capacity-transfer indicators for the illustrative deployment (Section 4.6). Figures are illustrative stipulations within a hypothetical composite, not measurements of a live system.

Indicator	Base	Month 24	Status
Local-initiated retraining runs	0%	35%	Plateaued
Docs in regional languages	0%	60%[†]	Partial
Local admins with access	0	8[‡]	Stalled
Budget reinvested locally (commit. 45%)	—	23%	Below target

[†] 0% of internal technical documentation needed for retraining. [‡] Reached by month 12, unchanged since.

Specification. A defined set of indicators, kept deliberately small to avoid measurement burden. For example: (i) share of model updates initiated by the local team versus the external partner; (ii) share of system documentation produced and maintained in local languages; (iii) number of local engineers, clinicians, and practitioners with administrative access to the system; (iv) proportion of project budget reinvested in local infrastructure, training, or community-controlled work; (v) decision points at which the local team holds and exercises veto authority; and (vi) the trajectory of dependency for critical maintenance, retraining, and adaptation tasks. Each indicator is baselined at deployment and reviewed at fixed intervals (typically quarterly), with named accountable parties and reporting to the community-controlled governance entity established under Step 2. Table 2 shows a worked instantiation drawn from the illustration in Section 4.

Construction. Indicators defined jointly with deployment partners and the community governance entity; baseline measurement at deployment; and integration into operational rhythm rather than treatment as one-off reports.

Audit criteria. Trajectory, are indicators moving in the direction of redistribution, or has the deployment plateaued in dependency? Comparison, do indicators match contractual capacity-transfer commitments? Action linkage, when indicators stagnate, what is done about it? The artifact maps to protocol Step 6.

3.5. The Stack as a Whole

The four artifact classes are minimal in that each layer answers a question the others cannot: provenance (a) does not say what counts as impact, a taxonomy (b) does not say how to measure it, a rubric (c) does not say whether author-

ity over the first three is being redistributed, and capacity-transfer metrics (d) presuppose the artifacts whose authority they track. Removing any layer leaves a structural gap; further sub-categorization is possible but optional rather than load-bearing.

Crucially, all four artifacts extend existing technical instrumentation: data lineage tooling, documentation conventions, evaluation harnesses, and deployment telemetry (the operational monitoring streams, request logs, override rates, uptime, and model-version history already collected in production). Implementation does not require new categories of infrastructure. It requires extending what gets recorded, who participates in defining it, and what gets reported to whom.

3.6. Minimum Viable Instantiation

The full protocol is demanding, and not every deployment can bear it before going live. We therefore distinguish the full instantiation from a minimum viable one, and identify which elements are load-bearing under time and resource constraints. Two carve-outs matter in practice.

Emergency and rapid-deployment settings. Where delay itself carries a mortality cost, a lite instantiation defers the labour-intensive elements (full workshop-based taxonomy co-construction, deep ethnographic rubrics) and retains three things that can be stood up quickly and guard against the least reversible harms: (i) a provenance trail seeded with consent, sovereignty, and downstream-restriction fields, because data commitments made at deployment are expensive to unwind later; (ii) a named community governance contact with a defined escalation path, even before a full advisory body is constituted; and (iii) a small baseline set of capacity-transfer indicators, because a baseline not captured at deployment cannot be reconstructed. The remaining elements become post-deployment obligations with dated milestones rather than being dropped.

Small and single-institution deployments. Where developer, deployer, and affected community are less separated, several artifacts collapse into lighter forms: the power map may be a short structured memo, and the rubric may reuse a validated proxy instrument rather than a bespoke one.

The distinction is between deferral and omission. A lite instantiation defers effortful elements against dated commitments; it does not omit the provenance and capacity baselines, whose value is largely destroyed if not captured early. Its adequacy is an empirical question we flag in Section 5.5.

4. Illustration: A Maternal Health System in Sub-Saharan Africa

To make the protocol concrete, we work through an illustrative deployment scenario constructed as a composite of documented systems in the region (drawing on Kimeu, 2024, and the landscape mapped in Section 1). The scenario is hypothetical and is not a primary report on any specific live system. All quantities given below, including the percentages in Section 4.6, are illustrative stipulations chosen to make the protocol's evaluation surface concrete; they are not measurements.

4.1. The Deployment Scenario

Consider an integrated AI-powered maternal health platform deployed across a regional partnership spanning multiple Sub-Saharan African countries, combining four AI components: (i) clinical decision support for prenatal care, integrated with existing electronic health records; (ii) AI-augmented fetal monitoring during labor; (iii) predictive analytics flagging high-risk pregnancies; and (iv) an mHealth interface for community health workers and patients. The partnership includes Ministries of Health, a non-local AI developer, regional health centers, networks of traditional birth attendants and community health workers, and an academic medical institution providing technical expertise. After two years, the platform is stipulated to report substantial gains in standard outcome metrics: reduced stillbirth rates, improved prenatal visit completion, and faster triage of high-risk cases.

4.2. What a Standard AIA Reports

A standard AIA on this deployment produces subgroup performance breakdowns; fairness audits showing accuracy gaps across rural versus urban populations and age cohorts; a model card documenting intended uses, demographic-level training composition, and known limitations; monitoring reports tracking adverse events, uptime, and override rates; and a privacy assessment confirming compliance with applicable data-protection regulation. That covers core technical performance but does not reach the dimensions Section 1 names. The next subsections walk through what each R-AIA artifact would surface here.

4.3. Provenance Audit Trail

The provenance trail traces every data asset feeding the system. It surfaces, for example: that 87% of training data for the diagnostic component originates from urban referral hospitals in two countries, with rural birth contexts underrepresented despite being the primary use environment; that consent forms for that urban data covered "improvement of clinical care" but not cross-border training of AI systems de-

ployed by external partners; that no benefit-sharing terms attach to the data despite substantial commercial value created; that no community-controlled entity asserts sovereignty over the traditional-birth-practice data the mHealth component has begun collecting; and that a recent fine-tuning run used data from a country whose regulation prohibits the resulting model's deployment outside it, yet the model runs regionally. None appears in the standard fairness audit; all are queryable and trigger named Step 3 indicators, and the consent violation is exactly what the query in Figure 1 returns.

4.4. Co-Constructed Impact Taxonomy

The taxonomy, co-developed through structured engagement with patients, traditional birth attendants, community health workers, and clinicians, names dimensions including preservation of TBA decision authority; integration with rather than replacement of customary birth practices; avoidance of pathologizing language in patient-facing outputs; alignment with collective conceptions of maternal wellbeing alongside clinical outcome; preservation of patient agency against confidence-weighted recommendations; and language accessibility across major regional languages. Several would not appear on a researcher-set risk taxonomy, particularly those concerning relationships between AI-mediated decisions and existing community authority structures. The taxonomy also documents disagreement: clinicians and traditional birth attendants weight "preservation of TBA authority" differently, and the taxonomy makes that explicit and binds the system to monitoring it rather than resolving it by fiat.

This is salient where standardized guidance diverges from infrastructural reality. Consider an infant-feeding recommendation a biomedical guideline would endorse for an HIV-positive mother but that presumes reliable formula, potable water, and sterilisation equipment the household may not have; where those presuppositions fail, following it can raise the very risk it aims to reduce (a case discussed further in Racine, 2025). A taxonomy dimension naming this context-dependence, and a rubric scoring against it, catch what a demographic accuracy check does not.

4.5. Community-Defined Evaluation Rubric

For each taxonomy dimension, the rubric specifies measurement. Table 1 gives the full row for "preservation of TBA decision authority," combining structured birth-event observation with system logs of clinician overrides. A second dimension, "avoidance of pathologizing language," is scored by community evaluators reviewing system-generated patient communications against criteria co-developed in the taxonomy, with deviations triggering content review. The rubric does not displace fairness or clinical accuracy metrics; it adds a parallel channel for dimensions those metrics cannot reach.

4.6. Capacity-Transfer Metrics

Tracked over the deployment's first 24 months, the indicators show a mixed picture (Table 2): local retraining autonomy rose then plateaued short of parity, documentation coverage stalled entirely for the internal materials retraining requires, administrative access froze after month 12, and local budget reinvestment ran at roughly half the committed share. Read together, the indicators flag a deployment sliding toward dependency rather than redistribution, despite the system's strong outcome metrics. Standard AIA reporting does not reach this layer.

4.7. What R-AIA Surfaces That Standard AIA Misses

Read together, the two assessments diverge. By standard fairness-and-performance assessment the system is doing well: it meets its outcome targets and its subgroup performance sits within acceptable bands. By R-AIA assessment the same system generates substantial value from data its partners neither hold sovereignty over nor share benefit from, plateaus in technical dependency, trends unfavorably on traditional birth attendant authority, and serves a context that does not match its training distribution. None of this makes the deployment unworthwhile on balance; all of it is what partners, the community governance entity, and the funder need in order to course-correct, in the same register as the technical metrics already reported.

5. Discussion

The R-AIA protocol is designed for deployments where developers, deployers, and affected communities are institutionally and geographically separated, where existing inequities shape who benefits and at what cost, and where governance gaps would otherwise leave significant harm unevaluated. It is not designed for short-cycle internal deployments and would over-specify lighter-touch contexts. As emphasized throughout, it augments existing AIA practice, of which fairness auditing is one component, rather than replacing it.

5.1. Implementation Burden, Incentives, and Failure Modes

The protocol imposes real cost: skilled local time, administrative labour, and calendar delay, precisely where all three are scarce. Two responses are in order. First, the burden is not uniform: provenance and capacity baselines are cheap to seed at deployment and expensive to reconstruct later, whereas taxonomy and rubric co-construction are labour-intensive, and Section 3.6 uses this asymmetry to defer the latter against dated commitments rather than omitting the

former. Second, the protocol only functions if incentives are aligned with it. Each artifact has a characteristic failure mode when they are not: provenance trails populated only where records are convenient; taxonomies quietly pruned to dimensions the system already satisfies; rubrics scored by evaluators who depend on the deployment partner; and capacity-transfer metrics reported narratively rather than against baselines with consequences. Identifying these failure modes does not prevent them, and the mitigations we discuss next are partial.

5.2. Capture and Co-optation

Any structured assessment instrument can be co-opted as compliance theatre, and the four artifacts here are no exception. Two capture pathways deserve attention because they can turn the protocol against the people it is meant to protect.

The first is capture of the governance mechanism itself. "Binding veto" assumes a body that genuinely represents the affected community; in practice, local elites or intermediaries may occupy it and wield its authority against the least powerful, including minorities within the community. A veto is only as reparative as the legitimacy of the body holding it, which the protocol cannot manufacture. Composition requirements and disaggregated rubric reporting that flags when a subgroup is outvoted reduce this risk without removing it.

The second is dilution by the parties who hold the resources. Where a deployment is funded by powerful donors or large technology firms, "binding" veto rights without legal backing in national law can erode into a symbolic gesture, retained on paper while real decisions track funding. Tying veto and capacity-transfer provisions to funding consequences, and making provenance trails and taxonomies publicly auditable, raises the cost of dilution without guaranteeing against it. The artifacts make capture visible and more costly, but visibility is not prevention, and a protocol that claimed otherwise would overclaim.

5.3. Interoperability with Existing Governance Frameworks

The R-AIA augments, not replaces, instruments such as the EU AI Act's conformity assessment regime, the NIST AI Risk Management Framework, and ISO/IEC 42001. Two questions follow. First, where can an artifact serve as evidence within an existing framework? Provenance trails map onto data-governance and record-keeping obligations, capacity-transfer metrics onto post-market monitoring. Second, and harder, community-defined rubrics introduce locally-derived criteria that do not standardize across regions, in tension with regimes that prize comparable metrics. The resolution is not to abandon local definitions but a two-layer structure: a standardized outer layer (which dimensions were assessed, by whom, with what decision rules and actions) comparable across deployments, wrapping a locally-defined inner layer that is not. This keeps process auditable and comparable across sites while keeping substance local. Working these mappings out by domain is an open problem.

5.4. Other Open Problems

Disagreement adjudication. When community-defined rubrics produce findings that conflict with technical performance metrics, for instance a system meeting clinical accuracy targets while scoring poorly on TBA authority preservation, the R-AIA surfaces the conflict but does not resolve it, and resolution-by-default in favour of either side is inadequate.

Metrics across heterogeneous deployments. Indicators that suit a multi-country partnership may not transfer to single-country or structurally different deployments; generalizable templates with deployment-specific instantiation would help.

5.5. Limitations

This paper specifies the protocol and required artifacts and gives concrete instantiations for each, but does not empirically validate them. Validation across live deployments, including feasibility, stakeholder acceptance, cost, and reliability, is a crucial next step, and the tiering in Section 3.6 is a hypothesis about sequencing under constraint rather than a validated result. The case in Section 4 is a hypothetical composite whose quantities are stipulated to make the evaluation surface concrete; it shows what the artifacts would surface, not that a real deployment would accept or sustain them. Finally, the protocol is grounded in decolonial and intersectional theory whose commitments may not be uniformly held across the technical AI governance community; our claim is that the artifacts are useful instrumentation regardless of full theoretical adoption, a claim that itself warrants empirical examination.

6. Conclusion

The Reparative AIA protocol gives the technical AI governance community a concrete instrument for assessing healthcare AI deployments where existing inequities shape who benefits, who bears risk, and who retains authority. Its four artifact classes, provenance audit trails, co-constructed impact taxonomies, community-defined evaluation rubrics, and capacity-transfer metrics, form a minimal stack across data, definition, measurement, and governance, and each catches failure modes that fairness audits do not reach. Together they let partners, communities, and assessors evaluate

a deployment's structural conditions in the register already used for its technical performance. The instrumentation is technically tractable, and for the provenance trail this paper gives a schema and query concrete enough to build from.

Impact Statement

This paper proposes an assessment protocol intended to surface harms that standard algorithmic impact assessments and fairness audits do not currently reach, in particular harms tied to data sovereignty, the epistemic exclusion of non-Western knowledge systems, and the long-term distribution of technical authority over deployed systems in the Global Majority. The protocol is designed to augment rather than replace existing instrumentation and to make structurally-mediated harms visible in a register comparable to that of technical performance metrics. As Section 5.2 discusses at length, any structured assessment instrument can be co-opted as compliance theatre or captured by parties who do not represent the affected community, and the safeguards we propose reduce rather than eliminate that risk. We do not foresee specific additional risks from the publication of this work beyond those normally associated with AI governance scholarship.

References

Ahmed, K. South Sudan medics trial AI app to identify snakes and improve bite treatment. The Guardian, September 2024. https://www.theguardian.com/global-development/2024/sep/25/south-sudan-medics-trial-ai-software-to-identify-snake-and-improve-bite-treatment.

Ashar, A., Ginena, K., Cipollone, M., Barreto, R., and Cramer, H. Algorithmic impact assessments at scale: Practitioners' challenges and needs. *Journal of Online Trust and Safety*, 2(4):1–27, 2024. doi: 10.54501/jots.v2i4.206.

Benjamin, R. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, Cambridge, UK, 2019.

Birhane, A. The limits of fairness. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*. Association for Computing Machinery, 2022. doi: 10.1145/3514094.3539568.

Birhane, A., Isaac, W., Prabhakaran, V., Diaz, M., Elish, M. C., Gabriel, I., and Mohamed, S. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, pp. 1–8. Association for Computing Machinery, 2022. doi: 10.1145/3551624.3555290.

Costanza-Chock, S. *Design Justice: Community-Led Practices to Build the Worlds We Need*. MIT Press, Cambridge, MA, 2020. ISBN 978-0-262-04345-8.

Davis, J. L., Williams, A., and Yang, M. W. Algorithmic reparation. *Big Data & Society*, 8(2), 2021. doi: 10.1177/20539517211044808.

Delgado, F., Yang, S., Madaio, M., and Yang, Q. The participatory turn in AI design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*, pp. 1–23. Association for Computing Machinery, 2023. doi: 10.1145/3617694.3623261.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., and Crawford, K. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021. doi: 10.1145/3458723.

Hu, C. Aparna hegde and aparna taneja: Reaching out to women. TIME100 Health 2025, May 2025. https://time.com/collections/time100-health-2025/7279670/aparna-hegde-aparna-taneja/.

Kimeu, C. How AI monitoring is cutting stillbirths and neonatal deaths in a clinic in Malawi. The Guardian, December 2024. https://www.theguardian.com/global-development/2024/dec/06/how-ai-monitoring-is-cutting-stillbirths-and-neonatal-deaths-in-a-clinic-in-malawi.

Korom, R., Kiptinness, S., Adan, N., Said, K., Ithuli, C., Rotich, O., Kimani, B., King'ori, I., Kamau, S., Atemba, E., Aden, M., Bowman, P., Sharman, M., Soskin Hicks, R., Distler, R., Heidecke, J., Arora, R. K., and Singhal, K. AI-based clinical decision support for primary care: A real-world study. OpenAI & Penda Health, 2025. arXiv:2507.16947. https://openai.com/index/ai-clinical-copilot-penda-health/.

Longpre, S., Mahari, R., Chen, A., Obeng-Marnu, N., Sileo, D., Brannon, W., Muennighoff, N., Khazam, N., Kabbara, J., Perisetla, K., Wu, X., Shippole, E., Bollacker, K., Wu, T., Villa, L., Pentland, S., and Hooker, S. The data provenance initiative: A large scale audit of dataset licensing & attribution in AI. *arXiv preprint arXiv:2310.16787*, 2023.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, pp. 220–229, 2019. doi: 10.1145/3287560.3287596.

Mohamed, S., Png, M.-T., and Isaac, W. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33(4):659–684, 2020. doi: 10.1007/s13347-020-00405-8.

Moss, E., Watkins, E. A., Singh, R., Elish, M. C., and Metcalf, J. Assembling accountability: Algorithmic impact assessment for the public interest. Technical report, Data & Society, 2021. https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/.

Partnership on AI. Guidelines for participatory and inclusive AI, 2024. https://partnershiponai.notion.site/1e8a6131dda045f1ad00054933b0bda0.

Racine, E. E. Reparative algorithmic impact assessments: A decolonial, justice-oriented accountability framework for AI and the global majority. In Belli, L. and Britto Gaspar, W. (eds.), *AI from the Global Majority: Official Outcome of the UN IGF Data and Artificial Intelligence Governance Coalition*, pp. 137–150. FGV Direito Rio, 2024.

Racine, E. E. Harnessing artificial intelligence to meet global healthcare challenges: Reparative algorithmic impact assessments in context. *Emerging Media*, 3(3):414–427, 2025. doi: 10.1177/27523543251372031.

Reuel, A., Bucknall, B., Casper, S., Fist, T., Soder, L., Aarne, O., Hammond, L., Ibrahim, L., Chan, A., Wills, P., Anderljung, M., Garfinkel, B., Heim, L., Trask, A., Mukobi, G., Schaeffer, R., Baker, M., Hooker, S., Solaiman, I., Luccioni, A. S., Rajkumar, N., Moes, N., Ladish, J., Guha, N., Newman, J., Bengio, Y., South, T., Pentland, A., Koyejo, S., Kochenderfer, M. J., and Trager, R. Open problems in technical AI governance. *arXiv preprint arXiv:2407.14981*, 2024.

Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., Yilla-Akbari, N., Gallegos, J., Smart, A., Garcia, E., and Virk, G. Sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*, pp. 723–741, 2023. doi: 10.1145/3600211.3604673.

Sloane, M., Moss, E., Awomolo, O., and Forlano, L. Participation is not a design fix for machine learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, pp. 1–6. Association for Computing Machinery, 2022. doi: 10.1145/3551624.3555285.

Stahl, B. C., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., Kirichenko, A., Marchal, S., Rodrigues, R., Santiago, N., Warso, Z., and Wright, D. A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review*, 56:12799–12831, 2023. doi: 10.1007/s10462-023-10420-8.

Verma, S., Lalan, A., Diaz, P. R., Danassis, P., Mahale, A., Sudan, K. M., Hegde, A., Tambe, M., and Taneja, A. Leveraging AI to improve health information access in the world's largest maternal mobile health program. *AI Magazine*, 45:526–536, 2024. doi: 10.1002/aaai.12206.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Rimell, L., Isaac, W., Haas, J., Legassick, S., Irving, G., and Gabriel, I. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, pp. 214–229, 2022. doi: 10.1145/3531146.3533088.

The appendices that follow are illustrative reference implementations rather than normative technical standards; they show one concrete way to build the artifacts of Section 3, not the only admissible one.

A. Full Provenance-Trail Schema and Example Audit Queries

Figure 1 in the body gives the three core tables and one audit query. This appendix completes the schema with the three remaining field-groups from the Section 3.1 specification (sovereignty assertion, benefit-sharing terms, and downstream restrictions) and adds three further audit queries. As in the body, the schema is illustrative and simplified rather than a production data model; types and constraints are shown to make the misalignment indicators of Step 3 executable as queries rather than manual reading tasks.

```
-- Remaining core tables (the three omitted from Figure 1) --
CREATE TABLE sovereignty_assertion (
  assertion_id  TEXT PRIMARY KEY,
  asset_id      TEXT REFERENCES data_asset(asset_id),
  governing_entity  TEXT,          -- state, community council, data trust, ...
  entity_type   TEXT CHECK (entity_type IN
                  ('state', 'community', 'council', 'data_trust', 'other')),
  asserted_when DATE,
  entity_active BOOLEAN);          -- is the governing entity still recognised?

CREATE TABLE benefit_sharing (
  term_id       TEXT PRIMARY KEY,
  asset_id      TEXT REFERENCES data_asset(asset_id),
  beneficiary   TEXT,              -- who value flows back to
  form          TEXT CHECK (form IN
                  ('revenue_share', 'attribution', 'model_access', 'none')),
  share_pct     NUMERIC,           -- may be 0 for non-revenue forms; use form='none' for no arrangement
  agreed_when   DATE);

CREATE TABLE downstream_restriction (
  restriction_id TEXT PRIMARY KEY,
  asset_id      TEXT REFERENCES data_asset(asset_id),
  restriction_type TEXT CHECK (restriction_type IN
                  ('jurisdiction', 'prohibited_use', 'finetune', 'third_party')),
  value         TEXT,              -- e.g. an ISO country code, a use-case label
  binding       BOOLEAN);

-- Query 1: Right to erasure (where applicable) ----------------------------
-- All assets and downstream uses linked to a given consent record, so a
-- subject's contribution can be located and removed.
SELECT da.asset_id, ul.use_id, ul.model_id, ul.system_id
FROM   consent_record cr
JOIN   data_asset     da ON da.asset_id = cr.asset_id
LEFT JOIN usage_log   ul ON ul.asset_id = da.asset_id
WHERE  cr.consent_id = :target_consent_id;

-- Query 2: Sovereignty still in force ---------------------------------------
-- Assets whose asserted governing entity is no longer recognised as active:
-- the trail claims sovereignty that no live entity is exercising.
SELECT da.asset_id, sa.governing_entity, sa.entity_type
FROM   data_asset da
JOIN   sovereignty_assertion sa ON sa.asset_id = da.asset_id
WHERE  sa.entity_active = FALSE;

-- Query 3: Value created without benefit-sharing ----------------------------
-- Assets used operationally but carrying no benefit-sharing arrangement:
-- commercial or clinical value is created with nothing flowing back. A zero
-- revenue share is not tested here, since benefit-sharing may take non-revenue
-- forms (attribution, model access) that the schema records separately.
SELECT da.asset_id, COUNT(ul.use_id) AS n_uses
FROM   data_asset da
JOIN   usage_log  ul ON ul.asset_id = da.asset_id
LEFT JOIN benefit_sharing bs ON bs.asset_id = da.asset_id
WHERE  bs.term_id IS NULL
   OR  bs.form = 'none'
GROUP BY da.asset_id
HAVING COUNT(ul.use_id) > 0;
```

Each query corresponds to one of the Step 3 audit criteria in Section 3.1: Query 1 to right to erasure, Query 2 to sovereignty, and Query 3 to the consistency of operational value against recorded benefit-sharing. The consent-consistency query of Figure 1 covers the fourth. Coverage (do all assets carry complete records across all six tables?) is checked by outer-joining `data_asset` against each attribute table and flagging null groups; it is omitted here for space.

B. Indicative Mapping of R-AIA Artifacts to Existing Governance Frameworks

Section 5.3 argues that the R-AIA augments rather than replaces existing instruments, and that several artifacts can serve as evidence within them. Table 3 sketches where each artifact plausibly attaches to obligations in the EU AI Act (Regulation (EU) 2024/1689), the NIST AI Risk Management Framework (AI RMF 1.0), and ISO/IEC 42001:2023. **The mapping is indicative rather than authoritative**: it identifies the obligation each artifact is best positioned to evidence, not a certification of sufficiency, and the precise article and control references should be confirmed against the current text of each instrument before use in a compliance context. The community-defined inner layer of the rubric (Section 5.3) is, by design, the element least reducible to a standardized control.

Table 3. Indicative mapping of the four artifact classes to obligations in three governance frameworks. References are indicative and should be verified against the current text of each instrument.

Artifact	EU AI Act (2024/1689)	NIST AI RMF 1.0	ISO/IEC 42001:2023
Provenance audit trail (a)	Data and data governance (Art. 10); record-keeping and logging (Art. 12)	MAP (data provenance, context); GOVERN (accountability)	Data-management and record controls (Annex A); operational planning (Cl. 8)
Co-constructed impact taxonomy (b)	Risk-management system, hazard identification (Art. 9)	MAP (impacts, affected communities)	AI impact assessment; planning to address risks (Cl. 6, Annex A)
Community-defined evaluation rubric (c)	Accuracy and robustness (Art. 15); post-market monitoring (Art. 72, where applicable)	MEASURE (evaluation, metrics)	Performance evaluation, monitoring and measurement (Cl. 9)
Capacity-transfer metrics (d)	Can inform post-market monitoring (Art. 72)	MANAGE, GOVERN (ongoing oversight)	Continual improvement (Cl. 10)

Two integration patterns follow from the table. First, artifacts (a), (c), and (d) map onto obligations that are already largely procedural (record keeping, evaluation, monitoring), so the R-AIA mostly specifies *what* those procedures should record and report rather than adding a new obligation type. Second, the taxonomy (b) and the inner layer of the rubric introduce locally-defined content that the frameworks accommodate as inputs (impact identification, oversight criteria) but do not themselves standardize; this is the two-layer structure of Section 5.3, with the frameworks providing the comparable outer layer and the community providing the local inner one.

C. Consolidated Protocol Reference

Table 4 collects the six protocol steps of Section 2 into a single reference, pairing each step with its function, principal output, the artifact it produces or draws on, and one representative misalignment indicator. It is a synopsis of the body text rather than new content.

Table 4. The six R-AIA steps at a glance. The final column gives one representative misalignment indicator per step; the body lists the full set.

Step	Function	Principal output	Artifact	Representative misalignment indicator
1. Socio-historical context audit	Document the terrain the system enters	Research synthesis with named contributors	—	Synthesis cites only indexed, English-language sources
2. Power analysis	Make authority explicit; establish binding governance	Power map; governance instrument	—	Governance instrument grants consultation but not binding veto
3. Reparative data sovereignty	Capture and audit data governance conditions	Provenance trail; data governance instrument	a	Consent or benefit-sharing fields empty for sensitive assets
4. Community-led impact co-construction	Define what counts as impact, with authority	Impact taxonomy; engagement protocol	b	Community-defined dimensions dropped where they conflict with the technical plan
5. Ongoing monitoring	Score community-defined criteria over time	Evaluation rubric; adaptation log	c	Rubric dimensions frozen; cannot absorb new community criteria
6. Active redress and capacity transfer	Track redistribution of technical authority	Capacity-transfer metrics; commitments	d	Transfer milestones passed without being met