The Deployable AI Index

A Structured Decision Model for Health Systems

A framework for assessing when AI systems are ready for real-world deployment

Contents

Executive Summary

Most AI projects in health systems fail not due to inadequate technology, but because organizations are unprepared to deploy them. The Deployable AI Index (DAI) provides a structured, operational model that synthesizes eight critical factors—Purpose, People, Process, Performance, Protection, Reliability, Economics, and Friction—into a single 0–1 score with stage-appropriate safety gates. Unlike existing readiness assessments that measure capability across dimensions, DAI specifically addresses the deployment decision: Should we proceed now, and if not, what must improve first? This paper presents the conceptual framework, demonstrates its application in a low- and middle-income country (LMIC) health context, and positions it as a decision tool pending empirical validation.

Who Should Use This Index?

The DAI is designed for decision-makers across multiple levels of health systems, with primary applications for:

- **Health systems evaluating AI innovations:** Ministries of health, hospital networks, and health system administrators assessing whether AI solutions are ready for deployment and identifying specific readiness gaps that must be addressed
- **Funders and donors:** Development partners, digital health funders, and implementing organizations benchmarking specific interventions to determine appropriate investment levels, de-risk portfolios, and stage funding decisions based on deployment readiness
- **AI vendors and implementers:** Solution providers and technical partners using DAI to self-assess readiness, prioritize improvement areas, and demonstrate deployment viability to prospective clients

While developed with insights from LMIC health system implementations, the DAI framework is globally applicable with context-specific calibration of thresholds (e.g., cost normalization, infrastructure baselines, regulatory requirements).

What is this?

The Deployable AI Index (DAI) is a single number—between 0 and 1—that quantifies how ready an AI system is for real-world deployment in health settings.

> 💡 **Think of it as a credit score for AI deployment readiness:**
>
> - = Not ready (critical gaps exist)
> - **0.25** = Ready for small pilot tests
> - **0.50** = Ready for limited rollout
> - **0.75+** = Ready for full production

Why This Matters

The AI Deployment Crisis in Health Systems

Across global health systems, AI pilots proliferate while production deployments remain rare. MIT research indicates that approximately 90-95% of AI pilot projects fail to reach production deployment, with organizations struggling to scale beyond proof-of-concept stages.[1] In health systems specifically, this challenge is amplified by sector-unique constraints including regulatory requirements, patient safety mandates, clinical workflow integration complexity, and resource limitations in low- and middle-income settings. Common failure patterns transcend technical performance:

- **Strategic misalignment:** AI solutions disconnected from health system priorities or clinical workflows
- **Workforce unpreparedness:** Clinicians lack training or confidence to use AI tools effectively
- **Process incompatibility:** AI requires workflow redesigns that organizations cannot absorb
- **Safety gaps:** Inadequate security, privacy controls, or reliability monitoring for clinical use
- **Economic unsustainability:** Unit costs or total cost of ownership exceed available budgets
- **Deployment friction:** Technical integration barriers or organizational resistance create insurmountable drag

The Missing Framework for Health AI Deployment

Existing AI readiness tools—including the Cisco AI Readiness Index, Singapore's AIRI, and the IMF AI Preparedness Index—assess general organizational capability across dimensions but provide limited guidance on the deployment decision itself. These frameworks were designed for enterprise AI adoption broadly, not for the unique constraints of health systems where patient safety, regulatory compliance, clinical workflow integration, and public health mandates create deployment challenges absent from other sectors.

Why conventional health intervention frameworks fall short for AI:

Traditional implementation science frameworks (RE-AIM, CFIR) and health systems strengthening models (WHO building blocks) were designed for static interventions—vaccines, protocols, medical devices—where the intervention itself remains constant once deployed. AI systems are fundamentally different: they are dynamic (models drift and require continuous monitoring), algorithmically complex (performance varies with local data distributions), technically dependent (requiring specialized infrastructure and integration expertise), and subject to unique governance challenges (explainability, bias monitoring, automated decision-making accountability). While these frameworks address organizational readiness and implementation fidelity, they do not account for algorithmic reliability, model governance, continuous performance monitoring, or the technical integration friction specific to AI deployment in resource-constrained health settings.

Policymakers using existing tools receive multidimensional scores without clear deployment thresholds: When is capability sufficient? What level of readiness justifies investment? Which gaps are deployment-blocking versus improvable post-pilot? No framework specifically addresses the health AI deployment decision with sector-appropriate safety gates, cost models calibrated for health budgets, and friction assessments tuned to clinical workflow realities.

The DAI addresses this gap *by operationalizing "deployability" as a function of readiness, safety, economics, and friction, with explicit stage gates that distinguish pilot-appropriate from production-ready systems. It extends established implementation science and health systems frameworks with AI-specific dimensions while maintaining compatibility with existing assessment approaches used by health system decision-makers.*

Theoretical Grounding and Design Principles

The DAI synthesizes insights from established research domains. The table below presents the core components of key frameworks that inform the DAI's design:

Framework Components Comparison: Thematic Alignment Across Domains

Thematic Dimension	RE-AIM (Implementation Science)	CFIR (Implementation Research)	WHO Building Blocks (Health Systems)	Rogers (Diffusion Theory)	TAM (Technology Acceptance)	NIST AI RMF (AI Governance)
Framework Links:	https://re-aim.org/	https://cfirguide.org/	https://extranet.who.int/nhptool/BuildingBlock.aspx	https://en.wikipedia.org/wiki/Diffusion_of_innovations	https://en.wikipedia.org/wiki/Technology_acceptance_model	https://www.nist.gov/itl/ai-risk-management-framework
Strategic Alignment & Value Does the innovation address a priority need? Is there evidence of benefit?	✓ Effectiveness	✓ Innovation: Evidence strength, relative advantage	✓ Service Delivery: Access, quality, coverage	✓ Relative Advantage, Observability	✓ Perceived Usefulness	✓ Map: Context & impact analysis
Human Capability & Workforce Do users have skills, training, and capacity to operate the system?	✓ Adoption (staff)	✓ Individuals: Knowledge, self-efficacy	✓ Health Workforce: Competence, availability	—	✓ Perceived Ease of Use	✓ Govern: Resources & competence (ISO 42001: Support)
Workflow & Processes Are workflows documented, mature, and compatible with the innovation?	✓ Implementation: Fidelity, consistency	✓ Process: Planning, executing; Inner Setting: Readiness	✓ Service Delivery: Workflow integration	✓ Compatibility	—	✓ Govern: Processes & procedures (ISO 42001: Operation)
Technical Performance & Quality Does the system meet technical specifications and deliver outcomes?	✓ Effectiveness: Outcomes	✓ Innovation: Design quality, complexity	✓ Medical Products: Quality, efficacy	✓ Trialability (testing)	—	✓ Measure: Tracking performance (ISO 42001: Performance Evaluation)

Thematic Dimension	RE-AIM (Implementation Science)	CFIR (Implementation Research)	WHO Building Blocks (Health Systems)	Rogers (Diffusion Theory)	TAM (Technology Acceptance)	NIST AI RMF (AI Governance)
Security, Privacy & Safety **Are data protection, security controls, and safety mechanisms in place?**	—	—	✓ Governance: Regulation, accountability	—	—	✓ Govern, Map, Measure: Risk management
Reliability & Uptime **Does the system operate consistently with minimal failure?**	—	—	✓ Health Information Systems: Reliability	—	—	✓ Measure, Manage: Continuous monitoring
Economics & Affordability **Are costs reasonable relative to budget and expected value?**	✓ Implementation: Cost of delivery	✓ Innovation: Cost; Outer Setting: Resources	✓ Health Financing: Purchasing, cost-effectiveness	—	—	✓ Manage: Resource allocation
Deployment Friction & Complexity **How difficult is integration, training, and organizational change?**	—	✓ Innovation: Complexity, adaptability; Inner Setting: Culture, climate	—	✓ Complexity (inverse)	✓ Perceived Ease of Use (inverse)	—
Governance & Leadership **Is there executive commitment, policy support, and accountability?**	—	✓ Outer Setting: Policies, incentives; Inner Setting: Leadership engagement	✓ Leadership & Governance: Policy, oversight	—	—	✓ Govern: Organizational structure, policies (ISO 42001: Leadership)
Reach & Scale **What proportion of the target population or setting adopts the innovation?**	✓ Reach, Adoption	✓ Outer Setting: Patient needs	✓ Service Delivery: Coverage	—	✓ Behavioral Intention	—
Sustainability & Maintenance **Can the system be maintained and sustained over time?**	✓ Maintenance	✓ Process: Reflecting, evaluating	✓ Health Financing: Sustainability	—	—	✓ Manage: Continuous improvement (ISO 42001: Improvement)

These frameworks collectively inform the eight-factor structure of the DAI, as described in the following sections:

Implementation Science

The model's structure reflects RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) and CFIR (Consolidated Framework for Implementation Research) dimensions, which consistently identify organizational readiness, outer setting constraints, and implementation process as critical determinants of health intervention success. Purpose aligns with CFIR's "intervention source" and "evidence strength"; People with "available resources" and "access to knowledge"; Process with "implementation climate" and "readiness for implementation"; Performance with "intervention effectiveness" in RE-AIM terms.

Health Systems Strengthening

The eight factors map to WHO health systems building blocks—particularly governance (Protection), service delivery (Process, Performance), health workforce (People), and financing (Economics)—extended with AI-specific considerations around reliability and technical integration friction absent from traditional frameworks.

Technology Adoption Theory

The balanced treatment of Core Readiness reflects Rogers' diffusion theory dimensions (relative advantage, compatibility, complexity) and the Technology Acceptance Model's dual emphasis on perceived usefulness (Performance, Purpose) and ease of use (inverse of Friction). The stage-gate structure aligns with Moore's technology adoption lifecycle, distinguishing early adopters (pilot) from early majority (limited production) and late majority (full production) readiness thresholds.

AI Governance Standards

Safety gates align with NIST AI Risk Management Framework's govern-map-measure-manage structure and ISO/IEC 42001's emphasis on continuous monitoring and risk mitigation as prerequisites for deployment. The binary gate logic reflects regulatory practice in medical device approval and clinical AI certification, where safety and reliability thresholds are non-negotiable.

Design Assumption

> ⚠ This model assumes partial independence among factors—e.g., strong Purpose does not automatically generate strong Process—while recognizing documented interaction effects (high friction reduces adoption regardless of technical performance). Geometric aggregation deliberately amplifies weak-link effects because empirical case studies consistently show that severe deficiency in any single dimension predicts deployment failure, even when other dimensions are strong. This design choice prioritizes balanced readiness over average readiness.

The Core Idea

$$\textbf{Deployable AI Index} = \frac{\text{(How prepared you are} \times \text{How safe it is} \times \text{Whether you can afford it)}}{\text{How hard it is to actually use}}$$

DAI (Deployable AI Index) increases when organizations have prepared the fundamentals well, built safe and affordable systems, and decreases when deployment faces excessive organizational or technical barriers.

The Four Building Blocks of DAI

1. Core Readiness — The Foundation

Question: Have you prepared the organizational and technical basics?

This combines four factors that must all be reasonably strong:

Purpose — Strategic clarity

Purpose begins with problem identification. AI deployment must directly address a measurable health system gap—not simply because the technology is available, but because it solves a priority problem that cannot be addressed as effectively through non-AI interventions.

Requirements:

- Clear linkage to health system priorities (disease burden reduction, access expansion, efficiency gains)
- Defined, measurable success metrics with baseline data
- Executive sponsorship with committed budget and change mandate

Example:

> ***Problem:*** TB cases detected 21+ days after symptom onset; current case detection rate 65% vs. national target 90%
>
> ***Solution:*** Chest X-ray AI-assisted screening to reduce diagnostic delay to 10 days and increase detection by 20%
>
> ***Alignment:*** District Medical Officer champion, donor funding committed for 3 years, explicit targets in National TB Strategic Plan.

People — Human capability

Requirements:

- Clinical and technical staff trained to use, maintain, and improve the AI system
- Clear roles, accountability structures (RACI matrices), and escalation paths
- Organizational culture demonstrating successful change initiatives and growth mindset

Example:

> ***Challenge:*** Only 12 of 30 radiographers trained (40%); high turnover (18-month average tenure); clinicians skeptical of AI accuracy.
>
> ***Action needed:*** Complete training for remaining staff, establish peer champion network, create escalation path to regional referral hospital.
>
> ***Score impact:*** People score 2.8 → below optimal but above pilot gate (≥2.5).

Process — Workflow maturity

Requirements:

- Well-designed clinical or administrative workflows that accommodate AI
- Documented standard operating procedures, decision rights, governance cadence
- Feedback loops and continuous improvement mechanisms (plan-do-study-act cycles)

Example:

> ***Current state:*** 70% of TB diagnostic pathway is informal/paper-based; no EMR; X-ray results manually transcribed; monthly quality reviews just started
>
> ***Gap:*** Workflow immaturity blocks AI integration—cannot embed digital tool in paper-based system
>
> ***Score impact:*** Process score 2.2 → **fails pilot gate (≥2.5 required) → deployment blocked**

Performance — Demonstrated value

Performance measures whether the AI performs as claimed through independent validation on local data. This distinguishes vendor-reported performance (test datasets, controlled conditions) from validated performance (your context, your data, your infrastructure). Performance assesses technical capability only—not clinical outcomes, adoption rates, or usability (those are captured in Purpose and Friction respectively).

Requirements:

- Independent validation: Assessment of accuracy, sensitivity, specificity, AUC-ROC on representative local data (not just vendor test sets or benchmark datasets)
- Performance adequacy: Validated performance meets minimum thresholds for the intended clinical or administrative task
- Stability: Performance remains consistent across data subgroups, time periods, and operating conditions (no unexpected degradation)

Example:

> ***Vendor claim:*** 90% sensitivity, 85% specificity on public TB screening benchmark dataset (ImageNet-derived).
>
> ***Independent validation:*** Pilot with 200 local patients showed 88% sensitivity, 83% specificity—close to claim but slightly lower on local X-ray quality.
>
> ***Adequacy assessment:*** 88% sensitivity exceeds minimum clinical threshold (≥85% for TB screening per WHO guidelines); validated performance adequate for intended use.
>
> ***Score impact:*** Performance score 3.5 → technology validated and adequate; minor gap between claimed and real-world performance but within acceptable range.

> Behind the scenes, we use a balanced (geometric-style) average:
>
> **Core Readiness = $\sqrt[4]{\text{Purpose} \times \text{People} \times \text{Process} \times \text{Performance}}$**.
>
> This ensures that one very low score in any dimension pulls down overall readiness significantly, reflecting the real-world experience that weak links break deployments.

> 💡 **Key insight:** You need all four to be decent. One weak area drags down the overall score—because that's what happens in reality. Excellent technology with untrained clinicians fails. Clear goals with broken workflows fail. High accuracy with zero adoption fails.

2. Safety — The Guardrails

Question: Is it secure and reliable enough for clinical use?

This is the minimum safety bar you must clear:

Protection — Security and compliance

- Encryption at rest and in transit, multi-factor authentication, regular penetration testing
- Privacy regulations compliance (GDPR, HIPAA, local data protection laws)
- Documented policies, immutable audit trails, tested incident response plans

Reliability — Consistent operation

- Uptime meets deployment-stage needs (pilot: 95%+, limited: 98%+, full production: 99%+)
- Model accuracy remains stable with low drift (<5% degradation monitored continuously)
- Automated monitoring with real-time alerting for anomalies or failures

> Safety uses the **weakest-link rule**: Safety = min (Protection, Reliability). If either score falls below the stage gate threshold, deployment is blocked. You cannot compensate for weak security with high reliability, or vice versa. Both must meet the bar.

> ⚠ This is a gate, not a score. Below the threshold → DAI = 0 (non-deployable), regardless of other strengths.

3. Economics — The Reality Check

Question: Can you afford this at scale overtime?

Total Cost of Ownership (TCO)

Economics assesses **Total Cost of Ownership (TCO)**—the full lifetime cost averaged over a 3-year period. This time horizon reflects the rapid evolution of AI technologies, shorter model lifecycles, and typical donor funding commitment cycles in LMICs, while capturing both initial deployment costs and ongoing operational expenses. The 3-year window balances the need to amortize fixed costs while acknowledging that AI systems often require significant updates or replacement within this timeframe due to algorithmic advances, changing regulatory requirements, or shifting health system priorities. TCO includes three cost categories:

Fixed costs (one-time): Initial setup, infrastructure buildout, regulatory compliance, system integration, data migration—amortized over 3 years.

Recurring costs (annual): Software licensing, cloud compute/storage, maintenance contracts, support services, connectivity fees.

Variable costs (scale-dependent): Training (per user onboarded), per-transaction inference costs, incremental hardware as volume grows.

> Example TCO Calculation (3-year horizon):
>
> Fixed: $50K setup + $30K infrastructure = $80K ÷ 3 = $26.7K/year Recurring: $18K licensing + $3K cloud + $3K support = $24K/year
>
> Variable: 30 users × 8 hrs × $15/hr = $3.6K training (year 1), ongoing $1K/year retraining
>
> Total annualized TCO: $26.7K + $24K + $2.3K = $53K/year average

Value for Money: Performance-Adjusted Cost Assessment

Higher performance can justify higher cost: Economics reflects value-for-money, not absolute cost minimization. A system that costs 2× more but delivers 3× better outcomes (e.g., 90% sensitivity vs. 70%) may represent superior value. The DAI does not prescribe specific cost-effectiveness thresholds but requires organizations to define "reasonable cost" relative to expected value and budget capacity.

Two common normalization approaches:

Budget-relative approach (organizational sustainability):

Express TCO as % of relevant budget (IT budget, digital health program budget, disease-specific vertical budget). Heuristic: 5% of relevant budget = "reasonable" baseline. Adjust based on strategic priority (higher-priority initiatives may justify 10–15%; lower-priority <5%).

> Cost-effectiveness approach (health programs): An alternative to the above would be to calculate cost per health outcome (cost per case detected, cost per QALY or DALY averted). Compare against willingness-to-pay thresholds:
>
> LMICs: Typically, 1–3× GDP per capita ($500–$3,000/DALY in Tanzania, Kenya) | High-income countries: $50,000–$150,000/QALY (USA, UK NICE guidelines).

Cost Burden

We recommend following the cost burden calculation, which calculates the relative burden attributable to the AI innovation over and above the budgeted cost.

Cost_Burden = max (0, [Actual_TCO ÷ Reasonable_Threshold] - 1)

Funding Sustainability: Who Pays and for How Long?

Economics assesses both affordability (TCO burden) and sustainability (funding security). A system may be affordable today but unsustainable if funding depends on short-term grants or single donors with uncertain renewal. Different payment models carry different sustainability risks.

Payment Model	Description	Sustainability Risk
Cost Recovery	User fees, insurance reimbursement, or service charges cover operating costs	Low – **Self-sustaining revenue model**
Permanent Budget Line	Government operational budget allocation with multi-year commitment	Low – **Institutionalized funding**
Multi-Year Commitment	Donor or program funding committed for 3+ years with established relationship	⚠Moderate – **Stable but time-limited**
Annual Funding	Year-to-year budget approval; single donor with uncertain renewal	✗ High – **Vulnerable to budget cuts**
Pilot-Only Funding	**Grant covers pilot phase only; no post-pilot funding plan**	✗ Very High – **Deployment cliff likely**

Funding Sustainability Score

Assess funding security on a 1–5 scale:

5 (Highly Sustainable): Cost-recovery model (user fees, insurance reimbursement) or permanent government budget line with diversified funding sources

4 (Sustainable): Multi-year committed funding (3+ years); institutional budget with high confidence of renewal; blended funding (government + donor)

3 (Moderate Risk): 2–3 year funding commitment; single donor but established partnership; institutional budget but subject to annual approval

2 (High Risk): Annual funding only; pilot-phase grants with uncertain renewal; single donor with no history of sustained support

1 (Unsustainable): Pilot-only funding with no post-pilot plan; donor exit announced or expected; no alternative funding sources identified

Combined Economics Calculation

The final Economics score combines cost burden (affordability) with funding sustainability (long-term viability):

Step 1: Calculate cost-burden

Cost Burden = (Annualized TCO of innovation ÷ Attributed Budget) – 1

Step 2: Calculate Economic Cost

Economics_cost = 1 ÷ (1 + Cost_Burden)

Step 3: Normalize Funding Sustainability

Funding_Sustainability_normalized = (FS_score - 1) ÷ 4

Step 4: Apply sustainability multiplier

Economics_final = Economics_cost × Funding_Sustainability_normalized

Example scenarios:

Scenario	TCO vs Budget	Econ_cost	Funding Model	FS Score	Econ_final	Interpretation
District hospital, donor pilot	1.2× budget	0.83	1-year grant, no renewal plan	2	0.83 × 0.25 = 0.21	**Affordable now but unsustainable**
National system, budget line	2× budget	0.50	Govt permanent allocation	5	0.50 × 1.00 = 0.50	**Moderate cost, secure funding**
Private clinic, patient pays	3× budget	0.33	Fee-for-service	5	0.33 × 1.00 = 0.33	**High cost but self-sustaining**
Regional program, 3-year donor	1.5× budget	0.67	Multi-year commitment	4	0.67 × 0.75 = 0.50	**Moderate cost, stable funding**
Pilot with exit strategy	**Equal to budget**	**1.00**	**Pilot only, donor leaving**	**1**	**1.00 × 0.00 = 0.00**	**Affordable but no future → blocked**

⚠ **Equity Consideration:** Cost-recovery models (user fees, fee-for-service) may score high on Economics (self-sustaining) but create access barriers for low-income populations. A system that is "deployable" from a sustainability perspective may still fail equity goals.

Recommendation: Organizations using user-pay models should conduct separate equity impact assessments and consider tiered pricing, subsidies for vulnerable populations, or blended funding to ensure equitable access. The DAI measures deployment sustainability, not fairness—both must be assessed.

Why Use a Multiplicative Approach?

Funding sustainability acts as a **multiplier rather than an additive factor** because unsustainable funding creates systemic risk that compounds cost burden. Consider:

Scenario A: Low-cost system (Economics_cost = 1.0) funded by 1-year pilot grant (FS = 2) → Economics_final = 1.0 × 0.25 = 0.25. Affordable but will fail when grant ends.

Scenario B: Moderate-cost system (Economics_cost = 0.5) with permanent budget line (FS = 5) → Economics_final = 0.5 × 1.0 = 0.50. Higher cost but sustainable long-term.
The multiplicative structure ensures that **unsustainable funding penalizes deployment even when costs are low**, reflecting the empirical reality that donor-dependent pilots frequently collapse post-funding regardless of technical success. This prevents the common LMIC failure pattern: "Worked beautifully for 18 months, then vanished when the grant expired."

Smooth Penalty Curve vs. Hard Gate

Unlike Safety (binary gate), Economics uses a **smooth penalty curve** because costs are often negotiable, scalable, or addressable through optimization. Organizations can:

- Negotiate volume discounts or tiered pricing with vendors
- Optimize infrastructure (edge deployment vs. cloud, open-source alternatives)
- Phase deployment to spread fixed costs over time
- Secure additional funding or reallocate budget priorities

Economics reflects a **sliding scale of sustainability risk**, not an absolute deployment blocker. A system with high cost (Economics = 0.3) can still deploy if other factors are strong—but sustainability is at risk without cost optimization.

When Economics Becomes a Hard Gate (LMIC Context)

Important caveat: In resource-constrained settings with inflexible budgets, economics can function as a **de facto hard gate** despite the smooth penalty structure. When:

- Annual budget is fixed by government allocation or donor commitment (no flexibility)
- Costs are non-negotiable (regulatory compliance, infrastructure prerequisites)
- No alternative funding sources exist

> ⚠ **LMIC Budget Reality:** If your TCO exceeds available budget by >50% and funding is non-negotiable, treat Economics as a gate failure regardless of calculated score. The smooth penalty assumes cost flexibility—if that assumption doesn't hold in your context, deployment is blocked until funding or cost structure changes.

> ⚠ A system that exceeds budget may be mathematically deployable (Economics = 0.3 yields DAI > 0) but practically blocked due to absolute affordability constraints. Organizations should flag TCO > 150% of budget as a practical deployment blocker requiring funding solutions before proceeding, even if the DAI calculation technically permits pilot deployment.

4. Friction — The Reality Tax

Question: How painful is deployment?

Friction captures the organizational and technical drag that slows or derails implementation, even when the AI itself is excellent. This includes usability challenges that increase training burden, cognitive load, and user resistance. Two components:

Workflow Friction

- **Training burden:** Hours of training required per user to achieve competency
- **Usability and cognitive load:** Interface complexity, task completion time, mental effort required, user satisfaction with system design
- **Workflow redesign:** Percentage of existing workflows requiring modification to accommodate AI
- **Approval cycles:** Number of governance/approval layers needed for deployment authorization
- **Cultural resistance:** Change fatigue, skepticism, fear of job displacement, or perceived threat to professional autonomy

Integration Complexity

- Age and architecture of legacy systems (paper-based, outdated EMRs, disconnected databases)
- Number of system integrations required (EMR, PACS, LIS, billing, reporting, DHIS2)
- Data pipeline challenges (format mismatches, missing fields, quality issues, incomplete metadata)
- Infrastructure gaps (connectivity, compute capacity, storage, power reliability)

> **Note on Usability:** Poor usability manifests as deployment friction—higher training burden, increased cognitive load, longer task completion times, and greater user resistance. While usability issues may also affect real-world outcomes (systems that are difficult to use may be used incorrectly or inconsistently), the DAI captures usability explicitly as a Friction factor to diagnose deployment barriers.

Formula:

Friction = 1 + (0.5 × Workflow_Friction + 0.5 × Integration_Complexity)

Where Workflow_Friction and Integration_Complexity are each normalized 0–1 (higher = worse)

- Base friction = 1.0 (no extra difficulty beyond inherent task complexity)
- Maximum friction = 2.0 (deployment twice as hard as baseline)

Friction appears in the denominator, so it reduces your DAI score proportionally. High friction doesn't make deployment impossible—it just makes it slower, more expensive, and more likely to fail without sustained effort.

Putting It All Together: The Complete DAI Formula

DAI = (Core Readiness × Safety × Economics) / Friction

*With stage-specific safety gates that can set DAI = 0 if minimum thresholds aren't met.

Step-by-step calculation:

1. Assess each of the eight factors on a 1–5 scale using structured rubrics
2. Normalize scores to 0–1 range: normalized = (score - 1) ÷ 4
3. Calculate Core Readiness = $\sqrt[4]{\text{Purpose} \times \text{People} \times \text{Process} \times \text{Performance}}$
4. Calculate Safety = min(Protection, Reliability)
5. Check stage gates: If Core Readiness factors or Safety components below threshold → DAI = 0
6. Calculate Economics = 1 ÷ (1 + Cost_Burden)
7. Calculate Friction = 1 + (0.5 × Workflow + 0.5 × Integration)
8. Calculate DAI = (Core Readiness × Safety × Economics) / Friction

> Note: We define exactly how we turn 1–5 ratings into 0–1 scores and combine them in our detailed scoring rubric in the technical appendix, which ensures consistent assessment across different organizations and country contexts while keeping this framework accessible.

Handling Uncertainty in DAI Scoring

Real-world assessments involve measurement uncertainty stemming from incomplete evidence, subjective judgment, and variability among evaluators. The DAI addresses this through structured protocols for scoring ranges and sensitivity analysis.

Scoring as Ranges (When Evidence Is Incomplete)

When evidence is insufficient to assign a precise score, we recommend expressing factors as ranges (e.g., "Process is 2.5–3.0").

Calculate DAI using both bounds:

- **Lower bound DAI:** Use minimum score for each uncertain factor → conservative estimate
- **Upper bound DAI:** Use maximum score for each uncertain factor → optimistic estimate

Report DAI as range: "DAI = 0.15–0.22" indicates pilot-ready with uncertainty

> **Example:**
>
> Purpose: 3.5–4.0 (strong but baseline data incomplete)
>
> People: 2.5–3.0 (training underway, final competency unknown) Process: 2.2 (precise: workflows documented and assessed) Performance: 3.5 (precise: pilot validation complete)
>
> **Lower bound:** Use Purpose 3.5, People 2.5 → DAI_{min}
>
> **Upper bound:** Use Purpose 4.0, People 3.0 → DAI_{max}

Decision rules for uncertain scores:

Gate check: If lower bound passes gate, system is robustly above threshold; if upper bound fails, robustly below; if bounds straddle gate, collect more evidence before deciding.

Deployment decision: If DAI range entirely within one stage (e.g., 0.18–0.22 = pilot-ready), proceed; if range spans stages (e.g., 0.19–0.31 = pilot to limited rollout), treat as lower stage until uncertainty resolves.

Applied Example: TB Screening AI in a District Hospital

Context: A 200-bed district hospital in the rural region of an LMIC serving a catchment population of 500,000. Annual operational budget: $2.4 million. Local TB prevalence: 295 per 100,000 population. Hospital diagnoses ~600 TB cases per year through passive case finding.

AI System: Chest X-ray computer-aided detection (CAD) for TB screening. Cloud-based inference. Vendor claims 90% sensitivity, 85% specificity on validation data.

Costs:

- Setup Costs for on time integration in system: $25,000
- Annual licensing: $18,000
- Cloud inference (estimated 5,000 screens/year): $3,000
- Training (30 clinical staff × 8 hours × $15/hour): $3,600
- Local IT support (0.25 FTE): $3,000
- **Total annual cost: $27,600**

Assessment Using DAI

P1 - Purpose (Score: 4 → 0.75 normalized)

- ✓ Strong alignment: National TB Strategic Plan target is 90% case detection by 2028 (currently 65%)
- ✓ Clear metrics: Increase TB case detection by 20%, reduce diagnostic delay from 21 to 10 days
- ✓ Executive sponsorship: District Medical Officer champion, donor funding committed for 3 years
- ⚠ Minor gap: Success metrics defined but baseline data collection incomplete

P2 - People (Score: 3 → 0.50 normalized)

- ⚠ Training incomplete: Only 12 of 30 staff trained (40%); high turnover among radiographers (18-month average tenure)
- ✓ Roles defined: RACI matrix complete, escalation path to regional referral hospital established
- ✗ Culture resistance: Clinicians skeptical of AI accuracy, fear of liability if AI misses cases
- Bottleneck: Staff readiness and change acceptance

P3 - Process (Score: 2 → 0.25 normalized)

- ✗ Workflows largely undocumented: 70% of TB diagnostic pathway is informal/paper-based; basic SOPs exist but not followed consistently
- ✗ No EMR: Patient tracking via paper registers; X-ray results manually transcribed into log books
- ⚠ Governance nascent: Monthly quality review meetings started 3 months ago but action tracking inconsistent
- Critical weakness: Low process maturity substantially reduces Core Readiness via geometric mean penalty

P4 - Performance (Score: 4 → 0.75 normalized)

- ✓ Independent validation: Pilot with 200 local patients showed 88% sensitivity, 83% specificity (close to vendor claim of 90%/85%)
- ✓ Adequacy: 88% sensitivity exceeds WHO TB screening guideline (≥85% required)
- ✓ Stability: Performance consistent across age groups and HIV status subgroups
- Assessment: Technology validated and adequate for intended clinical task

SP - Protection (Score: 3.0 → 0.50 normalized)

- ✓ Compliance: Meets Country Data Protection Act requirements; patient consent process defined
- ⚠ Security: Cloud vendor SOC 2 certified, but no local penetration testing; basic access controls (password-only, no MFA)
- ✓ Governance: Incident response plan documented, tested annually
- Assessment: → **GATE CLEARED**, but needs strengthening for scale

SR - Reliability (Score: 3 → 0.50 normalized)

- ✓ Uptime: 96% availability during pilot phase; edge device with offline capability deployed to handle power outages
- ✓ Accuracy stability: No drift observed in 6-month pilot; monthly validation shows consistent performance
- ⚠ Monitoring: Basic automated alerting configured; manual review of performance logs monthly; drift detection being refined
- Assessment: → **GATE CLEARED**

Economics (EC - Cost component = 0.25; FS - Funding sustainability = 0.50)

- 3-year TCO calculation:
  - Fixed costs (amortized): Setup $15K, integration $10K → $25K ÷ 3 years = $8,333/year
  - Recurring annual costs = $27,600K/year
  - **Total annualized TCO: $35,933/year**
- Budget benchmark: TB program budget = $180K/year; 5% reasonable = $9K
  - Cost burden: ($35,933 ÷ $9,000) - 1 = 2.99
  - Economics_cost = 1 ÷ (1 + 2.99) = 0.25 (75% penalty for being ~4× reasonable cost)
- **Funding sustainability:** 3-year donor grant with no renewal commitment (Score: 3 → Moderate risk)
  - Funding_sustainability_normalized = (3-1) ÷ 4 = 0.50
- **Economics_final = 0.25 × 0.50 = 0.125** (affordable short-term but sustainability uncertain)

WF - Workflow Friction (Score: 4 → 0.75 normalized)

- ✗ Training burden: 8 hours per user, high turnover requires continuous retraining (score: 3)
- ✗ Redesign: 60% of diagnostic workflow requires change (manual → digital, new decision points) (score: 4)
- ✗ Approval cycles: Required approvals from District Health Management Team, Regional Health Secretariat, Ministry ICT division, ethics committee (4 levels, 6 months) (score: 5)
- High organizational friction

IC - Integration Complexity (Score: 4 → 0.75 normalized)

- ✗ Legacy: Paper-based systems, no EMR, DHIS2 reporting disconnected (score: 4)
- ⚠ Integrations: Must connect to PACS (not yet implemented), DHIS2, lab system (score: 4)
- ⚠ Data: Single source (X-ray images), but quality variable (old machine, inconsistent technique) (score: 3)
- High technical friction

Calculation

Core Readiness = geometric mean (P1, P2, P3, P4) = $\sqrt[4]{}$(0.75 × 0.50 × 0.25 × 0.75) = $\sqrt[4]{}$(0.070313) = 0.515

Safety = min (SP, SR) = min(0.50, 0.50) = 0.50

Economics = EC x FS = 0.25 x 0.5 = 0.125

Friction = 1+ 0.5*(WF+IC) = 1 + (0.5 × 0.75 + 0.5 × 0.75) = 1 + 0.75 = 1.75

Result: DAI = (0.516 × 0.50 × 0.125) ÷ 1.75 ≈ 0.018 (Pilot-ready)

Both safety gates pass (Protection 3.0, Reliability 3.0 meet pilot threshold ≥3.0), so deployment is NOT blocked. However, the low Process score (2.0 → normalized 0.25) substantially reduces Core Readiness through the geometric mean: CR = $\sqrt[4]{}$(0.75 × 0.50 × 0.25 × 0.75) = 0.516 (instead of 0.75 if all factors were equal). This demonstrates how the geometric mean naturally penalizes weak factors without requiring hard gate cutoffs.

Interpretation: The AI system can proceed to pilot deployment (safety adequate), but the low DAI score (0.019) signals severe constraints: Process immaturity (2.0) drags down readiness, Economics (0.13) threatens sustainability (~4× reasonable budget), and high Friction (1.75) slows adoption. The system is deployable but requires systematic improvements to succeed.

Diagnosis and Action Plan

Critical Constraints:

1. **Process (2.0 → normalized 0.25):** Weakest Core Readiness factor; 70% undocumented workflows, no EMR, nascent governance. Geometric mean penalty reduces CR from potential ~0.69 to actual 0.516 (−25% due to this single weak factor)
2. **Economics (0.125):** Cost ~4× reasonable budget allocation AND 3-year donor funding with no renewal plan; threatens long-term sustainability
3. **Friction (1.75):** High workflow redesign burden (60%), complex integration needs (4 systems), high approval overhead (4 layers) compound to slow adoption

Strengths:

- **Purpose (4.0):** Strong strategic alignment, clear metrics, executive sponsorship
- **Performance (4.0):** Validated technology with 88% sensitivity meeting clinical thresholds
- **Safety (Protection 3.0, Reliability 3.0):** Both factors meet pilot gate thresholds; adequate security and uptime for pilot deployment

Recommended Action Sequence

The roadmap should focus on addressing the weak Process factor, then strengthening economics and reducing friction:

Why This Formula Works

It Mirrors Real Failure Modes

Failure Pattern	How DAI Captures It
Weak link kills deployment	**Geometric mean in Core Readiness penalizes imbalance**
Safety is non-negotiable	**Binary gate: below threshold → DAI = 0**
Cost compounds over time	**Smooth penalty curve reflects sustainability risk**
Friction slows everything	**Divisor structure: 2× friction → ½ DAI**

It Balances Optimism and Realism

Numerator (Readiness × Safety × Economics) = What you've built "on paper" Denominator (Friction) = Real-world barriers "in practice".

This prevents the classic failure mode: "Looks great on paper, fails in practice." If your numerator is strong but friction is high, DAI correctly predicts deployment difficulty.

It's Actionable

Unlike multidimensional scorecards, DAI tells you:

✓ What's blocking you → Which gate failed

✓ What's limiting you → Weakest component in the numerator

✓ What to fix next → Improve the bottleneck

✓ When you're ready → Score crosses stage threshold

Example diagnosis: "We cannot deploy because Reliability is 2.6 and we need 3.0 — specifically, uptime is the problem (92% vs. 95% required). Fix that first, then address Process gaps."

Three Embedded Principles

Principle 1: Balance Beats Perfection

A system with all factors at 3.5 (balanced, solid) will outscore a system with three 5.0s and one 1.0 (unbalanced disaster).

Example:

- System A: Purpose 3.5, People 3.5, Process 3.5, Performance 3.5 → Core Readiness = 3.5 (0.625 normalized)
- System B: Purpose 5.0, People 5.0, Process 5.0, Performance 1.0 → Core Readiness = 2.66 (0.415 normalized)
- System A wins, even though B has three perfect scores, because the disaster in Performance kills deployment. This reflects reality: you cannot compensate for one critical gap by excelling elsewhere.

Principle 2: Safety Is Binary, Capability Is Continuous

If your capability (Purpose, People, Process, Performance) is moderately weak, you can still deploy at small scale with close monitoring. But if your safety (Protection, Reliability) is below threshold, you cannot deploy at all, not even for pilots.

Why? Regulatory and ethical norms.

Medical devices don't get "partial approval" or "deploy but be careful" clearance. Either they meet safety standards, or they don't ship. DAI enforces this through binary gates.

Principle 3: Every Context Is Different

$50,000 annual cost means different things to:

- A $2M district hospital (2.5% of budget → high burden)
- A $500M national health system (0.01% of budget → trivial)

Similarly, 95% uptime is:

- Acceptable for a pilot in a research setting
- Inadequate for a diagnostic tool in a referral hospital (needs 99%+)

40% user adoption is:

- Fine for an optional clinical decision support tool
- Disastrous for a compliance-mandated reporting system

DAI handles this through context-specific normalization of costs, stage-appropriate gate thresholds, and user-defined "reasonable" baselines. There is no universal "good" score—only "good enough for this context and stage."

Design Assumptions and Limitations

What the Model Assumes

Measurement validity: Assumes 1–5 Likert scales can be assessed consistently with structured rubrics. In practice, scoring introduces subjectivity (is this a 3.4 or 3.6?). We address this through scorer calibration, documenting assumptions, and treating scores as bands rather than precise points.

Factor independence: Treats eight factors as separable, though they are correlated (e.g., strong Purpose often predicts strong Process). The model enables diagnosis ("Fix Process, then People") while acknowledging interactions. Geometric structure prevents full substitution (can't offset Process failure with Purpose excellence).

Multiplicative error: Geometric mean + multiplicative numerator amplify measurement error. If you underestimate each factor by 10%, the compound error in DAI is larger. This reflects reality: failures are concentrated, not distributed. Small miscalibrations in foundational areas cascade into deployment failures.

Smooth cost penalty: Assumes costs are partially addressable through optimization, negotiation, or scale economies. Some costs are fixed (regulatory compliance, required infrastructure); others are reducible (licensing, cloud compute). The penalty curve averages these, which may underweight fixed costs in resource-constrained LMICs.

What DAI Does NOT Capture

Limitation	Implication
Regulatory compliance	**High DAI ≠ FDA/CE Mark approval. Protection scores reflect security and governance capability, not formal certification. Separate regulatory processes required.**
Political economy	**Ministerial opposition, procurement corruption, donor policy shifts, or geopolitical factors can override technical readiness. DAI assesses system capability, not political feasibility.**
External shocks	**Point-in-time assessment. Pandemics, funding cuts, currency devaluation, or infrastructure collapse can invalidate scores overnight. Requires periodic reassessment.**
Context thresholds	**Model doesn't prescribe universal "reasonable cost." Users must define via stakeholder consensus, budget analysis, or willingness-to-pay studies for their specific context.**
Equity and access	**High DAI doesn't ensure equitable deployment (e.g., urban-only rollout, elite facility concentration). Requires separate equity impact analysis and deliberate inclusive design.**

Positioning

DAI is complementary to existing frameworks:

- Cost-effectiveness analysis (CEA) and health technology assessment (HTA) for economic evaluation
- Regulatory compliance and certification processes for safety and efficacy
- Change management and stakeholder engagement for implementation
- AI ethics frameworks (WHO FPAR, IEEE EAD) for responsible innovation

DAI answers one specific question: "Should we deploy now, and if not, what must improve first?"

It does not replace comprehensive evaluation but provides a structured decision gate within the deployment lifecycle.

This framework is a structured heuristic pending empirical validation. The next research phase requires:

✓ Prospective field testing across diverse contexts (high-income, LMIC, public, private)

✓ Validation of gate thresholds (do systems with DAI < 0.2 actually fail more often?)

✓ Calibration of factor weights (are all eight equally important, or context-dependent?)

✓ Predictive validity testing (do DAI scores at time T predict deployment success at T+1?)

✓ Refinement of cost and friction normalization approaches for resource-constrained settings

Call to Action

We invite health system implementers, researchers, and policymakers to pilot the DAI in their AI deployment planning, document lessons learned, and contribute to the evidence base that will refine this model into a validated, globally applicable standard.

Conclusion

Core Contribution

The Deployable AI Index offers health system leaders, policymakers, and implementation teams a structured model for navigating the gap between AI capability and AI deployment readiness. By synthesizing readiness, safety, economics, and friction into a single score with stage-appropriate gates, DAI translates multidimensional assessments into actionable deployment decisions.

The Formula in One Sentence

Your AI is deployable when you've prepared the fundamentals in balance, built a safe and affordable system, and can deploy it without excessive organizational or technical pain—and the DAI quantifies exactly that.

Appendix 1: The Eight Factors

Factor	Type	Role	If Weak
Purpose	Readiness	Numerator (balanced)	**Drags down Core Readiness → lower DAI; deployment lacks strategic anchor**
People	Readiness	Numerator (balanced)	**Drags down Core Readiness → lower DAI; users cannot operate effectively**
Process	Readiness	Numerator (balanced)	**Drags down Core Readiness → lower DAI; workflows incompatible with AI**
Performance	Readiness	Numerator (balanced)	**Drags down Core Readiness → lower DAI; does not deliver promised value**
Protection	Safety	Numerator (gated min)	**Below gate → DAI = 0; security/privacy risk unacceptable**
Reliability	Safety	Numerator (gated min)	**Below gate → DAI = 0; cannot be trusted for clinical decisions**
Economics	Constraint	Numerator (smooth penalty)	**Reduces DAI gradually as cost rises; threatens long-term sustainability**
Friction	**Constraint**	**Denominator (divisor)**	**Increases deployment difficulty → reduces DAI proportionally; slows adoption, raises cost**

Resources and Contact

Contact Information:

Anunaya Jain
Global Technical Director, Digital and Data Analytics Hub (DIGIT)

Jhpiego
Johns Hopkins University Affiliate
E: anunaya.jain@jhpiego.org

Technical Appendix

Normalization and Calculation Formulas

Likert to normalized score: normalized_score = (likert_score - 1) ÷ 4 where likert_score $\in \{1, 2, 3, 4, 5\}$

Core Readiness:
CR = (Purpose × People × Process × Performance)^(1/4) where all factors $\in [0,1]$

Safety:
S = min (Protection, Reliability) where Protection, Reliability $\in [0,1]$

Economics:
C_norm = max (0, (Actual_Cost ÷ Reasonable_Threshold) - 1) E = 1 ÷ (1 + C_norm)

Friction:
F = 1 + (0.5 × WF + 0.5 × IC)
where WF = Workflow_Friction $\in [0,1]$ IC = Integration Complexity $\in [0,1]$

Deployable AI Index:
if any gatecondition fails:
DAI = 0
else:
DAI = (CR × S × E) ÷ F

Suggested Data Inputs for DAI Assessment

The tables below provide **guidance on data inputs** that can inform each factor assessment. **These are suggestions, not requirements.** In practice, use whatever data is readily available in your context. The goal is to enable defensible scoring with reasonable effort—ease of assessment should trump perfection. A "good enough" assessment completed quickly is more useful than a perfect assessment that never happens.

Pragmatic approach: For each factor, identify which data sources you already have (existing reports, system logs, staff records, project documentation) and use those to assign 1–5 scores. If you lack data for a factor, use informed judgment with clearly documented assumptions. If a score is uncertain or contentious, then—and only then—invest effort in collecting additional evidence for that specific factor.

Purpose — Strategic Alignment Data

Data Input	Definition	Acceptable Sources	Minimum Standard
Health system problem definition	Specific, measurable gap that AI addresses (e.g., TB diagnostic delay >21 days, case detection rate 65% vs. 90% target)	Disease burden reports, health facility data, national health strategic plans, epidemiological surveillance	**Quantified baseline with timeframe**
Strategic alignment documentation	Evidence that AI solution addresses a documented health system priority	National/regional health strategic plans, Ministry of Health priorities, facility strategic plans, donor funding priorities	**Explicit reference in strategy document OR executive mandate**
Success metrics	Defined, measurable outcomes with targets (e.g., reduce delay from 21→10 days, increase detection by 20%)	Project proposal, M&E framework, performance contract, donor agreement	**≥2 quantified metrics with baseline and target values**
Executive sponsorship	**Documented leadership commitment with resource allocation authority**	**Signed commitment letter, budget allocation, RACI matrix with named executive sponsor, project charter**	**Named sponsor with budget approval authority**

People — Human Capacity Data

Data Input	Definition	Acceptable Sources	Minimum Standard
Training completion rate	Percentage of target users (clinical, technical, administrative) who have completed required AI system training	Training records, attendance registers, competency assessments, learning management system data	**Documented training records for ≥80% of intended users**
Staff turnover rate	Annual attrition rate for key user roles (higher turnover increases training burden and continuity risk)	HR records, facility staffing reports, payroll data	**12-month rolling average**
RACI matrix	Documented roles and accountability (Responsible, Accountable, Consulted, Informed) for AI system operation, maintenance, escalation	Project documentation, governance charter, organizational chart with named individuals	**Complete RACI for ≥3 critical processes (operation, troubleshooting, escalation)**
Change readiness indicators	**Evidence of organizational culture supporting change (past successful innovations, staff engagement surveys, resistance assessment)**	**Staff surveys, prior project success cases, change management assessments, stakeholder engagement logs**	**≥1 documented example of successful change initiative OR staff survey with >60% positive readiness**

Process — Workflow Maturity Data

Data Input	Definition	Acceptable Sources	Minimum Standard
Workflow documentation	Standard operating procedures (SOPs) for clinical/administrative processes that AI will augment or modify	Written SOPs, process maps, clinical protocols, approved workflow diagrams	**≥70% of affected workflows documented and approved**
EMR/HIS status	Electronic medical record or health information system availability and maturity (paper-based, basic digital, integrated EMR)	System inventory, IT architecture documentation, user access logs, integration architecture	**Digital data capture for ≥50% of relevant data OR documented integration plan**
Governance cadence	Regular, documented review meetings for quality, performance monitoring, issue escalation (frequency, attendance, action tracking)	Meeting minutes, governance charter, quality improvement logs, action item tracking system	**Monthly meetings with documented minutes for ≥3 months**
Continuous improvement mechanisms	**Evidence of PDSA (Plan-Do-Study-Act) cycles, feedback loops, process refinement based on data**	**PDSA documentation, quality improvement project records, before/after process comparisons, user feedback logs**	**≥1 completed PDSA cycle with documented improvement**

Performance — Technical Validation Data

Data Input	Definition	Acceptable Sources	Minimum Standard
Vendor-claimed performance	Published model performance metrics (accuracy, sensitivity, specificity, AUC-ROC, F1 score) on test datasets	Vendor documentation, peer-reviewed publications, model cards, regulatory submission documents	**Published metrics with dataset description and sample size**
Independent validation results	Performance assessment on local data with representative patient population, infrastructure, and workflows	Pilot study reports, validation study publications, internal evaluation data with statistical analysis	**n ≥ 100 local cases with sensitivity/specificity + 95% CI**
Clinical adequacy threshold	Minimum performance required for intended clinical task (e.g., WHO guideline: TB screening ≥85% sensitivity)	Clinical guidelines (WHO, national protocols), regulatory standards, clinical expert consensus, published benchmarks	**Documented threshold from authoritative source**
Subgroup performance	Model performance across relevant subgroups (age, sex, comorbidities, disease severity) to detect disparities	Disaggregated validation results, fairness/bias assessment reports, subgroup analyses	**Performance reported for ≥2 clinically relevant subgroups**
Model stability evidence	**Performance consistency over time periods, data batches, or operating conditions (demonstrates lack of drift)**	**Longitudinal validation data, drift monitoring reports, temporal performance analysis**	**≥2 time points OR ≥2 data batches with consistent performance (variance <10%)**

Protection — Security & Privacy Data

Data Input	Definition	Acceptable Sources	Minimum Standard
Encryption status	Data encryption at rest (stored data) and in transit (network transmission) with encryption standards documented	Security audit reports, system architecture documentation, vendor security specifications, penetration test results	**AES-256 (or equivalent) for data at rest; TLS 1.2+ for data in transit**
Access control mechanisms	Multi-factor authentication (MFA), role-based access control (RBAC), principle of least privilege implementation	Access control policy, user access logs, authentication system configuration, security audit trails	**MFA enabled for ≥80% of administrative users; RBAC with audit logs**
Regulatory compliance evidence	Documentation of compliance with applicable data protection regulations (GDPR, HIPAA, local data protection acts)	Data protection impact assessment (DPIA), legal compliance certificates, regulatory approval letters, privacy policy	**Completed DPIA OR legal compliance certification**
Incident response capability	Documented incident response plan with defined procedures, responsible parties, communication protocols, tested exercises	Incident response plan document, test/drill reports, incident logs, escalation procedures	**Documented plan tested within past 12 months**
Penetration testing results	**Independent security assessment identifying vulnerabilities and remediation status**	**Penetration test reports, vulnerability assessment results, remediation tracking logs**	**Test within past 12 months with critical vulnerabilities resolved**

Reliability — System Stability Data

Data Input	Definition	Acceptable Sources	Minimum Standard
System uptime percentage	Percentage of time system is operational and accessible (excludes scheduled maintenance windows)	Server logs, uptime monitoring reports (e.g., Pingdom, UptimeRobot), cloud provider SLA reports, incident logs	**≥3 months continuous monitoring data**
Model drift metrics	Change in model performance over time measured via continuous monitoring (accuracy degradation, prediction distribution shifts)	Model monitoring dashboards, drift detection alerts, performance tracking logs, retraining trigger events	**Monthly drift assessments with <5% performance degradation**
Infrastructure reliability	Network connectivity stability, power availability, backup systems, failover capability	Network monitoring logs, power outage records, generator logs, backup system test results, disaster recovery plans	**Documented infrastructure baseline + redundancy plan**
Automated monitoring status	Real-time alerting for system failures, performance anomalies, security incidents with defined thresholds and response workflows	Monitoring system configuration, alert logs, incident tickets, escalation procedures, automated testing results	**Automated alerts configured with <15 minute detection time**
Mean time to recovery (MTTR)	**Average time from system failure detection to restoration of full functionality**	**Incident management logs, service desk tickets, downtime analysis reports**	**≥3 incidents documented with MTTR calculation**

Economics — Cost & Funding Data

Data Input	Definition	Acceptable Sources	Minimum Standard
Total Cost of Ownership (TCO) — Fixed costs	One-time costs: initial setup, infrastructure buildout, regulatory compliance, system integration, data migration (amortized over 3 years)	Vendor contracts, procurement records, project budgets, IT infrastructure invoices, consulting fees	**Itemized cost breakdown with supporting documentation**
TCO — Recurring costs	Annual ongoing costs: software licensing, cloud compute/storage, maintenance contracts, technical support, connectivity fees	Annual contracts, subscription invoices, cloud billing statements, support service agreements, telecom bills	**12-month cost projection with historical data (if available)**
TCO — Variable costs	Scale-dependent costs: training per user ($, hours), per-transaction inference costs, incremental hardware as volume grows	Training cost calculations, usage-based billing data, projected user/transaction volumes, hardware scaling estimates	**Cost-per-unit calculation with volume assumptions**
Budget capacity	Relevant organizational budget against which TCO is normalized (IT budget, program budget, operational budget, or total facility budget)	Approved annual budget documents, financial statements, donor grant agreements, government budget allocations	**Documented annual budget with authority confirmation**
Cost-effectiveness data (optional)	Health outcome costs: cost per case detected, cost per QALY/DALY averted, cost per life saved (for CEA approach)	Health economic modeling, pilot outcome data, published CEA studies for comparable interventions	**Credible estimate with assumptions documented OR pilot data (n ≥ 50)**
Willingness-to-pay threshold (optional)	Maximum acceptable cost per health outcome for the decision-making context (LMIC: 1-3× GDP per capita; HIC: $50-150K/QALY)	National HTA guidelines, WHO cost-effectiveness thresholds, published economic evaluations, stakeholder consensus	**Documented threshold from authoritative source OR stakeholder consensus process**
Funding source and duration	Documented funding commitment: source (government, donor, user fees), amount, duration, renewal conditions	Grant agreements, budget approval letters, memoranda of understanding, revenue projections (for cost-recovery models)	**Signed funding commitment document with ≥1 year duration**
Funding sustainability score	**1-5 rating based on funding model: 1=pilot-only, 2=annual budget, 3=2-3 year commitment, 4=multi-year/blended, 5=permanent/cost-recovery**	**Funding agreement analysis, institutional budget classification, revenue stream assessment**	**Evidence supporting funding model classification**

Friction — Deployment Difficulty Data

Data Input	Definition	Acceptable Sources	Minimum Standard
Training burden	Hours of training required per user to achieve competency; total training investment (users × hours × cost per hour)	Training curriculum, competency assessment data, time-tracking logs, training completion records	**Documented training hours per role with competency criteria**
Usability metrics	Interface complexity, task completion time, cognitive load assessment, user satisfaction scores (SUS, PSSUQ, custom surveys)	Usability testing reports, user surveys, task time observations, cognitive walkthrough assessments, user feedback logs	**Usability assessment with ≥15 representative users OR SUS score**
Workflow redesign scope	Percentage of existing workflows requiring modification to accommodate AI; number of workflow touch-points affected	Workflow analysis, process maps (before/after), change impact assessment, SOPs requiring revision	**Workflow inventory with % requiring change quantified**
Approval/governance layers	Number of organizational approval cycles required for deployment authorization (committees, authorities, procurement stages)	Organizational governance structure, procurement process documentation, approval workflow charts, historical approval timelines	**Documented approval pathway with average timeline**
Cultural resistance indicators	Staff concerns about AI (job displacement fears, professional autonomy, trust in AI accuracy); change fatigue; stakeholder opposition	Stakeholder surveys, focus group reports, change readiness assessments, resistance pattern documentation, union feedback	**Stakeholder engagement record with resistance level (low/medium/high)**
Legacy system age/architecture	Age of existing IT infrastructure; architecture type (paper-based, standalone systems, siloed databases, integrated digital); technical debt	IT asset inventory, system architecture diagrams, integration maturity assessments, technical debt analysis	**System inventory with deployment year and integration capability**
Required integrations count	Number of external systems AI must connect to (EMR, PACS, LIS, billing, DHIS2, etc.); API availability; data exchange standards	Integration architecture documentation, API specifications, data exchange agreements, HL7/FHIR implementation guides	**Integration inventory with API status (available/requires development)**
Data pipeline complexity	Data format mismatches, missing required fields, data quality issues, metadata completeness, interoperability barriers	Data quality assessments, ETL (extract-transform-load) logs, data completeness reports, field mapping documents	**Data quality report with % completeness for ≥5 critical fields**
Infrastructure gaps	**Network connectivity reliability, compute capacity adequacy, storage availability, power stability, environmental conditions**	**Infrastructure assessments, network speed tests, server capacity reports, power outage logs, facility assessments**	**Infrastructure baseline, assessment with gap identification**

Data Quality Principles:

- **Recency:** Data should be ≤12 months old for dynamic factors (Performance, Reliability, Economics) and ≤24 months for stable factors (Purpose, Process documentation)
- **Representativeness:** Performance data must reflect local patient populations, infrastructure, and workflows—not only benchmark datasets
- **Verifiability:** All data inputs should be documented with sources, dates, and responsible parties for audit trail purposes
- **Sufficiency:** Minimum data standards (sample sizes, time periods, coverage rates) are deployment-stage dependent; pilot requires less rigorous evidence than full production

FOR OPEN CIRCULATION

2026