# AI-Enabled Digital Public Infrastructure: Responsible AI Governance and DPI Safeguards - Revised edition 1.1, November 2026

Independent Technical Working Paper

Version 1.1 | November 22, 2026

Nasir Rehman Shaikh, Independent Advisor, AI Governance and Human Oversight, Dubai, United Arab Emirates

---

AI-ENABLED DIGITAL PUBLIC INFRASTRUCTURE

An Operational Framework for Responsible AI Governance, Human Oversight and

DPI Safeguards

www.nasirrehmanshaikh.com

## Purpose

An independent practitioner contribution examining how responsible AI governance can complement DPI safeguards as AI becomes embedded within population-scale digital infrastructure.

## Status

This is an independent publication. It is not an official publication of the United Nations, UNDP, the DPI Safeguards Initiative or any other international organisation.

## Abstract

Digital Public Infrastructure (DPI) is becoming foundational to digital identity, payments, data exchange and service delivery at population scale, while artificial intelligence moves from standalone applications into the systems and decision processes that operate on top of that infrastructure. This working paper examines the governance question created by this convergence: how AI can be integrated into DPI without weakening the safety, inclusion, agency, accountability and trust that DPI requires. Positioned as a complementary operational layer to the Universal DPI Safeguards Framework rather than a replacement, the paper proposes a six-layer AI-DPI governance model covering purpose and public value, risk and impact assessment, human oversight architecture, technical and procurement controls, accountability and redress, and continuous monitoring and learning; a four-level AI-DPI risk classification; a mapping of DPI safeguard areas to operational AI controls; an AI-DPI Control

Loop running Identify, Assess, Design, Test, Approve, Deploy, Monitor, Intervene, Learn; and a deployment readiness assessment. The framework is conceptual: it is grounded in established international instruments but has not been empirically validated. Its objective is managed, visible, accountable and continuously improving risk rather than zero risk.

## Executive Summary

Digital Public Infrastructure (DPI) is becoming foundational to how societies establish digital identity, make payments, exchange data and deliver public and private services at population scale. At the same time, artificial intelligence is moving from standalone applications into the systems, workflows and decision processes that operate on top of digital infrastructure.

This convergence creates an important governance question: How can AI be integrated into DPI without weakening the safety, inclusion, agency, accountability and trust that DPI requires?

The Universal DPI Safeguards Framework provides a rights-based and practical foundation for addressing this challenge. This paper does not seek to replace or reproduce that framework. Instead, it proposes an additional operational layer focused specifically on the governance of AI when AI systems interact with or operate on top of DPI.

Central proposition: AI governance should become an operational component of DPI safeguards wherever AI can materially influence access, decisions, services, data, security, economic participation or public trust.

The proposed framework introduces six components: AI-DPI impact assessment; risk-proportionate governance; human oversight architecture; technical and procurement controls; accountability and redress; and continuous monitoring, measurement and learning.

These components operate across five DPI lifecycle stages: conception and scoping; strategy and design; development; deployment; and operations and maintenance. The paper also introduces an AI-DPI Control Loop:

**Identify → Assess → Design → Test → Approve → Deploy → Monitor → Intervene → Learn.**

The objective is not zero risk. The objective is managed, visible, accountable and continuously improving risk.

## 1. The Emerging AI-DPI Convergence

The first generation of digital transformation focused primarily on digitising processes. The second connected systems, institutions, people and data. The emerging phase is increasingly characterised by intelligent systems that can interpret information, generate content, make recommendations and perform actions.

AI can become a decision-support layer, service-delivery layer, fraud-detection layer, cybersecurity layer, identity-verification layer, data-analysis layer, automation layer or autonomous action layer.

The opportunity is significant. AI can improve public-service efficiency, detect anomalous transactions, assist officials, improve accessibility and support better allocation of scarce resources.

But the risks change when AI operates through infrastructure used by millions of people. A model error in an isolated enterprise application may affect a department. A model error integrated into foundational DPI may affect an entire population.

### Principle: The governance intensity of AI should increase as the scale, autonomy, interconnectedness and consequence of the system increase.

## 2. Relationship with the Universal DPI Safeguards Framework

The Universal DPI Safeguards Framework is a multi-stakeholder framework intended to guide safe and inclusive DPI implementation and to address risks at individual and societal levels. This paper is deliberately positioned as a complementary AI governance layer, not a competing framework.

The broader DPI safeguards architecture establishes the foundation. This paper asks what additional controls are required when AI becomes embedded within that architecture.

### DPI Safeguards → AI-DPI Risk Assessment → AI Governance Controls → Technical Implementation →

### Monitoring & Evidence → Human Accountability & Redress

The DPI Safeguards Initiative describes its approach as minimising risks across technical, normative and organisational layers while making implementation secure, inclusive, practical and adaptable. The Initiative also describes a Discover → Design → Implement → Evolve process, reinforcing the importance of learning from implementation. (DPI Safeguards Initiative, 2026).

## 3. The Six-Layer AI-DPI Governance Model

### Layer 1: Purpose and Public Value

Before introducing AI, establish the problem, beneficiaries, potential harms, public value, alternatives and consequences of failure. AI should not be introduced simply because it is technologically available.

### Layer 2: Risk and Impact Assessment

Assess individual, institutional and societal impacts, considering scale, severity, reversibility, autonomy and interconnectedness.

### Layer 3: Human Oversight Architecture

Design meaningful oversight with awareness, authority, capacity and intervention. A person merely appearing in a workflow is not sufficient.

### Layer 4: Technical and Procurement Controls

Translate safeguards into architecture, data governance, security, auditability, supplier obligations, contracts and lifecycle controls.

### Layer 5: Accountability and Redress

Assign decision rights and create accessible mechanisms to challenge consequential AI-assisted outcomes.

### Layer 6: Continuous Monitoring, Measurement and Learning

Measure safeguard effectiveness, detect drift and incidents, and feed evidence back into governance and system design.

## 4. AI-DPI Risk Classification

| Level | Characteristics | Minimum governance |
|---|---|---|
| 1: Limited | Low consequence, reversible, limited users | Basic documentation, security and monitoring |
| 2: Material | Meaningful operational or individual impact | Formal risk assessment and accountable owner |
| 3: High | Significant rights, financial or service-access implications | Enhanced testing, independent review and human approval |
| 4: Systemic | Population-scale, highly consequential or difficult to reverse | Senior approval, continuous monitoring, tested intervention and independent assurance |

Classification should remain context-sensitive. A system that is low-risk in an isolated environment can become high-risk when connected to foundational DPI.

## 5. Mapping DPI Safeguards to AI Controls

| DPI safeguard area | AI-DPI operational control |
|---|---|
| Do no harm | AI impact assessment, harm modelling and risk thresholds |

| | |
|---|---|
| Non-discrimination | Bias testing, representative evaluation and disparity monitoring |
| Non-exclusion | Accessibility testing, alternative channels and inclusive user journeys |
| Transparency and accountability | AI disclosure, decision ownership and audit trails |
| Rule of law | Legal basis, regulatory review and documented decision authority |
| Autonomy and agency | User choice, meaningful human review and contestability |
| Community engagement | Participatory testing and stakeholder feedback |
| Remedy and redress | Appeals, complaints, correction and human review |

| **DPI safeguard area** | **AI-DPI operational control** |
|---|---|
| Future sustainability | Lifecycle cost, institutional capacity and environmental assessment |
| Evolving with evidence | Continuous monitoring, audits and model reassessment |
| Privacy/security by design | Data minimisation, access controls, threat modelling and secure architecture |
| Inclusive governance | Multi-stakeholder oversight and documented decision rights |
| Financial viability | Sustainable operating model and lifecycle cost assessment |
| Open/interoperable assets | Open standards, interoperability and portability where appropriate |

## 6. AI Controls Across the DPI Lifecycle

The AI governance layer should operate across the same lifecycle rather than being treated as a separate compliance activity.

| **Lifecycle stage** | **Key AI-DPI questions and controls** | **Primary output** |
|---|---|---|
| Conception & Scoping | Is AI necessary? Who could be affected? What services or rights could be impacted? | AI-DPI Impact Assessment |
| Strategy & Design | Define governance, data, security, human oversight, procurement, monitoring and redress. | Governance & Technical Design |
| Development | Test data, model performance, security, fairness, accessibility, failure modes and auditability. | Verified Release Candidate |
| Deployment | Controlled rollout, capacity-building, monitoring, communication, redress and incident response. | Approved Operational Deployment |

| | | |
|---|---|---|
| Operations & Maintenance | Monitor drift, incidents, complaints, security, model changes, impact and retirement readiness. | Continuous Safeguard Assurance |

## 7. The AI-DPI Control Loop

## AI-DPI

## Control Loop

## Identify → Assess → Design → Test → Approve → Deploy → Monitor → Intervene → Learn

The lifecycle provides the governance structure. The Control Loop provides the operational mechanism. The loop should repeat whenever material model, data, use-case, risk or context changes.

A responsible AI system must have an explicit pathway for intervention, escalation, modification or retirement. A theoretical override that cannot be executed under operating conditions is not a meaningful safeguard.

## 8. Responsible Authorities and Accountability

| **Responsible authority** | **Core AI-DPI responsibility** |
|---|---|
| Government / DPI owner | Public purpose, policy and ultimate accountability |

| **Responsible authority** | **Core AI-DPI responsibility** |
|---|---|
| DPI programme leadership | Programme governance, coordination and risk management |
| Regulator | Rights, legal compliance, competition and independent oversight |
| Technology provider | Technical safeguards, security, testing and documentation |
| Procurement authority | Contractual safeguard requirements and supplier accountability |
| Funder | Safeguard conditions and sustainability requirements |
| Civil society / advocates | Independent scrutiny and community representation |
| Independent assurance | Testing, audit and evidence-based challenge |
| Service operator | Operational monitoring and incident response |
| Users / communities | Feedback, participation and contestability |

The purpose of this model is to prevent accountability from disappearing behind the statement that “the algorithm made the decision.” Algorithms do not replace institutional accountability.

## 9. Responsible AI-DPI Procurement Control Set

Procurement is an early opportunity to make safeguards enforceable. AI-enabled DPI procurement should address:

- Data: ownership, permitted processing, retention, deletion and secondary use.
- Model: provenance, versioning, material changes, performance expectations and limitations.
- Security: security standards, vulnerability management, incident reporting and access control.
- Accountability: audit rights, documentation, assurance requirements and responsible personnel.
- Continuity: availability, disaster recovery, business continuity and transition arrangements.
- Interoperability: open standards, APIs, portability and avoidance of unnecessary vendor lock-in.
- Human oversight: intervention requirements, escalation procedures and suspension mechanisms.
- Exit: data return/deletion, transition support, knowledge transfer and replacement readiness.

**Procurement principle: safeguards should become contractual obligations rather than post-deployment aspirations.**

## 10. Inclusion, Agency and Redress

AI can unintentionally increase barriers for people already experiencing digital exclusion. An AI-DPI Inclusion Test should examine language, literacy, disability, age, connectivity, affordability, digital skills, alternative channels and contestability.

The objective is not merely equal access to technology. It is equitable access to the outcome the technology is intended to enable.

For consequential decisions, transparency should support contestability. Affected people should understand, where appropriate, whether AI is involved, what role it plays, relevant limitations, who is accountable and how an outcome can be challenged.

This approach is consistent with the Global Digital Compact's emphasis on a responsible, accountable, transparent and human-centric approach across the technology lifecycle, with effective human oversight, and with its emphasis on equitable AI governance and capacity-building, particularly for developing countries (United Nations, 2024).

The United Nations General Assembly has since given this agenda institutional form, establishing the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance (United Nations, 2025).

## 11. Security, Privacy and Technical Assurance

AI-enabled DPI creates a combined privacy and cybersecurity challenge. Controls should include data minimisation, purpose limitation, access controls, retention limits, privacy impact assessment, threat modelling, encryption, authentication, vulnerability management, monitoring, incident response and appropriate adversarial testing.

AI-specific threats can include prompt manipulation, data poisoning, model extraction, adversarial inputs, unauthorised model changes, training-data leakage and excessive autonomous action.

Security and privacy should be reassessed as systems evolve. The OECD AI Principles similarly emphasise human agency and oversight, transparency and explainability, robustness, security and safety, and accountability throughout the AI lifecycle (OECD, 2024).

## 12. Sustainability

Sustainability should be assessed across four dimensions: financial, institutional, technical and environmental.

- Financial sustainability: Can the system be funded over its full lifecycle?
- Institutional sustainability: Does the country or organisation have the capability to operate and govern it?
- Technical sustainability: Can the system be maintained, upgraded and integrated without excessive dependency?
- Environmental sustainability: What are the infrastructure, compute, energy and resource implications?

AI-enabled DPI should be evaluated not only on whether it can be launched, but whether it can remain safe, useful and governable over time.

## 13. Measurement, Evaluation and Learning

| Dimension | Example indicator |
|---|---|
| Governance | % of AI-DPI systems with named accountable owners |
| Risk | % completing documented impact assessments |
| Human oversight | % of high-impact systems with tested intervention mechanisms |
| Testing | % completing appropriate pre-deployment testing |
| Transparency | % of relevant services disclosing AI involvement |
| Redress | Average time to resolve AI-related complaints |
| Inclusion | Performance disparity across relevant user groups |
| Accessibility | % of priority user journeys passing accessibility testing |
| Security | Number of material AI-related security incidents |
| Privacy | % of systems completing privacy assessment |
| Procurement | % of critical suppliers meeting safeguard clauses |
| Monitoring | % of high-impact systems subject to periodic reassessment |
| Sustainability | Lifecycle cost and environmental assessment completed |
| Learning | % of material incidents resulting in documented control changes |

Indicators should be adapted to national context rather than treated as a universal compliance checklist. The objective is to connect safeguard principles with observable evidence and learning.

## 14. AI-DPI Deployment Readiness Assessment

Before population-scale deployment, the responsible authority should be able to demonstrate:

- Purpose: clear public/service purpose; AI necessity established; alternatives considered.
- Risk: impact assessment completed; high-risk populations identified; residual risk explicitly accepted.
- Human oversight: accountable authority identified; human review established; override tested.
- Technical safeguards: security testing, privacy controls, auditability and model/version management.

- Inclusion: accessibility tested; relevant language needs addressed; alternative channels considered.
- Accountability: supplier responsibilities documented; regulatory obligations addressed; escalation established.
- Redress: complaint mechanism available; appeal process defined; human review accessible.
- Monitoring: indicators and intervention thresholds established; continuous monitoring activated.
- Sustainability: lifecycle financing, institutional capacity, technical maintenance and environmental implications assessed.

If critical controls are absent, deployment should be delayed, restricted or redesigned.

## 15. Implementation Roadmap and Maturity

| Step | Action |
|---|---|
| 1. Establish | Create AI-DPI governance ownership and identify existing DPI risks. |
| 2. Assess | Map AI systems against DPI principles, risks and lifecycle stages. |
| 3. Embed | Integrate safeguards into architecture, procurement, contracts and operating procedures. |
| 4. Measure | Introduce indicators, assurance mechanisms, user feedback and independent review. |
| 5. Adapt | Use evidence to modify governance, technology and policy continuously. |

A practical maturity model is: Reactive → Policy-Based → Managed → Adaptive. The strategic objective is to move from compliance with safeguards to institutional capability to operate safely.

## 16. Worked Example (Illustrative)

Consider a hypothetical ministry introducing AI-assisted eligibility screening into an existing social-protection DPI serving two million people. Under Layer 1, the purpose test asks whether AI is necessary: the ministry must show that backlogs and error rates cannot be resolved through process improvement alone. Under Layer 2, the system is classified Level 3, High, because errors affect service access and household income; this triggers enhanced testing, independent review and human approval before deployment. Layer 3 gives caseworkers real authority: they can override, escalate and suspend, with workload and training designed so review is feasible rather than nominal. Layer 4 moves controls into procurement: bias testing across demographic groups, documented model limitations, audit rights and suspension mechanisms become contractual obligations.

Layer 5 establishes redress: an applicant refused a benefit can obtain human review within a defined service standard. Layer 6 monitors disparity in overturn rates, complaint volumes and model drift, feeding quarterly evidence back to the governance board. The Control Loop then governs every material change: a retrained model re-enters the loop at Assess, not at Deploy. This example is hypothetical and illustrates the framework's application; it reports no real programme.

### Strategic Recommendations

1. Treat AI as a DPI safeguard issue wherever AI can materially affect DPI outcomes.
2. Establish accountability before deployment and identify an accountable authority for high-impact systems.
3. Make human oversight operational by defining authority, capability, intervention and escalation.
4. Embed safeguards into procurement so critical controls become enforceable supplier obligations.
5. Design for inclusion and test systems against real-world barriers rather than assuming equal access.
6. Make redress part of the architecture for consequential AI-assisted decisions.
7. Measure safeguard outcomes and not only policy completion.
8. Build local technical, legal, governance and operational capability alongside infrastructure.
9. Design for financial, institutional, technical and environmental sustainability.
10. Govern continuously and evolve controls with evidence, technology, threats and societal needs.

## Conclusion

Digital Public Infrastructure has the potential to become one of the most important foundations of inclusive digital economies and public services. Artificial intelligence may significantly increase that potential. It may also increase the consequences of failure.

The answer is not to prevent innovation, nor to treat safeguards as a compliance exercise after technology has already been deployed. The answer is to make safeguards part of the architecture of digital transformation.

**From principles to operational controls.**

**From policy documents to accountable implementation.**

**From human presence to meaningful human oversight.**

**From equal access to equitable outcomes.**

**From procurement as purchasing to procurement as governance.**

**From static compliance to continuous learning.**

The ultimate test of responsible AI is not how intelligent the system becomes. It is whether people remain protected, informed, empowered and able to seek meaningful human accountability when the system affects their lives.

Responsible AI is not an obstacle to digital transformation. Properly designed, it is part of the infrastructure that makes transformation sustainable.

## Limitations

This paper is an independent conceptual contribution. The six-layer model, risk classification, control loop, indicator set and readiness assessment are practitioner proposals: they have not been applied in a live DPI programme, and no empirical validation is claimed or implied. The framework is aligned with the Universal DPI Safeguards Framework, the OECD AI Principles and the Global Digital Compact, but it is not derived from them, and alignment does not constitute endorsement by any institution. The example indicators are proposed measures for adaptation to national context, not reported baselines. Sources were reviewed in August 2026, and the referenced frameworks continue to evolve. The author writes as an independent practitioner; the perspectives of governments, providers and affected communities are represented through cited public materials rather than primary consultation.

## References

DPI Safeguards Initiative (2026) Universal DPI Safeguards Framework, Version 2.0. Available at: https://www.dpi-safeguards.org/framework (Accessed: 16 August 2026).

Organisation for Economic Co-operation and Development (2024) OECD AI Principles. Paris: OECD. Available at: https://www.oecd.org/en/topics/ai-principles.html (Accessed: 16 August 2026).

United Nations (2024) Global Digital Compact: Annex I to the Pact for the Future. New York: United Nations. Available at: https://www.un.org/digital-emerging-technologies/global-digital-compact (Accessed: 16 August 2026).

United Nations (2025) Terms of reference and modalities for the establishment and functioning of the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance. General Assembly Resolution A/RES/79/325, adopted 26 August 2025. Available at: https://docs.un.org/en/A/RES/79/325 (Accessed: 22 November 2026).

Author's Note: This working paper is an independent practitioner contribution. It is not an official publication of the United Nations, UNDP, the DPI Safeguards Initiative or any other international organisation. It does not claim authorship, endorsement or institutional affiliation with the Universal DPI Safeguards Framework. Its purpose is to contribute a practitioner perspective on the governance of AI-enabled systems within the broader DPI safeguards ecosystem.

## Research and citation note

The publication was reviewed against current public materials available on 16 August 2026. Claims about the DPI Safeguards Initiative, the Global Digital Compact and OECD AI Principles were checked against primary institutional sources. The paper intentionally distinguishes the author's proposed framework from official frameworks and does not represent the proposed controls as UN-endorsed recommendations.