Technical Annex

Sovereign Education-AI Reference Architecture

Teacher-Led, Localised and Frugal AI for Equitable Education

(Focus is on the Global South, especially on Small Island Developing States and Low- and Middle-Income Countries)

Template for consideration by	Ministries of Education in the Global South, especially SIDS and LMICs
Developed by	Commonwealth of Learning (COL) as a reference resource
Purpose	Reference architecture and options menu for sovereign, Frugal AI in education; countries may adopt, adapt, or set aside as appropriate
Intended audience	Senior officials, education management information system (EMIS) and information and communications technology (ICT) units, curriculum authorities, teacher education institutions, public universities, regulators
Status	For discussion and adaptation
Version /date	v1 | February 2026
Scope note	A practical baseline for teacher-support and system enablement. High-stakes automated decision-making (e.g., student progression, certification) is out of scope for this baseline.
Document conventions	This annex presents options, and also defines a Minimum Government Baseline of safeguards for any public deployment that (i) processes education data or (ii) produces content intended for learners. Components beyond the baseline are presented as optional modules.

Contents

1. Executive Summary for Ministers and Officials

This technical annex provides a reference architecture — a structured set of options and guidelines for deploying artificial intelligence in education. It is designed for developing countries, especially for small island developing states (SIDS) and low- and middle-income countries (LMICs) that wish to explore AI while maintaining national control over education data and policy.

What problem does this address?

Many countries face a common challenge: how to benefit from AI in education without surrendering control of education data and policy, or deploying tools that fail when connectivity is poor. Commercial AI subscriptions can be procured quickly, but they introduce dependency risks (service availability, latency, cost volatility, and data jurisdiction). Fully sovereign systems are feasible but require deliberate capacity-building. This architecture sets out a sovereignty-by-design baseline: national control where it matters most, with pragmatic options for capability and scale.

What choices do countries have?

This annex presents options, and also sets out a Minimum Government Baseline of safeguards for public deployments (especially learner-facing uses). Countries may:

- Adopt the full architecture as a starting point for national implementation
- Adapt selected components to fit existing systems and capacities
- Use hybrid approaches that combine local processing with controlled cloud access for complex tasks
- Use it as a reference for evaluating vendor proposals and procurement decisions
- Set it aside entirely if alternative approaches better suit national circumstances

What about the quality gap?

Quality is managed through task specialisation and hybrid options. Locally hosted models are optimised for core education workflows on national infrastructure - predictable latency, predictable costs, offline continuity, and data residency, while remaining grounded in ministry-approved knowledge bases. Where additional capability is required for specific high-complexity tasks, this architecture supports controlled cloud burst within a defined sovereignty envelope (data minimisation, jurisdictional controls, audit logging, and offline fallback). The appropriate split between local and cloud processing should be established empirically through pilots, reflecting language, curriculum, and connectivity constraints.

How does this scale?

The architecture includes a risk-tiered approach to teacher oversight. Not every AI output requires full teacher approval before release. Low-risk tasks that do not directly reach learners (e.g., formatting) can be automated with logging. Medium-risk tasks intended for teacher-only use can use post-hoc audit with publish-controls. Any output intended for learners defaults to high-risk gating with strict approval. This makes the system viable at scale while maintaining safety where it matters most.

What resources are required?

Pilots can begin with modest infrastructure: a single server at a ministry or public university, existing school devices for offline access, and open-source software with no licensing fees. Countries with

existing Digital Public Infrastructure (DPI) may leverage those investments. Scaling requires additional planning for connectivity, device management, teacher training, and technical capacity.

2. Design Principles and Options

2.1 Guiding Principles

The following principles inform the architecture. Countries may weigh these differently according to national priorities.

Teacher-led accountability

AI assists professional tasks; teachers maintain oversight through a risk-tiered workflow. High-risk, learner-facing content requires explicit approval. This preserves professional judgment while enabling operational efficiency.

Sovereign operation

Education data and audit logs remain under national jurisdiction and policy control. Countries determine what data is collected, where it is stored, and who may access it. Hybrid cloud options maintain sovereignty through data governance controls. Where hybrid cloud is used, Ministries should define a sovereignty envelope specifying what may be processed externally (if anything), under what transformations, and with what contractual controls.

Frugal resilience

Systems operate offline-first with store-and-forward synchronisation to support low bandwidth and climate-induced disruptions. Hybrid topologies allow cloud burst for complex tasks while maintaining local fallback.

Local relevance

Alignment to national curricula, languages and pedagogies through explicit localisation workflows. AI outputs are grounded in ministry-approved content.

Safety-by-design

Safeguards for learners include incident reporting, traceability and mechanisms for error correction. Risk classification determines the level of human oversight required.

2.2 Minimum Government Baseline (Non-Functional Requirements)

For any public deployment aligned to this architecture, Ministries are expected to define and implement the following baseline requirements (thresholds adapted to local context):

- Availability under intermittent connectivity, including defined offline service levels
- Privacy controls: data minimisation, personally identifiable information (PII) redaction, controlled retention of logs
- Security baseline: role-based access control, encryption, patch management, incident response readiness
- Auditability: immutable or tamper-evident logs, versioned knowledge base and model configurations
- Scalability pathway: design for pilot-to-scale transition from the outset

Optional modules (e.g., controlled cloud burst, advanced tuning, DPI integrations) should only be adopted once baseline safeguards are demonstrably in place.

3. Reference Architecture (Layered Model)

The reference architecture is organised into eight layers. Countries may implement these as shared services or modular components. Each layer should be addressed explicitly in design and procurement, though implementation approaches will vary.

Layer	Objective	Core components (minimum)
A. Access & Identity	Controlled access and accountability	Role-based access; least privilege; multi-factor authentication (MFA) for administrators; school/district tenancy where applicable
B. Teacher-in-the-Loop Workflow	Pedagogy leads technology	Risk-tiered approval workflow (see Section 5.1); feedback capture; audit trail
C. Application Services	Deliver education value safely	Lesson planning; assessment drafting; open educational resources (OER) adaptation; ministry drafting support
D. Knowledge Layer (RAG)	Ground outputs in approved sources	Ministry-vetted knowledge base; ingestion and versioning; retrieval policies; citation display
E. Privacy Airlock	Minimise and protect personal data	PII detection/redaction; context minimisation; policy filters; retention control
F. Model Layer	Localised intelligence and cost control	Locally hostable models for routine tasks; hybrid cloud burst for complex tasks (see Section 4.2)
G. Infrastructure	Frugal resilience and continuity	Edge/hub nodes; offline caching; store-and-forward sync; DPI-compatible registries where available
H. Operations (model lifecycle and security operations)	Continuous assurance	Model and knowledge-base registries; evaluation pipelines; monitoring; incident response; audit reporting

3.1 Interface Expectations

For implementations that follow this architecture, the following interface behaviours support interoperability and auditability:

- Knowledge ingestion and retrieval interfaces support versioning and provenance tracking
- Audit logging is consistent across applications, model services and knowledge services
- Offline-first synchronisation follows defined data minimisation and retention policies
- Integration with existing Education Management Information System (EMIS), Learning Management System (LMS), and DPI systems follows national data exchange standards where these exist

Minimum interoperability behaviours (procurement-relevant):

- Single sign-on (SSO) / role-based access integration (where available)
- Export/import of approved content artefacts (lesson plans, items, rubrics)
- Audit log export in a machine-readable format (for regulator/oversight)

- Knowledge base ingestion supports versioning, provenance, and licensing metadata
- LMS/EMIS integration must not require ingestion of unnecessary learner records into prompts

3.2 DPI Compatibility

Countries with existing Digital Public Infrastructure investments may leverage those foundations. The architecture is designed to be compatible with federated, unbundled DPI approaches:

- Registry and identity layers can integrate with national digital identity systems
- Credentialing components can align with national qualification frameworks
- Data exchange can follow national interoperability standards

Where national DPI components exist, countries may integrate rather than rebuild infrastructure layers. Example (illustrative): India's National Digital Education Architecture (NDEAR) is one national approach; other countries will apply their own DPI frameworks and legal requirements.

4. Hosting Topologies and Hybrid Options

4.1 Reference Hosting Topologies

Countries may select a topology that matches connectivity, institutional capacity and disaster risk. Hybrid combinations are common and often advisable. The following are illustrative options:

Topology	Contexts where it may fit	Considerations
A. Ministry hub + district nodes	National or multi-district scale	Central governance and audit; district inference and caching; supports offline schools via periodic sync
B. Public university shared service	Where universities can host public digital services	Builds national capacity; ministry retains governance and audit control; useful for teacher education integration
C. School edge devices	Remote areas and high-outage contexts	Local inference and content access; periodic updates; strongest resilience but requires device management
D. Hybrid (Frugal + Cloud Burst)	Balancing sovereignty with capability	Local processing for routine tasks; controlled cloud access for complex reasoning; requires data governance for cloud tier; requires a defined sovereignty envelope for any external processing

4.2 The Hybrid Model: Addressing the Quality Gap

Locally hosted models are selected for resilience, cost predictability, and data residency. Frontier cloud models can offer broader general capability, but they also create external dependency and jurisdictional risk. A hybrid topology provides strategic autonomy: local inference as the default for routine workflows, with an option for controlled cloud burst for narrowly defined tasks where it adds measurable value.

Local processing for routine tasks (illustrative planning assumption; validate in pilots)

- Formatting and document preparation
- Basic quiz and question generation
- Translation and localisation

- Fact retrieval from approved knowledge bases
- Administrative task support

Controlled cloud burst for computationally intensive drafting support (illustrative planning assumption; validate in pilots)

- Complex mathematical reasoning and problem-solving
- Computationally intensive drafting support subject to Tier 1 safeguards if the output is intended for learners
- Tasks requiring frontier model capabilities

Cloud burst requires additional governance controls: data minimisation before transmission, approved cloud providers (sovereign cloud or trusted partners), audit logging of cloud interactions, and fallback to local processing when connectivity fails. Risk tiering and approval gates apply regardless of whether processing is local or cloud-based.

4.3 Open-Source Options (Illustrative)

Open-source components can reduce dependency and support sovereign operation. Selection should follow national procurement rules, licensing checks, security due diligence, and evaluation against local language and curriculum requirements. Categories to consider include:

- Models: open-weights small language models suitable for local inference; multilingual variants where available
- Inference serving: lightweight model servers supporting central/graphics processor (CPU/GPU) inference and offline packaging
- RAG stack: document ingestion, vector search and retrieval services hosted in-country
- Safety and moderation: policy filters for prohibited content and child safeguarding
- Observability: standard metrics and logging, dashboards for latency and availability

4.4 Sovereignty Envelope for Hybrid/Cloud Burst

This subsection defines a practical sovereignty envelope for hybrid deployments. It specifies what data may be processed externally (if any), what must remain in-country, and what controls are required to make external processing auditable.

Data classification (education-specific, illustrative):

- Child-related / learner data (highest protection category)
- Sensitive (administrative records, staff data)
- Internal (non-sensitive operational content)
- Public (curriculum, public guidance)

Transformation required before any external processing:

Minimisation, redaction, and aggregation thresholds.

Permitted cloud burst payloads (default):

- Permitted: de-identified, curriculum-only prompts; generic pedagogy queries; non-personal content drafting subject to Tier controls

- Prohibited by default: learner free text, identifiers, or small-school quasi-identifiers; special category data

Contractual and operational controls (minimum):

- Right to terminate and retrieve/securely delete data upon exit
- Sub-processor disclosure and location controls (where applicable)
- Retention and deletion commitments for prompts and derived data
- Breach notification timelines and incident coordination procedures
- Audit rights and independent evaluation rights for education authorities

5. Governance Guardrails: Risk-Tiered Teacher-in-the-Loop

The original Teacher-in-the-Loop (TiL) concept—requiring teacher approval for all AI outputs—creates a scalability bottleneck. Training 50 teachers for a pilot is manageable; training 50,000 teachers to review every AI output is not. This section introduces a risk-tiered approach that maintains safety while enabling scale.

5.1 Risk-Tiered Approval Framework

The following framework calibrates human oversight to the risk level of the task:

Risk Tier	Use Cases	Approval Requirement	Rationale
Tier 1: High Risk	Any output intended for learners; assessment items and marking guidance; sensitive topics (as locally defined); high-consequence communications	Mandatory human approval before learner release	Direct impact on learners; errors may cause harm
Tier 2: Medium Risk	Teacher professional support drafts not intended for learner release (planning, internal notes, resource scaffolding)	Immediate release to teacher with publish-controls + periodic quality-assurance (QA) sampling	Supports teacher efficiency while preventing direct learner release; sampling enables quality improvement
Tier 3: Low Risk	Non-instructional automation that does not reach learners (formatting, administrative templates)	Automated release with logging	Minimal consequence; oversight not cost-effective

Tier classification quick test (default to higher tier)

Will this output be shown to learners? → Tier 1
Is it assessment-related, grading-related, or high-consequence? → Tier 1
Is it sensitive-topic content? → Tier 1
Is it teacher-only draft material not intended for learners? → Tier 2
Is it purely administrative and non-instructional? → Tier 3
If uncertain → Tier 1

Tier 1: High Risk — Strict Teacher Approval

Any content that will be presented directly to learners, or that addresses sensitive topics, requires explicit teacher review and approval before release. The system enforces this technically through an approval gate. Teachers edit, approve, and take professional responsibility for the content.

Tier 2: Medium Risk — Post-Hoc Audit

Content intended for teacher use (not direct learner consumption) can be released to the requesting teacher immediately, with systematic sampling and audit. Tier 2 outputs must be clearly labelled "Not learner-ready" and cannot be exported to learner channels without an explicit Tier 1 confirmation step. If a Tier 2 output is to be shared with learners, it must pass through Tier 1 approval. A defined percentage of outputs are reviewed for quality assurance. Errors identified through audit feed back into system improvement.

Tier 3: Low Risk — Automated Release

Routine tasks with minimal risk of harm that do not reach learners (e.g., formatting and administrative templates) can be automated with logging. All outputs are logged for potential review, but no human approval is required before release. This tier handles high-volume, low-stakes tasks efficiently.

5.2 Classification and Governance

Countries should establish a classification process to assign tasks to tiers:

- Default to higher tiers when classification is uncertain
- Review and adjust classifications based on operational experience
- Document classification decisions and rationale
- Establish escalation paths for edge cases

5.3 Data Protection and Privacy Airlock

Regardless of tier, all processing passes through the privacy airlock:

- Data minimisation by default: only what is necessary for the task is processed
- PII detection and redaction: personal identifiers are removed before model processing

Quasi-identifiers (e.g., school name + grade + rare attribute) should be treated as potentially re-identifying in small schools and islands.

- Retention and deletion: define default retention for prompts/outputs/logs; apply secure deletion; restrict access to logs; enable redress workflows.
- Controlled retention: documented retention periods and secure deletion procedures
- Audit trail: tamper-evident logging for regulatory review and individual redress

5.4 Security Baseline

The following controls represent a minimum-security posture:

- Role-based access control; privileged access management; MFA for administrators
- Encryption in transit and at rest; secure key management
- Secure configuration and patch management for servers and school edge devices
- Incident response procedures: detection, triage, remediation, and reporting

6. Implementation Challenges and Mitigations

Sovereign education-AI deployments are programmes, not single procurements. The following implementation challenges are predictable across contexts; the mitigation patterns are intended to reduce pilot risk, shorten time-to-value, and preserve safety and sovereignty during scale-up.

Instrument (illustrative)	Core expectation	How reflected in this architecture
UNESCO Recommendation on the Ethics of AI (2021)	Human-centred values; accountability; transparency	Teacher-led accountability; audit trails; traceability and redress; governance guardrails
UNESCO Guidance for Generative AI in Education and Research (2023)	Education-specific safeguards; responsible use	Teacher-in-the-Loop for learner-facing content; privacy airlock; vetted knowledge sources
Organisation for Economic Co-operation and Development (OECD) AI Principles (2019)	Robustness, security, transparency, accountability	Security baseline; risk-tiered oversight; logging; evaluation and incident response
National Institute of Standards and Technology (NIST) AI Risk Management Framework (2023)	Lifecycle risk management (govern–map–measure–manage)	Operations layer; evaluation pipelines; monitoring; incident response; continuous improvement

6.1 Procurement Paralysis

Building a sovereign stack requires procuring and maintaining servers, operating open-source model lifecycles (MLOps), and curating knowledge bases. This requires different capabilities than purchasing a commercial subscription. Without a phased plan, ministries may delay implementation while seeking a "perfect" solution.

Mitigations: Start with managed pilots using university partnerships or regional shared services. Adopt phased implementation that builds capacity progressively. Initial deployments should be treated as controlled pilots with explicit safeguards and an iterative improvement cycle based on measured evidence. Consider regional cooperation to share technical burden across multiple small states.

6.2 Quality Gap

Quality management is a programme requirement, not a one-off model selection. Open-weights models that run locally are designed for efficient, curriculum-grounded workflows and may not match frontier cloud models on every open-ended task. The key risk is capability mismatch: deploying the wrong model for the wrong task, or creating expectations that cannot be met offline. The architecture mitigates this through task-model fit, hybrid cloud burst for specific high-complexity workflows, localisation and fine-tuning for national curricula and languages, and continuous evaluation using pilot telemetry.

Mitigations: Adopt hybrid topology with cloud burst for complex tasks. Match model capability to task requirements—many educational tasks do not require frontier intelligence. Invest in model fine-tuning for local languages and curricula. Participate in initiatives to improve open-source educational models (see Frugal AI Challenge concept below).

6.3 Pilot-to-Scale Chasm

Many education technology initiatives demonstrate strong pilots but stall at scale. Teacher-in-the-Loop, in particular, can become a capacity constraint: training teachers to be effective "AI editors" requires sustained professional development and usable workflow design.

Mitigations: Implement risk-tiered approval to reduce the volume requiring teacher review. Design cascade training models where trained teachers train others. Integrate AI literacy into existing teacher professional development rather than creating parallel programmes. Simplify the teacher interface to minimise training burden. Plan for scale from pilot design stage.

6.4 The Frugal AI Challenge Concept

One mechanism to address the quality gap is a coordinated innovation challenge, building on similar frugal innovation models used internationally, focused specifically on education. A possible framing:

"Create the best Mathematics Tutor that runs on an Aptus Pi (or equivalent edge device) without internet."

Such a challenge would incentivise the global developer community to solve the specific constraints of resource-limited educational contexts, driving innovation in Layer F (Model Layer) of this architecture.

7. Policy Alignment

This architecture aligns with major international frameworks for AI governance in education. The following table maps core principles:

Definitions (for consistent cross-country reporting)
Active teacher: a teacher who completes at least one substantive workflow action per month (e.g., generates + reviews/edits + saves/exports).
Usage frequency: median sessions per active teacher per month (aggregated).
Tier 2 audit pass rate: % sampled outputs meeting quality and safety criteria.

Principle	This Architecture	UNESCO Guidance	OECD AI Principles	India DPI/NDEAR
Governance	Sovereign Operation (Data Residency)	Data Dignity	Human-Centric & Trustworthy	Data Sovereignty
Ethics	Teacher-in-the-Loop	Human-Centred AI	Public AI Infrastructure	Responsible AI
Infrastructure	Frugal Resilience (Offline First)	Green AI	Reliability & Accuracy	Federated Architecture
Knowledge	Epistemic Control (Vetted RAG)	Epistemic Diversity	Proportionate Governance	Curated Content
Scale	Risk-Tiered Human Review	Caution over Speed	Quarterly	Population Scale via DPI
Safety & Trust	Safety incidents (#, severity, time-to-remediate) by tier	Incident logs aggregated; time-to-remediate tracked		

This alignment supports countries in positioning national implementations within international policy discourse and in engaging partners on governance, safety, and sustainability requirements.

7.1 DPI Alignment

The architecture is designed to be compatible with Digital Public Infrastructure (DPI) approaches discussed in multiple international fora. Countries may explore alignment with relevant partner financing modalities subject to national priorities and partner requirements. Where national DPI exists, infrastructure layers can integrate rather than duplicate.

8. Monitoring and Evaluation (Suggested 12-Month Indicators)

The following indicators are suggested for cross-country pilots and system scale-up decisions. Countries should adapt these to national monitoring frameworks and disaggregate by region and school type where feasible.

Domain	Indicator (suggested)	Measurement example	Reporting
Teacher uptake	Active teachers per month; usage frequency	Aggregated usage logs; participation by district	Monthly
Teacher oversight	Share of Tier 1 drafts edited; audit sample pass rate for Tier 2	Workflow logs; sample audits of approvals	Monthly
Language coverage	Coverage of core learning resources in local languages	Knowledge-base inventory; localisation backlog	Quarterly
Frugal performance	Latency (median/95th); availability offline; cloud burst frequency	Monitoring and offline service tests	Monthly
Scale readiness	Teacher training completion; support ticket volume	Training records; helpdesk data	Quarterly

Telemetry should be privacy-preserving by design: collect the minimum necessary; aggregate at district/national level; apply suppression rules for small cohorts; restrict access to authorised teams; and prohibit collection of learner free text for monitoring unless explicitly authorised and protected.

Note: Indicators should be reviewed at month 6 to assess pilot health and at month 12 for scale-up decisions. Early warning indicators (teacher uptake, support ticket volume) can signal implementation problems before they become critical.

9. Pilot Blueprint Template

Countries planning pilots may use the following template to support comparability and reduce implementation risk. Each element should be specified in pilot design documentation:

Pilot scope

Grades/levels, subjects, and targeted teacher cohorts (e.g., national science curriculum, grades 7–10, 50 teachers in three districts). Include explicit scale-up pathway.

Topology selection

A (hub + district), B (university shared service), C (school edge), or D (hybrid), including connectivity assumptions, sovereignty envelope for any external processing (if applicable), and fallback provisions.

Include a one-page data flows diagram covering prompt path, knowledge base retrieval, logging, sync points, and retention.

Risk tier classification

Which use cases fall into Tier 1, 2, and 3; classification rationale; review schedule; and the Tier 2→Tier 1 promotion rule and publish-controls for any output that may be shared with learners.

Knowledge base

Curriculum documents, approved textbooks/guides, vetted OER; language coverage targets; licensing notes and provenance documentation.

Teacher capacity building

Training approach for Tier 1 approval workflow; cascade training model for scale; integration with existing professional development.

Safeguards

Synthetic or de-identified demo content for initial testing; privacy airlock configuration; sovereignty envelope configuration (if any external processing is used); publish-controls and escalation paths; incident response procedures and escalation paths.

Measurement

Baseline values for selected indicators; targets at month 6 and month 12; data collection methods and responsibilities; early warning thresholds.

Governance

Named focal points at ministry and implementing institutions; steering committee composition; reporting schedule and decision points; scale-up criteria.

Timeline

Key milestones from inception to month 12 review, including procurement, technical setup, teacher training, classroom deployment, evaluation, and scale-up decision.

10. Glossary of Key Terms

Airlock: A control layer that detects and removes personal identifiers from data before it is processed by AI models, minimising privacy exposure.

Cloud burst: Temporary use of cloud computing resources for tasks that exceed local processing capacity, with governance controls to maintain data sovereignty.

DPI (Digital Public Infrastructure): Shared digital systems (identity, payments, data exchange) that enable public and private services at population scale. Examples include India's Aadhaar, Unified Payments Interface (UPI), and NDEAR.

Edge device: A computing device located at the point of use (e.g., in a school) rather than in a central data centre, enabling local processing and offline operation.

Frugal AI: AI systems designed for resource-constrained environments, optimising for low cost, low energy consumption, and operation under poor connectivity.

LLM (Large Language Model): AI models with billions of parameters, typically requiring cloud infrastructure and offering frontier capabilities.

RAG (Retrieval-Augmented Generation): A technique that grounds AI outputs in retrieved documents from a knowledge base, improving accuracy and enabling citation of sources.

SLM (Small Language Model): AI models with fewer parameters that can run on modest hardware, suitable for edge deployment but with reduced capability compared to LLMs.

Sovereign operation: Operation of AI systems under national jurisdiction and policy control, including in-country governance of knowledge bases and audit logs, and a defined sovereignty envelope governing any external processing (if used).

Quasi-identifier: A combination of attributes that may re-identify an individual in small populations (e.g., school + grade + rare characteristic).

Learner-facing content: Any content intended to be shown directly to learners or used for assessment, feedback, or progression decisions.

Sensitive topics: Content areas defined nationally as requiring heightened care (e.g., health, religion, conflict history), especially in learner-facing materials.

Store-and-forward: A synchronisation method where data is stored locally when connectivity is unavailable and transmitted when connection is restored.

Teacher-in-the-Loop (TiL): A workflow design that requires teacher oversight of AI-generated content, with the level of oversight calibrated to the risk level of the task.

Appendix A: Self-Assessment Checklist for Ministries (Optional)

The following indicators may assist ministries in assessing alignment with this reference architecture. Countries should adapt these to national legal frameworks and institutional capacity. This checklist is offered as a tool for self-assessment, not as a compliance requirement.

Risk-Tiered Teacher-in-the-Loop

- Tasks are classified into risk tiers with documented rationale
- Tier 1 (high-risk) content requires teacher approval before learner release
- Tier 2 (medium-risk) content is subject to systematic post-hoc audit and includes publish-controls
    - *Tier 2 publish-controls* prevent direct export to learner channels without Tier 1 confirmation
    - *Tier 2 →Tier 1 promotion rule is enforced for any learner-facing use*
- Tier 3 (low-risk) tasks that do not reach learners are automated with logging
- Classification is reviewed and adjusted based on operational experience

Knowledge Management

- All knowledge sources are vetted by relevant curriculum authorities
- Knowledge base is versioned with change history
- AI outputs cite retrieved sources where RAG is used

Privacy and Data Protection

- PII airlock is implemented; personal identifiers are redacted before model processing
- Data minimisation is the default; only necessary data is collected
- Cloud burst (if used) has appropriate data governance controls
- Sovereignty envelope is defined for any external processing (permitted/prohibited payloads, transformations, contractual controls)
- Retention and deletion policies are documented and followed
- Audit logs are tamper-evident and accessible for authorised review

Infrastructure and Resilience

- Offline-first operation is supported at defined service levels
- Hybrid topology (if used) has fallback to local processing
- Synchronisation and continuity procedures are documented
- DPI integration points are identified (where applicable)

Security

- Role-based access control is implemented
- Encryption is used in transit and at rest
- Patch management and incident response procedures are documented

Scale Readiness

- Teacher training approach is designed for cascade scaling
- Pilot design includes explicit scale-up pathway and criteria
- Implementation challenges (procurement, quality, scale) have documented mitigations

Monitoring and Evaluation

- Risk-tiered oversight indicators are collected (Tier 1 approval rate, Tier 2 audit results)
- Teacher uptake and training completion are tracked
- Cloud burst frequency is monitored (if applicable)
- Early warning thresholds are defined and monitored

Document end. *This reference architecture was developed by the Commonwealth of Learning as a resource for Commonwealth education ministries. Countries retain full discretion over adoption, adaptation, or alternative approaches.*