

ODA3 INSTITUTE

SECTOR IMPLEMENTATION GUIDANCE

GAISSF™ Education Sector Implementation Guide

Informative sector implementation guidance for the GAISSF™ v1.0 control baseline across education systems

GAISSF-SEC-043 | Public | Draft for Publication — External Review Open

Issued 30 June 2026 | Updated 07 October 2026

Document Control and Use

All 59 authoritative control identifiers and titles are preserved. Sector interpretations are informative.

- Select applicable use cases and threats.
- Use the cross-reference matrix to identify controls, tasks and evidence.
- Execute the 0–90 day stabilisation checklist, then the 0–180 day roadmap.
- Record justified non-applicability and evidence limits.

Public-package boundary. This public guide must not contain internal review administration, unpublished approval records, credentials, secrets, or protected student/research data. Internal review logs and publication-gate records remain outside the public publication package.

Certification boundary. GAISSF certification is not currently offered. Any future certification language is conditional on a separately approved and published scheme.

Physical AI relationship. Where education AI includes robotics, laboratory automation, connected access systems, vehicles, drones or other cyber-physical actuation, apply GAISSF™ D9 alongside PAI-SF™ v1.0 as applicable. CSX-PAISF-004 is a mapping only and does not establish equivalence, evidence sufficiency or automatic conformance.

Evidence taxonomy. This guide does not redefine ODA3's controlling T1-T4 evidence-sourcing taxonomy. Local threat-source descriptors use descriptive evidence-source classes rather than T1-T4 labels.

D7 boundary. D7 is the canonical GAISSF domain Human & Societal Harms. Its five controls address AI-generated phishing, deepfake detection training, out-of-band authentication, AI social-engineering incident response and AI-enhanced external attack defence. Accessibility, bias, safeguarding and human-factors concerns remain important sector risks but do not redefine D7.

Applicable-Control Baseline. GAISSF™ v1.0 contains 59 controls across nine domains. The canonical baseline is all 52 controls in D1-D8. D9 contains seven additional/conditional controls and is mandatory where Physical AI or cyber-physical actuation is in scope. Education-sector tailoring changes implementation context and evidence expectations; it does not remove D1-D8 controls from the baseline.

Document identity. This publication is GAISSF-SEC-043, Version 1.0. Corrections retain v1.0 and are recorded by update date rather than patch-version identities.

Framework baseline and sector boundaries

1. Executive Summary

Education institutions exercise consequential authority over learners while processing long-lived personal, assessment, safeguarding and research data. AI creates model, autonomy, supplier, academic-integrity and unequal-impact risks that conventional security alone does not address.

2. Scope and Key Definitions

High-impact decisions include admissions, enrolment, grading, progression, scholarships, discipline, safeguarding, accommodations and material student-record changes. A competent reviewer is trained, authorised, evidence-enabled and able to override. AI containment isolates the AI function and tools while preserving evidence and essential service.

3. Education Operating Context

Characteristic	Required response
Minors	Age-appropriate design, safeguarding escalation and stronger data restrictions; legal review determines local duties.
Open academic environment	Protect collaboration, assessments, IP and export-controlled material.
Accessibility	Test assistive workflows and maintain an effective alternative.
Third-party dependence	Control model, subprocessor, change, continuity and exit.
Asymmetric authority	Provide notice, meaningful human review, contestability and correction.

4. GAISSF Domain Guide

Domain	Coverage
D1	Model Integrity & Adversarial Robustness — Protects datasets, model artifacts, adaptations and behaviour against poisoning, extraction, drift and integrity failures.
D2	Runtime Security & Adversarial Defense — Protects live interactions and connected tools against injection, jailbreaks, multimodal manipulation and context hijacking.
D3	Agentic Risk & Autonomous System Security — Limits autonomous authority, secures agent communication and memory, and preserves effective human control.
D4	Supply Chain & Third-Party AI Security — Maintains assurance across models, registries, APIs, MCP servers, software components and shadow AI.
D5	Content Safety & Output Integrity — Controls harmful, privacy-invasive, infringing or misleading outputs and validates privacy-preserving techniques.
D6	Governance, Accountability & Human Oversight — Establishes accountable decision authority, auditability, incident readiness, lifecycle control and resilience.
D7	Human & Societal Harms — Prepares institutions for AI-enabled phishing, deepfakes, impersonation, social engineering and external attack amplification.
D8	Regulatory Alignment & Compliance — Maps applicable legal and standards obligations and verifies technical and governance documentation.
D9	Physical AI Safety — Protects people and facilities where AI controls robots, laboratories, vehicles, access systems or physical processes.

5. Threat Register

The register distinguishes observed/reported classes, demonstrated attacks and plausible scenarios. Evidence source class and confidence are separate. No row asserts a named institutional incident.

ID	Threat	Complexity	Detection	Evidence source class	Confidence	Primary controls
EDU-THR-001	Student data uploaded to public model	Low	Difficult	Primary / corroborated	High	D1-CTL-01; D1-CTL-09; D3-CTL-06
EDU-THR-002	Model trained on student work without approval	Medium	Moderate	Corroborated authoritative	High	D1-CTL-02; D2-CTL-01; D3-CTL-07
EDU-THR-003	Prompt injection reveals assessment guidance	Medium	Moderate	Corroborated / contextual	Moderate	D1-CTL-03; D2-CTL-02; D4-CTL-01
EDU-THR-004	AI detector produces false misconduct allegation	Medium	Moderate	Contextual	Moderate	D1-CTL-04; D2-CTL-03; D4-CTL-02
EDU-THR-005	Automated grading systematically mis-scores a group	Medium	Moderate	Contextual / unverified	Low	D1-CTL-05; D2-CTL-04; D4-CTL-03
EDU-THR-006	Examination content leakage	Medium	Moderate	Primary / corroborated	High	D1-CTL-06; D2-CTL-05; D4-CTL-04
EDU-THR-007	Deepfake impersonation of administrator	Low	Moderate	Corroborated authoritative	High	D1-CTL-07; D2-CTL-06; D4-CTL-05
EDU-THR-008	Account takeover alters student records	Medium	Moderate	Corroborated / contextual	Moderate	D1-CTL-08; D3-CTL-01; D4-CTL-06
EDU-THR-009	Wellbeing chatbot gives unsafe advice	Medium	Moderate	Contextual	Moderate	D1-CTL-09; D3-CTL-02; D4-CTL-07
EDU-THR-010	Safeguarding alert is missed	Medium	Moderate	Contextual / unverified	Low	D2-CTL-01; D3-CTL-03; D5-CTL-01
EDU-THR-011	Safeguarding alert causes unjustified intervention	Medium	Moderate	Primary / corroborated	High	D2-CTL-02; D3-CTL-04; D5-CTL-02

ID	Threat	Complexity	Detection	Evidence source class	Confidence	Primary controls
EDU-THR-012	Biometric attendance data reused	Low	Difficult	Corroborated authoritative	High	D2-CTL-03; D3-CTL-05; D5-CTL-03
EDU-THR-013	Facial recognition misidentifies student	Medium	Difficult	Corroborated / contextual	Moderate	D2-CTL-04; D3-CTL-06; D5-CTL-04
EDU-THR-014	Adaptive learning locks learner into low pathway	Medium	Difficult	Contextual	Moderate	D2-CTL-05; D3-CTL-07; D5-CTL-05
EDU-THR-015	Accessibility regression excludes disabled learners	Medium	Difficult	Contextual / unverified	Low	D2-CTL-06; D4-CTL-01; D5-CTL-06
EDU-THR-016	Shadow AI use exposes confidential material	Low	Difficult	Primary / corroborated	High	D3-CTL-01; D4-CTL-02; D6-CTL-01
EDU-THR-017	Vendor changes model without notice	Medium	Moderate	Corroborated authoritative	High	D3-CTL-02; D4-CTL-03; D6-CTL-02
EDU-THR-018	Subprocessor stores data in unapproved jurisdiction	Low	Difficult	Corroborated / contextual	Moderate	D3-CTL-03; D4-CTL-04; D6-CTL-03
EDU-THR-019	Research data leakage	Low	Difficult	Contextual	Moderate	D3-CTL-04; D4-CTL-05; D6-CTL-04
EDU-THR-020	Fabricated citations enter published research	Medium	Moderate	Contextual / unverified	Low	D3-CTL-05; D4-CTL-06; D6-CTL-05
EDU-THR-021	Malicious file compromises AI workflow	Medium	Moderate	Primary / corroborated	High	D3-CTL-06; D4-CTL-07; D6-CTL-06
EDU-THR-022	Autonomous agent modifies grades	High	Difficult	Corroborated authoritative	High	D3-CTL-07; D5-CTL-01; D6-CTL-07
EDU-THR-023	Agent sends unauthorised communication	High	Difficult	Corroborated / contextual	Moderate	D4-CTL-01; D5-CTL-02; D7-CTL-H01
EDU-THR-024	Model API outage disrupts examination	Low	Moderate	Contextual	Moderate	D4-CTL-02; D5-CTL-03; D7-CTL-H02
EDU-THR-025	Training-data poisoning affects tutor	Low	Difficult	Contextual / unverified	Low	D4-CTL-03; D5-CTL-04; D7-CTL-H03
EDU-THR-026	Secrets exposed in logs	Low	Difficult	Primary / corroborated	High	D4-CTL-04; D5-CTL-05; D7-CTL-H04
EDU-THR-027	Insecure plugin accesses unrelated records	Medium	Moderate	Corroborated authoritative	High	D4-CTL-05; D5-CTL-06; D7-CTL-H05
EDU-THR-028	Student manipulates recommender to target peers	Medium	Moderate	Corroborated / contextual	Moderate	D4-CTL-06; D6-CTL-01; D8-CTL-01
EDU-THR-029	Synthetic identity used for fraudulent enrolment	Low	Moderate	Contextual	Moderate	D4-CTL-07; D6-CTL-02; D8-CTL-02
EDU-THR-030	Scholarship decision lacks appeal	Medium	Moderate	Contextual / unverified	Low	D5-CTL-01; D6-CTL-03; D8-CTL-03
EDU-THR-031	Disciplinary tool amplifies historical bias	Medium	Moderate	Primary / corroborated	High	D5-CTL-02; D6-CTL-04; D8-CTL-04
EDU-THR-032	Language model disadvantages non-native speakers	Medium	Difficult	Corroborated authoritative	High	D5-CTL-03; D6-CTL-05; D8-CTL-05
EDU-THR-033	Parent targeted by AI-generated	Low	Moderate	Corroborated / contextual	Moderate	D5-CTL-04; D6-CTL-06; D9-CTL-

ID	Threat	Complexity	Detection	Evidence source class	Confidence	Primary controls
	fee scam					01
EDU-THR-034	Model update invalidates prior assurance	Medium	Moderate	Contextual	Moderate	D5-CTL-05; D6-CTL-07; D9-CTL-02
EDU-THR-035	Deleted student data persists in vendor backups	Low	Difficult	Contextual / unverified	Low	D5-CTL-06; D7-CTL-H01; D9-CTL-03

6. High-Risk Use Cases

ID	Use case	Impact rationale	Minimum data	Competent reviewer
EDU-UC-001	AI tutoring for minors	Direct interaction with minors creates learning, privacy and safeguarding consequences.	Student pseudonymous ID; age band; course context; current task; minimal learning history	Teacher or learning-support reviewer; safeguarding lead for high-risk disclosures
EDU-UC-002	Adaptive learning	Recommendations can constrain pathways and reinforce early predictions.	Pseudonymous learner ID; task outcomes; mastery indicators; accessibility preferences	Teacher or curriculum lead
EDU-UC-003	Automated grading	Scores directly affect grades, progression and academic opportunity.	Student ID; submission; rubric; authorised reference material	Qualified subject assessor or moderation panel
EDU-UC-004	AI-authorship detection	False positives can cause misconduct and disciplinary consequences.	Submission; assignment context; declared AI use; validated comparison data	Academic integrity investigator with corroborating evidence
EDU-UC-005	Remote examination proctoring	Monitoring can affect exam validity, privacy, disability access and sanctions.	Verified identity; session event flags; minimum necessary video/audio	Human proctor and appeal reviewer
EDU-UC-006	Admissions screening	Can determine access to education and reproduce historical disadvantage.	Application fields demonstrably relevant to published criteria	Admissions officer or panel
EDU-UC-007	Scholarship decision support	EDU-UC-007: The use case may affect educational access, quality, privacy, safety or records.	EDU-UC-007: Only data necessary for scholarship decision support, using pseudonymous identifiers where feasible.	Named process owner with subject expertise and authority to override.
EDU-UC-008	Student retention prediction	Predictions can stigmatise students and alter treatment.	Current academic/engagement indicators proven relevant	Student-support professional
EDU-UC-009	Academic advising	EDU-UC-009: The use case may affect educational access, quality, privacy, safety or records.	EDU-UC-009: Only data necessary for academic advising, using pseudonymous identifiers where feasible.	Named process owner with subject expertise and authority to override.
EDU-UC-010	Disciplinary decision support	EDU-UC-010: The use case may affect educational access, quality, privacy, safety or records.	EDU-UC-010: Only data necessary for disciplinary decision support, using pseudonymous identifiers where feasible.	Named process owner with subject expertise and authority to override.
EDU-UC-011	Safeguarding alerts	False negatives can leave harm unaddressed; false positives can trigger intrusive action.	Minimum necessary signal, source context and contact route	Designated Safeguarding Lead
EDU-UC-012	Student wellbeing chatbot	Students may rely on the system during distress or crisis.	Minimal conversation state; age band; service location for routing	Trained support service
EDU-UC-013	Biometric attendance	Sensitive biometric processing can affect access, discipline and surveillance.	Biometric template; class/session; attendance result	Attendance/records owner and DPO
EDU-UC-014	Facial recognition access	EDU-UC-014: The use case may affect educational access, quality, privacy, safety or records.	EDU-UC-014: Only data necessary for facial recognition access, using pseudonymous identifiers where feasible.	Named process owner with subject expertise and authority to override.
EDU-UC-015	Accessibility assistant	Failure may deny effective access to essential education or accommodations.	Content to transform; user-selected accessibility preferences	Accessibility specialist or educator
EDU-UC-016	Research assistant with	May expose participant,	Approved research	Principal investigator and

ID	Use case	Impact rationale	Minimum data	Competent reviewer
	confidential data	sponsor, export-controlled or unpublished data.	corpus; task instructions; project identity	research governance/security
EDU-UC-017	Generative AI in assessment	Affects assessment validity, equity and academic integrity.	Assignment prompt; declared AI use; output and process evidence	Course leader / academic integrity process
EDU-UC-018	Lesson planning assistant	EDU-UC-018: The use case may affect educational access, quality, privacy, safety or records.	EDU-UC-018: Only data necessary for lesson planning assistant, using pseudonymous identifiers where feasible.	Named process owner with subject expertise and authority to override.
EDU-UC-019	Question generation	EDU-UC-019: The use case may affect educational access, quality, privacy, safety or records.	EDU-UC-019: Only data necessary for question generation, using pseudonymous identifiers where feasible.	Named process owner with subject expertise and authority to override.
EDU-UC-020	Autonomous agent in student information system	Can change authoritative records and communicate at scale.	Task-specific fields and scoped transaction data	Records owner and IT approval

7. Control Interpretation Matrix

D1 — Protects datasets, model artifacts, adaptations and behaviour against poisoning, extraction, drift and integrity failures.

D1-CTL-01 — Dataset Provenance & Poisoning Prevention

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D1-CTL-01: Document and implement controls for dataset sources, licences, collection purpose, consent or lawful basis, transformation history and poisoning checks; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D1-CTL-01: For an education system using this capability, record dataset sources, licences, collection purpose, consent or lawful basis, transformation history and poisoning checks, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D1-CTL-01: System-specific design or configuration for dataset sources, licences, collection purpose, consent or lawful basis, transformation history and poisoning checks; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D1-CTL-01: Treating dataset sources, licences, collection purpose, consent or lawful basis, transformation history and poisoning checks as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D1-CTL-01: Inspect the live design and configuration, execute targeted tests of dataset sources, licences, collection purpose, consent or lawful basis, transformation history and poisoning checks, sample records and verify that failed results trigger containment, correction or rollback.

D1-CTL-02 — Model Extraction Resistance

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D1-CTL-02: Document and implement controls for query limits, output exposure, endpoint authentication and extraction-attempt monitoring; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D1-CTL-02: For an education system using this capability, record query limits, output exposure, endpoint authentication and extraction-attempt monitoring, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D1-CTL-02: System-specific design or configuration for query limits, output exposure, endpoint authentication and extraction-attempt monitoring; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D1-CTL-02: Treating query limits, output exposure, endpoint

Field	Implementation
	authentication and extraction-attempt monitoring as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D1-CTL-02: Inspect the live design and configuration, execute targeted tests of query limits, output exposure, endpoint authentication and extraction-attempt monitoring, sample records and verify that failed results trigger containment, correction or rollback.

D1-CTL-03 — Behavioral Drift Detection

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D1-CTL-03: Document and implement controls for accuracy, calibration, refusal behaviour and subgroup performance across terms, subjects and model versions; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D1-CTL-03: For an education system using this capability, record accuracy, calibration, refusal behaviour and subgroup performance across terms, subjects and model versions, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D1-CTL-03: System-specific design or configuration for accuracy, calibration, refusal behaviour and subgroup performance across terms, subjects and model versions; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D1-CTL-03: Treating accuracy, calibration, refusal behaviour and subgroup performance across terms, subjects and model versions as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D1-CTL-03: Inspect the live design and configuration, execute targeted tests of accuracy, calibration, refusal behaviour and subgroup performance across terms, subjects and model versions, sample records and verify that failed results trigger containment, correction or rollback.

D1-CTL-04 — Federated Learning Poisoning Prevention

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D1-CTL-04: Document and implement controls for participant authentication, signed updates, robust aggregation and anomalous-client quarantine; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D1-CTL-04: For an education system using this capability, record participant authentication, signed updates, robust aggregation and anomalous-client quarantine, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D1-CTL-04: System-specific design or configuration for participant authentication, signed updates, robust aggregation and anomalous-client quarantine; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D1-CTL-04: Treating participant authentication, signed updates, robust aggregation and anomalous-client quarantine as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D1-CTL-04: Inspect the live design and configuration, execute targeted tests of participant authentication, signed updates, robust aggregation and anomalous-client quarantine, sample records and verify that failed results trigger containment, correction or rollback.

D1-CTL-05 — Embedding Space Robustness

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D1-CTL-05: Document and implement controls for retrieval behaviour, demographic clustering, representation leakage and recommendation bias; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.

Field	Implementation
Example	D1-CTL-05: For an education system using this capability, record retrieval behaviour, demographic clustering, representation leakage and recommendation bias, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D1-CTL-05: System-specific design or configuration for retrieval behaviour, demographic clustering, representation leakage and recommendation bias; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D1-CTL-05: Treating retrieval behaviour, demographic clustering, representation leakage and recommendation bias as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D1-CTL-05: Inspect the live design and configuration, execute targeted tests of retrieval behaviour, demographic clustering, representation leakage and recommendation bias, sample records and verify that failed results trigger containment, correction or rollback.

D1-CTL-06 — Post-Quantum Model Signing & Crypto Hardening

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D1-CTL-06: Document and implement controls for artifact signing, key governance, cryptographic agility and migration planning; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D1-CTL-06: For an education system using this capability, record artifact signing, key governance, cryptographic agility and migration planning, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D1-CTL-06: System-specific design or configuration for artifact signing, key governance, cryptographic agility and migration planning; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D1-CTL-06: Treating artifact signing, key governance, cryptographic agility and migration planning as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D1-CTL-06: Inspect the live design and configuration, execute targeted tests of artifact signing, key governance, cryptographic agility and migration planning, sample records and verify that failed results trigger containment, correction or rollback.

D1-CTL-07 — LoRA/Adapter Integrity Verification

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D1-CTL-07: Document and implement controls for adapter provenance, hashes, training data, compatibility and regression behaviour; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D1-CTL-07: For an education system using this capability, record adapter provenance, hashes, training data, compatibility and regression behaviour, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D1-CTL-07: System-specific design or configuration for adapter provenance, hashes, training data, compatibility and regression behaviour; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D1-CTL-07: Treating adapter provenance, hashes, training data, compatibility and regression behaviour as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D1-CTL-07: Inspect the live design and configuration, execute targeted tests of adapter provenance, hashes, training data, compatibility and regression behaviour, sample records and verify that failed results trigger containment, correction or rollback.

D1-CTL-08 — Model Merge Attack Detection

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector

Field	Implementation
	implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D1-CTL-08: Document and implement controls for source checkpoints, merge manifests, differential testing and backdoor screening; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D1-CTL-08: For an education system using this capability, record source checkpoints, merge manifests, differential testing and backdoor screening, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D1-CTL-08: System-specific design or configuration for source checkpoints, merge manifests, differential testing and backdoor screening; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D1-CTL-08: Treating source checkpoints, merge manifests, differential testing and backdoor screening as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D1-CTL-08: Inspect the live design and configuration, execute targeted tests of source checkpoints, merge manifests, differential testing and backdoor screening, sample records and verify that failed results trigger containment, correction or rollback.

D1-CTL-09 — Quantization Backdoor Screening

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D1-CTL-09: Document and implement controls for pre/post-compression behaviour, refusal performance, subgroup accuracy and trigger activation; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D1-CTL-09: For an education system using this capability, record pre/post-compression behaviour, refusal performance, subgroup accuracy and trigger activation, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D1-CTL-09: System-specific design or configuration for pre/post-compression behaviour, refusal performance, subgroup accuracy and trigger activation; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D1-CTL-09: Treating pre/post-compression behaviour, refusal performance, subgroup accuracy and trigger activation as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D1-CTL-09: Inspect the live design and configuration, execute targeted tests of pre/post-compression behaviour, refusal performance, subgroup accuracy and trigger activation, sample records and verify that failed results trigger containment, correction or rollback.

D2 — Protects live interactions and connected tools against injection, jailbreaks, multimodal manipulation and context hijacking.

D2-CTL-01 — Direct Prompt Injection Prevention

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D2-CTL-01: Document and implement controls for instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D2-CTL-01: For an education system using this capability, record instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D2-CTL-01: System-specific design or configuration for instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D2-CTL-01: Treating instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection as a generic policy

Field	Implementation
	topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D2-CTL-01: Inspect the live design and configuration, execute targeted tests of instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection, sample records and verify that failed results trigger containment, correction or rollback.

D2-CTL-02 — Indirect Prompt Injection Prevention

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D2-CTL-02: Document and implement controls for instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D2-CTL-02: For an education system using this capability, record instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D2-CTL-02: System-specific design or configuration for instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D2-CTL-02: Treating instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D2-CTL-02: Inspect the live design and configuration, execute targeted tests of instruction hierarchy, untrusted user input, tool restrictions and suspicious-session detection, sample records and verify that failed results trigger containment, correction or rollback.

D2-CTL-03 — Jailbreak Resistance Testing

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D2-CTL-03: Document and implement controls for multi-turn, multilingual, encoded and role-play attempts against education-specific boundaries; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D2-CTL-03: For an education system using this capability, record multi-turn, multilingual, encoded and role-play attempts against education-specific boundaries, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D2-CTL-03: System-specific design or configuration for multi-turn, multilingual, encoded and role-play attempts against education-specific boundaries; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D2-CTL-03: Treating multi-turn, multilingual, encoded and role-play attempts against education-specific boundaries as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D2-CTL-03: Inspect the live design and configuration, execute targeted tests of multi-turn, multilingual, encoded and role-play attempts against education-specific boundaries, sample records and verify that failed results trigger containment, correction or rollback.

D2-CTL-04 — Multi-Modal Injection Defense

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D2-CTL-04: Document and implement controls for hidden instructions in images, audio, OCR layers, metadata and document structures; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D2-CTL-04: For an education system using this capability, record hidden instructions in images, audio, OCR layers, metadata and document structures, test the configured service before release, and block deployment when the approved threshold is not met.

Field	Implementation
Evidence	D2-CTL-04: System-specific design or configuration for hidden instructions in images, audio, OCR layers, metadata and document structures; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D2-CTL-04: Treating hidden instructions in images, audio, OCR layers, metadata and document structures as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D2-CTL-04: Inspect the live design and configuration, execute targeted tests of hidden instructions in images, audio, OCR layers, metadata and document structures, sample records and verify that failed results trigger containment, correction or rollback.

D2-CTL-05 — Function Call/Tool Call Injection Prevention

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D2-CTL-05: Document and implement controls for tool schemas, argument validation, least privilege, approval gates and rollback; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D2-CTL-05: For an education system using this capability, record tool schemas, argument validation, least privilege, approval gates and rollback, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D2-CTL-05: System-specific design or configuration for tool schemas, argument validation, least privilege, approval gates and rollback; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D2-CTL-05: Treating tool schemas, argument validation, least privilege, approval gates and rollback as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D2-CTL-05: Inspect the live design and configuration, execute targeted tests of tool schemas, argument validation, least privilege, approval gates and rollback, sample records and verify that failed results trigger containment, correction or rollback.

D2-CTL-06 — Cross-Context Hijacking Mitigation

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D2-CTL-06: Document and implement controls for tenant, user, course, session and memory isolation; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D2-CTL-06: For an education system using this capability, record tenant, user, course, session and memory isolation, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D2-CTL-06: System-specific design or configuration for tenant, user, course, session and memory isolation; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D2-CTL-06: Treating tenant, user, course, session and memory isolation as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D2-CTL-06: Inspect the live design and configuration, execute targeted tests of tenant, user, course, session and memory isolation, sample records and verify that failed results trigger containment, correction or rollback.

D3 — Limits autonomous authority, secures agent communication and memory, and preserves effective human control.

D3-CTL-01 — Least Agency Enforcement

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D3-CTL-01: Document and implement controls for minimum tools,

Field	Implementation
	records, time windows, transaction volumes and permissions; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D3-CTL-01: For an education system using this capability, record minimum tools, records, time windows, transaction volumes and permissions, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D3-CTL-01: System-specific design or configuration for minimum tools, records, time windows, transaction volumes and permissions; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D3-CTL-01: Treating minimum tools, records, time windows, transaction volumes and permissions as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D3-CTL-01: Inspect the live design and configuration, execute targeted tests of minimum tools, records, time windows, transaction volumes and permissions, sample records and verify that failed results trigger containment, correction or rollback.

D3-CTL-02 — Inter-Agent Communication Security

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D3-CTL-02: Document and implement controls for agent identities, message schemas, trust rules and replay/spoofing protection; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D3-CTL-02: For an education system using this capability, record agent identities, message schemas, trust rules and replay/spoofing protection, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D3-CTL-02: System-specific design or configuration for agent identities, message schemas, trust rules and replay/spoofing protection; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D3-CTL-02: Treating agent identities, message schemas, trust rules and replay/spoofing protection as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D3-CTL-02: Inspect the live design and configuration, execute targeted tests of agent identities, message schemas, trust rules and replay/spoofing protection, sample records and verify that failed results trigger containment, correction or rollback.

D3-CTL-03 — Agentic Prompt Chaining Detection

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D3-CTL-03: Document and implement controls for harmful objectives split across multiple steps, agents or apparently benign requests; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D3-CTL-03: For an education system using this capability, record harmful objectives split across multiple steps, agents or apparently benign requests, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D3-CTL-03: System-specific design or configuration for harmful objectives split across multiple steps, agents or apparently benign requests; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D3-CTL-03: Treating harmful objectives split across multiple steps, agents or apparently benign requests as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D3-CTL-03: Inspect the live design and configuration, execute targeted tests of harmful objectives split across multiple steps, agents or apparently benign requests, sample records and verify that failed results trigger containment, correction or rollback.

D3-CTL-04 — Embodied AI Safety Controls

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D3-CTL-04: Document and implement controls for safe operating envelopes, supervision, zones, speed, force and physical stop controls; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D3-CTL-04: For an education system using this capability, record safe operating envelopes, supervision, zones, speed, force and physical stop controls, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D3-CTL-04: System-specific design or configuration for safe operating envelopes, supervision, zones, speed, force and physical stop controls; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D3-CTL-04: Treating safe operating envelopes, supervision, zones, speed, force and physical stop controls as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D3-CTL-04: Inspect the live design and configuration, execute targeted tests of safe operating envelopes, supervision, zones, speed, force and physical stop controls, sample records and verify that failed results trigger containment, correction or rollback.

D3-CTL-05 — Multi-Agent Trust Chain Attestation

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D3-CTL-05: Document and implement controls for agent identities, versions, attestations and approval status across a workflow; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D3-CTL-05: For an education system using this capability, record agent identities, versions, attestations and approval status across a workflow, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D3-CTL-05: System-specific design or configuration for agent identities, versions, attestations and approval status across a workflow; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D3-CTL-05: Treating agent identities, versions, attestations and approval status across a workflow as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D3-CTL-05: Inspect the live design and configuration, execute targeted tests of agent identities, versions, attestations and approval status across a workflow, sample records and verify that failed results trigger containment, correction or rollback.

D3-CTL-06 — Persistent Memory Exfiltration Prevention

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D3-CTL-06: Document and implement controls for cross-user memory access, protected attributes and inferential extraction; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D3-CTL-06: For an education system using this capability, record cross-user memory access, protected attributes and inferential extraction, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D3-CTL-06: System-specific design or configuration for cross-user memory access, protected attributes and inferential extraction; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D3-CTL-06: Treating cross-user memory access, protected attributes and inferential extraction as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D3-CTL-06: Inspect the live design and configuration, execute targeted tests of cross-user memory access, protected attributes and

Field	Implementation
	inferential extraction, sample records and verify that failed results trigger containment, correction or rollback.

D3-CTL-07 — Secure Memory Lifecycle Management

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D3-CTL-07: Document and implement controls for memory purpose, schema, retention, correction, expiry and deletion propagation; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D3-CTL-07: For an education system using this capability, record memory purpose, schema, retention, correction, expiry and deletion propagation, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D3-CTL-07: System-specific design or configuration for memory purpose, schema, retention, correction, expiry and deletion propagation; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D3-CTL-07: Treating memory purpose, schema, retention, correction, expiry and deletion propagation as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D3-CTL-07: Inspect the live design and configuration, execute targeted tests of memory purpose, schema, retention, correction, expiry and deletion propagation, sample records and verify that failed results trigger containment, correction or rollback.

D4 — Maintains assurance across models, registries, APIs, MCP servers, software components and shadow AI.**D4-CTL-01 — AI Bill Of Materials (AI BOM) Maintenance**

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D4-CTL-01: Document and implement controls for models, datasets, adapters, libraries, APIs, plugins, safety filters and hosting dependencies; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D4-CTL-01: For an education system using this capability, record models, datasets, adapters, libraries, APIs, plugins, safety filters and hosting dependencies, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D4-CTL-01: System-specific design or configuration for models, datasets, adapters, libraries, APIs, plugins, safety filters and hosting dependencies; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D4-CTL-01: Treating models, datasets, adapters, libraries, APIs, plugins, safety filters and hosting dependencies as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D4-CTL-01: Inspect the live design and configuration, execute targeted tests of models, datasets, adapters, libraries, APIs, plugins, safety filters and hosting dependencies, sample records and verify that failed results trigger containment, correction or rollback.

D4-CTL-02 — Model File & Artifact Scanning

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D4-CTL-02: Document and implement controls for malware, unsafe serialization, embedded code, secrets, hashes and quarantine; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D4-CTL-02: For an education system using this capability, record malware, unsafe serialization, embedded code, secrets, hashes and quarantine, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D4-CTL-02: System-specific design or configuration for malware, unsafe serialization, embedded code, secrets, hashes and quarantine; targeted test results; approval record; exceptions; change and incident history.

Field	Implementation
Failure modes	D4-CTL-02: Treating malware, unsafe serialization, embedded code, secrets, hashes and quarantine as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D4-CTL-02: Inspect the live design and configuration, execute targeted tests of malware, unsafe serialization, embedded code, secrets, hashes and quarantine, sample records and verify that failed results trigger containment, correction or rollback.

D4-CTL-03 — Model Hub & Registry Vetting

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D4-CTL-03: Document and implement controls for publisher identity, provenance, licence, artifact integrity, security and moderation practices; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D4-CTL-03: For an education system using this capability, record publisher identity, provenance, licence, artifact integrity, security and moderation practices, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D4-CTL-03: System-specific design or configuration for publisher identity, provenance, licence, artifact integrity, security and moderation practices; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D4-CTL-03: Treating publisher identity, provenance, licence, artifact integrity, security and moderation practices as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D4-CTL-03: Inspect the live design and configuration, execute targeted tests of publisher identity, provenance, licence, artifact integrity, security and moderation practices, sample records and verify that failed results trigger containment, correction or rollback.

D4-CTL-04 — MCP Server Behavioral Monitoring

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D4-CTL-04: Document and implement controls for server inventory, tool capabilities, data access, permissions and behavioural monitoring; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D4-CTL-04: For an education system using this capability, record server inventory, tool capabilities, data access, permissions and behavioural monitoring, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D4-CTL-04: System-specific design or configuration for server inventory, tool capabilities, data access, permissions and behavioural monitoring; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D4-CTL-04: Treating server inventory, tool capabilities, data access, permissions and behavioural monitoring as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D4-CTL-04: Inspect the live design and configuration, execute targeted tests of server inventory, tool capabilities, data access, permissions and behavioural monitoring, sample records and verify that failed results trigger containment, correction or rollback.

D4-CTL-05 — Third-Party AI API Security Assessment

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D4-CTL-05: Document and implement controls for authentication, data use, retention, logs, resilience, provider change and exit; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D4-CTL-05: For an education system using this capability, record authentication, data use, retention, logs, resilience, provider change and exit, test the configured service before release, and block deployment when the approved threshold is not met.

Field	Implementation
Evidence	D4-CTL-05: System-specific design or configuration for authentication, data use, retention, logs, resilience, provider change and exit; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D4-CTL-05: Treating authentication, data use, retention, logs, resilience, provider change and exit as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D4-CTL-05: Inspect the live design and configuration, execute targeted tests of authentication, data use, retention, logs, resilience, provider change and exit, sample records and verify that failed results trigger containment, correction or rollback.

D4-CTL-06 — Shadow AI Discovery & Governance

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D4-CTL-06: Document and implement controls for unapproved SaaS, browser extensions, embedded features, pilots and approved alternatives; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D4-CTL-06: For an education system using this capability, record unapproved SaaS, browser extensions, embedded features, pilots and approved alternatives, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D4-CTL-06: System-specific design or configuration for unapproved SaaS, browser extensions, embedded features, pilots and approved alternatives; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D4-CTL-06: Treating unapproved SaaS, browser extensions, embedded features, pilots and approved alternatives as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D4-CTL-06: Inspect the live design and configuration, execute targeted tests of unapproved SaaS, browser extensions, embedded features, pilots and approved alternatives, sample records and verify that failed results trigger containment, correction or rollback.

D4-CTL-07 — AI Software Composition Analysis (SCA)

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D4-CTL-07: Document and implement controls for AI libraries, runtimes, parsers, containers, orchestration components and vulnerabilities; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D4-CTL-07: For an education system using this capability, record AI libraries, runtimes, parsers, containers, orchestration components and vulnerabilities, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D4-CTL-07: System-specific design or configuration for AI libraries, runtimes, parsers, containers, orchestration components and vulnerabilities; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D4-CTL-07: Treating AI libraries, runtimes, parsers, containers, orchestration components and vulnerabilities as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D4-CTL-07: Inspect the live design and configuration, execute targeted tests of AI libraries, runtimes, parsers, containers, orchestration components and vulnerabilities, sample records and verify that failed results trigger containment, correction or rollback.

D5 — Controls harmful, privacy-invasive, infringing or misleading outputs and validates privacy-preserving techniques.

D5-CTL-01 — Harmful Content Blocking

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control

Field	Implementation
	membership is not optional.
Minimum expectation	D5-CTL-01: Document and implement controls for age-appropriate harmful-content boundaries, safeguarding escalation and false-positive/negative review; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D5-CTL-01: For an education system using this capability, record age-appropriate harmful-content boundaries, safeguarding escalation and false-positive/negative review, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D5-CTL-01: System-specific design or configuration for age-appropriate harmful-content boundaries, safeguarding escalation and false-positive/negative review; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D5-CTL-01: Treating age-appropriate harmful-content boundaries, safeguarding escalation and false-positive/negative review as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D5-CTL-01: Inspect the live design and configuration, execute targeted tests of age-appropriate harmful-content boundaries, safeguarding escalation and false-positive/negative review, sample records and verify that failed results trigger containment, correction or rollback.

D5-CTL-02 — PII Leakage Prevention

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D5-CTL-02: Document and implement controls for student, staff, parent, applicant and research-participant data in outputs; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D5-CTL-02: For an education system using this capability, record student, staff, parent, applicant and research-participant data in outputs, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D5-CTL-02: System-specific design or configuration for student, staff, parent, applicant and research-participant data in outputs; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D5-CTL-02: Treating student, staff, parent, applicant and research-participant data in outputs as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D5-CTL-02: Inspect the live design and configuration, execute targeted tests of student, staff, parent, applicant and research-participant data in outputs, sample records and verify that failed results trigger containment, correction or rollback.

D5-CTL-03 — Copyright Detection

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D5-CTL-03: Document and implement controls for protected course materials, textbooks, question banks, student work and research outputs; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D5-CTL-03: For an education system using this capability, record protected course materials, textbooks, question banks, student work and research outputs, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D5-CTL-03: System-specific design or configuration for protected course materials, textbooks, question banks, student work and research outputs; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D5-CTL-03: Treating protected course materials, textbooks, question banks, student work and research outputs as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D5-CTL-03: Inspect the live design and configuration, execute targeted tests of protected course materials, textbooks, question banks, student work and research outputs, sample records and verify

Field	Implementation
	that failed results trigger containment, correction or rollback.

D5-CTL-04 — AI Watermarking Robustness

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D5-CTL-04: Document and implement controls for provenance signals after compression, screenshots, transcription and LMS processing; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D5-CTL-04: For an education system using this capability, record provenance signals after compression, screenshots, transcription and LMS processing, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D5-CTL-04: System-specific design or configuration for provenance signals after compression, screenshots, transcription and LMS processing; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D5-CTL-04: Treating provenance signals after compression, screenshots, transcription and LMS processing as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D5-CTL-04: Inspect the live design and configuration, execute targeted tests of provenance signals after compression, screenshots, transcription and LMS processing, sample records and verify that failed results trigger containment, correction or rollback.

D5-CTL-05 — Privacy-By-Design Verification

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D5-CTL-05: Document and implement controls for minimisation, purpose limitation, privacy defaults, rights and vendor configuration; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D5-CTL-05: For an education system using this capability, record minimisation, purpose limitation, privacy defaults, rights and vendor configuration, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D5-CTL-05: System-specific design or configuration for minimisation, purpose limitation, privacy defaults, rights and vendor configuration; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D5-CTL-05: Treating minimisation, purpose limitation, privacy defaults, rights and vendor configuration as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D5-CTL-05: Inspect the live design and configuration, execute targeted tests of minimisation, purpose limitation, privacy defaults, rights and vendor configuration, sample records and verify that failed results trigger containment, correction or rollback.

D5-CTL-06 — Privacy-Preserving ML Validation

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D5-CTL-06: Document and implement controls for differential privacy, synthetic data, federated methods, inference attacks and utility trade-offs; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D5-CTL-06: For an education system using this capability, record differential privacy, synthetic data, federated methods, inference attacks and utility trade-offs, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D5-CTL-06: System-specific design or configuration for differential privacy, synthetic data, federated methods, inference attacks and utility trade-offs; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D5-CTL-06: Treating differential privacy, synthetic data, federated methods, inference attacks and utility trade-offs as a generic policy

Field	Implementation
	topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D5-CTL-06: Inspect the live design and configuration, execute targeted tests of differential privacy, synthetic data, federated methods, inference attacks and utility trade-offs, sample records and verify that failed results trigger containment, correction or rollback.

D6 — Establishes accountable decision authority, auditability, incident readiness, lifecycle control and resilience.

D6-CTL-01 — Human-In-The-Loop For High-Risk Actions

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D6-CTL-01: Document and implement controls for named competent reviewers, evidence access, response deadlines, reasons, override and appeal; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D6-CTL-01: For an education system using this capability, record named competent reviewers, evidence access, response deadlines, reasons, override and appeal, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D6-CTL-01: System-specific design or configuration for named competent reviewers, evidence access, response deadlines, reasons, override and appeal; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D6-CTL-01: Treating named competent reviewers, evidence access, response deadlines, reasons, override and appeal as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D6-CTL-01: Inspect the live design and configuration, execute targeted tests of named competent reviewers, evidence access, response deadlines, reasons, override and appeal, sample records and verify that failed results trigger containment, correction or rollback.

D6-CTL-02 — Audit Trail Completeness

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D6-CTL-02: Document and implement controls for model/version, input reference, output or action, human decision, override and change records; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D6-CTL-02: For an education system using this capability, record model/version, input reference, output or action, human decision, override and change records, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D6-CTL-02: System-specific design or configuration for model/version, input reference, output or action, human decision, override and change records; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D6-CTL-02: Treating model/version, input reference, output or action, human decision, override and change records as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D6-CTL-02: Inspect the live design and configuration, execute targeted tests of model/version, input reference, output or action, human decision, override and change records, sample records and verify that failed results trigger containment, correction or rollback.

D6-CTL-03 — AI Model Card Completeness

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D6-CTL-03: Document and implement controls for purpose, users, data, limitations, prohibited uses, evaluations and oversight; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D6-CTL-03: For an education system using this capability, record

Field	Implementation
	purpose, users, data, limitations, prohibited uses, evaluations and oversight, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D6-CTL-03: System-specific design or configuration for purpose, users, data, limitations, prohibited uses, evaluations and oversight; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D6-CTL-03: Treating purpose, users, data, limitations, prohibited uses, evaluations and oversight as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D6-CTL-03: Inspect the live design and configuration, execute targeted tests of purpose, users, data, limitations, prohibited uses, evaluations and oversight, sample records and verify that failed results trigger containment, correction or rollback.

D6-CTL-04 — AI Incident Response Readiness

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D6-CTL-04: Document and implement controls for AI-specific detection, containment, evidence preservation, affected-person review and return-to-service; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D6-CTL-04: For an education system using this capability, record AI-specific detection, containment, evidence preservation, affected-person review and return-to-service, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D6-CTL-04: System-specific design or configuration for AI-specific detection, containment, evidence preservation, affected-person review and return-to-service; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D6-CTL-04: Treating AI-specific detection, containment, evidence preservation, affected-person review and return-to-service as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D6-CTL-04: Inspect the live design and configuration, execute targeted tests of AI-specific detection, containment, evidence preservation, affected-person review and return-to-service, sample records and verify that failed results trigger containment, correction or rollback.

D6-CTL-05 — Model Deprecation & Decommissioning

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D6-CTL-05: Document and implement controls for replacement, data export, notices, retention, deletion, dependency removal and rollback; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D6-CTL-05: For an education system using this capability, record replacement, data export, notices, retention, deletion, dependency removal and rollback, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D6-CTL-05: System-specific design or configuration for replacement, data export, notices, retention, deletion, dependency removal and rollback; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D6-CTL-05: Treating replacement, data export, notices, retention, deletion, dependency removal and rollback as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D6-CTL-05: Inspect the live design and configuration, execute targeted tests of replacement, data export, notices, retention, deletion, dependency removal and rollback, sample records and verify that failed results trigger containment, correction or rollback.

D6-CTL-06 — Third-Party AI Vendor Governance

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector

Field	Implementation
	implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D6-CTL-06: Document and implement controls for impact-level classification, due diligence, contract controls, change notices, incidents and exit readiness; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D6-CTL-06: For an education system using this capability, record impact-level classification, due diligence, contract controls, change notices, incidents and exit readiness, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D6-CTL-06: System-specific design or configuration for impact-level classification, due diligence, contract controls, change notices, incidents and exit readiness; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D6-CTL-06: Treating impact-level classification, due diligence, contract controls, change notices, incidents and exit readiness as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D6-CTL-06: Inspect the live design and configuration, execute targeted tests of impact-level classification, due diligence, contract controls, change notices, incidents and exit readiness, sample records and verify that failed results trigger containment, correction or rollback.

D6-CTL-07 — AI Resilience & Business Continuity

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D6-CTL-07: Document and implement controls for tested non-AI fallback, capacity, recovery, peak-period change freeze and supplier escalation; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D6-CTL-07: For an education system using this capability, record tested non-AI fallback, capacity, recovery, peak-period change freeze and supplier escalation, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D6-CTL-07: System-specific design or configuration for tested non-AI fallback, capacity, recovery, peak-period change freeze and supplier escalation; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D6-CTL-07: Treating tested non-AI fallback, capacity, recovery, peak-period change freeze and supplier escalation as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D6-CTL-07: Inspect the live design and configuration, execute targeted tests of tested non-AI fallback, capacity, recovery, peak-period change freeze and supplier escalation, sample records and verify that failed results trigger containment, correction or rollback.

D7 — Prepares institutions for AI-enabled phishing, deepfakes, impersonation, social engineering and external attack amplification.

D7-CTL-H01 — AI-Generated Phishing Simulation

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D7-CTL-H01: Document and implement controls for authorised, privacy-preserving role-specific simulations and follow-up learning; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D7-CTL-H01: For an education system using this capability, record authorised, privacy-preserving role-specific simulations and follow-up learning, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D7-CTL-H01: System-specific design or configuration for authorised, privacy-preserving role-specific simulations and follow-up learning; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D7-CTL-H01: Treating authorised, privacy-preserving role-specific

Field	Implementation
	simulations and follow-up learning as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D7-CTL-H01: Inspect the live design and configuration, execute targeted tests of authorised, privacy-preserving role-specific simulations and follow-up learning, sample records and verify that failed results trigger containment, correction or rollback.

D7-CTL-H02 — Deepfake Detection Training

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D7-CTL-H02: Document and implement controls for limitations of visual/audio indicators and trusted-channel verification; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D7-CTL-H02: For an education system using this capability, record limitations of visual/audio indicators and trusted-channel verification, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D7-CTL-H02: System-specific design or configuration for limitations of visual/audio indicators and trusted-channel verification; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D7-CTL-H02: Treating limitations of visual/audio indicators and trusted-channel verification as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D7-CTL-H02: Inspect the live design and configuration, execute targeted tests of limitations of visual/audio indicators and trusted-channel verification, sample records and verify that failed results trigger containment, correction or rollback.

D7-CTL-H03 — Out-Of-Band Authentication

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D7-CTL-H03: Document and implement controls for independent trusted-channel verification for sensitive requests; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D7-CTL-H03: For an education system using this capability, record independent trusted-channel verification for sensitive requests, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D7-CTL-H03: System-specific design or configuration for independent trusted-channel verification for sensitive requests; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D7-CTL-H03: Treating independent trusted-channel verification for sensitive requests as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D7-CTL-H03: Inspect the live design and configuration, execute targeted tests of independent trusted-channel verification for sensitive requests, sample records and verify that failed results trigger containment, correction or rollback.

D7-CTL-H04 — AI Social Engineering IR

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D7-CTL-H04: Document and implement controls for impersonation containment, payment holds, media preservation and corrective communication; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D7-CTL-H04: For an education system using this capability, record impersonation containment, payment holds, media preservation and corrective communication, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D7-CTL-H04: System-specific design or configuration for

Field	Implementation
	impersonation containment, payment holds, media preservation and corrective communication; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D7-CTL-H04: Treating impersonation containment, payment holds, media preservation and corrective communication as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D7-CTL-H04: Inspect the live design and configuration, execute targeted tests of impersonation containment, payment holds, media preservation and corrective communication, sample records and verify that failed results trigger containment, correction or rollback.

D7-CTL-H05 — AI-Enhanced External Attack Defense

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D7-CTL-H05: Document and implement controls for higher-volume credential, phishing, bot and synthetic-identity attacks; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D7-CTL-H05: For an education system using this capability, record higher-volume credential, phishing, bot and synthetic-identity attacks, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D7-CTL-H05: System-specific design or configuration for higher-volume credential, phishing, bot and synthetic-identity attacks; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D7-CTL-H05: Treating higher-volume credential, phishing, bot and synthetic-identity attacks as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D7-CTL-H05: Inspect the live design and configuration, execute targeted tests of higher-volume credential, phishing, bot and synthetic-identity attacks, sample records and verify that failed results trigger containment, correction or rollback.

D8 — Maps applicable legal and standards obligations and verifies technical and governance documentation.

D8-CTL-01 — EU AI Act Risk Tier Mapping

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D8-CTL-01: Document and implement controls for current official risk classification, dates, applicability and legal uncertainty; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D8-CTL-01: For an education system using this capability, record current official risk classification, dates, applicability and legal uncertainty, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D8-CTL-01: System-specific design or configuration for current official risk classification, dates, applicability and legal uncertainty; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D8-CTL-01: Treating current official risk classification, dates, applicability and legal uncertainty as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D8-CTL-01: Inspect the live design and configuration, execute targeted tests of current official risk classification, dates, applicability and legal uncertainty, sample records and verify that failed results trigger containment, correction or rollback.

D8-CTL-02 — ISO 42001 Gap Analysis

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D8-CTL-02: Document and implement controls for AI management-system gap analysis without false equivalence or certification claims; define an accountable owner, measurable acceptance threshold,

Field	Implementation
	change trigger and evidence-retention rule.
Example	D8-CTL-02: For an education system using this capability, record AI management-system gap analysis without false equivalence or certification claims, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D8-CTL-02: System-specific design or configuration for AI management-system gap analysis without false equivalence or certification claims; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D8-CTL-02: Treating AI management-system gap analysis without false equivalence or certification claims as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D8-CTL-02: Inspect the live design and configuration, execute targeted tests of AI management-system gap analysis without false equivalence or certification claims, sample records and verify that failed results trigger containment, correction or rollback.

D8-CTL-03 — GPAI Technical Documentation Verification

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D8-CTL-03: Document and implement controls for provider technical documentation, limitations, evaluations, changes and missing information; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D8-CTL-03: For an education system using this capability, record provider technical documentation, limitations, evaluations, changes and missing information, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D8-CTL-03: System-specific design or configuration for provider technical documentation, limitations, evaluations, changes and missing information; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D8-CTL-03: Treating provider technical documentation, limitations, evaluations, changes and missing information as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D8-CTL-03: Inspect the live design and configuration, execute targeted tests of provider technical documentation, limitations, evaluations, changes and missing information, sample records and verify that failed results trigger containment, correction or rollback.

D8-CTL-04 — DORA ICT Incident Reporting (Financial Sector)

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control membership is not optional.
Minimum expectation	D8-CTL-04: Document and implement controls for documented applicability or non-applicability for financial-sector entities or services; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D8-CTL-04: For an education system using this capability, record documented applicability or non-applicability for financial-sector entities or services, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D8-CTL-04: System-specific design or configuration for documented applicability or non-applicability for financial-sector entities or services; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D8-CTL-04: Treating documented applicability or non-applicability for financial-sector entities or services as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D8-CTL-04: Inspect the live design and configuration, execute targeted tests of documented applicability or non-applicability for financial-sector entities or services, sample records and verify that failed results trigger containment, correction or rollback.

D8-CTL-05 — NIST SP 800-218A Compliance Check

Field	Implementation
Applicability	Included in the canonical D1-D8 Applicable-Control Baseline. Sector implementation remains system- and risk-informed, but control

Field	Implementation
	membership is not optional.
Minimum expectation	D8-CTL-05: Document and implement controls for secure AI-development practices where the institution develops or materially modifies AI software; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D8-CTL-05: For an education system using this capability, record secure AI-development practices where the institution develops or materially modifies AI software, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D8-CTL-05: System-specific design or configuration for secure AI-development practices where the institution develops or materially modifies AI software; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D8-CTL-05: Treating secure AI-development practices where the institution develops or materially modifies AI software as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D8-CTL-05: Inspect the live design and configuration, execute targeted tests of secure AI-development practices where the institution develops or materially modifies AI software, sample records and verify that failed results trigger containment, correction or rollback.

D9 — Protects people and facilities where AI controls robots, laboratories, vehicles, access systems or physical processes.

D9-CTL-01 — Physical Harm Boundary Enforcement

Field	Implementation
Applicability	Additional/conditional. Determine applicability in the Applicable-Control Baseline; mandatory where Physical AI or cyber-physical actuation is in scope.
Minimum expectation	D9-CTL-01: Document and implement controls for enforceable physical limits based on hazard analysis; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D9-CTL-01: For an education system using this capability, record enforceable physical limits based on hazard analysis, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D9-CTL-01: System-specific design or configuration for enforceable physical limits based on hazard analysis; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D9-CTL-01: Treating enforceable physical limits based on hazard analysis as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D9-CTL-01: Inspect the live design and configuration, execute targeted tests of enforceable physical limits based on hazard analysis, sample records and verify that failed results trigger containment, correction or rollback.

D9-CTL-02 — Safe State And Graceful Degradation

Field	Implementation
Applicability	Additional/conditional. Determine applicability in the Applicable-Control Baseline; mandatory where Physical AI or cyber-physical actuation is in scope.
Minimum expectation	D9-CTL-02: Document and implement controls for safe behaviour under sensor, model, network or environmental failure; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D9-CTL-02: For an education system using this capability, record safe behaviour under sensor, model, network or environmental failure, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D9-CTL-02: System-specific design or configuration for safe behaviour under sensor, model, network or environmental failure; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D9-CTL-02: Treating safe behaviour under sensor, model, network or environmental failure as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D9-CTL-02: Inspect the live design and configuration, execute

Field	Implementation
	targeted tests of safe behaviour under sensor, model, network or environmental failure, sample records and verify that failed results trigger containment, correction or rollback.

D9-CTL-03 — Human Override And Emergency Stop

Field	Implementation
Applicability	Additional/conditional. Determine applicability in the Applicable-Control Baseline; mandatory where Physical AI or cyber-physical actuation is in scope.
Minimum expectation	D9-CTL-03: Document and implement controls for independent, accessible and tested local and remote override; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D9-CTL-03: For an education system using this capability, record independent, accessible and tested local and remote override, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D9-CTL-03: System-specific design or configuration for independent, accessible and tested local and remote override; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D9-CTL-03: Treating independent, accessible and tested local and remote override as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D9-CTL-03: Inspect the live design and configuration, execute targeted tests of independent, accessible and tested local and remote override, sample records and verify that failed results trigger containment, correction or rollback.

D9-CTL-04 — Cyber-Physical Attack Detection

Field	Implementation
Applicability	Additional/conditional. Determine applicability in the Applicable-Control Baseline; mandatory where Physical AI or cyber-physical actuation is in scope.
Minimum expectation	D9-CTL-04: Document and implement controls for correlated command, identity, sensor and safety-control anomaly detection; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D9-CTL-04: For an education system using this capability, record correlated command, identity, sensor and safety-control anomaly detection, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D9-CTL-04: System-specific design or configuration for correlated command, identity, sensor and safety-control anomaly detection; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D9-CTL-04: Treating correlated command, identity, sensor and safety-control anomaly detection as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D9-CTL-04: Inspect the live design and configuration, execute targeted tests of correlated command, identity, sensor and safety-control anomaly detection, sample records and verify that failed results trigger containment, correction or rollback.

D9-CTL-05 — Physical Environment Integrity Monitoring

Field	Implementation
Applicability	Additional/conditional. Determine applicability in the Applicable-Control Baseline; mandatory where Physical AI or cyber-physical actuation is in scope.
Minimum expectation	D9-CTL-05: Document and implement controls for sensor health, calibration and environmental preconditions for safe operation; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D9-CTL-05: For an education system using this capability, record sensor health, calibration and environmental preconditions for safe operation, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D9-CTL-05: System-specific design or configuration for sensor health, calibration and environmental preconditions for safe operation; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D9-CTL-05: Treating sensor health, calibration and environmental preconditions for safe operation as a generic policy topic, relying

Field	Implementation
	solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D9-CTL-05: Inspect the live design and configuration, execute targeted tests of sensor health, calibration and environmental preconditions for safe operation, sample records and verify that failed results trigger containment, correction or rollback.

D9-CTL-06 — Actuator Command Verification

Field	Implementation
Applicability	Additional/conditional. Determine applicability in the Applicable-Control Baseline; mandatory where Physical AI or cyber-physical actuation is in scope.
Minimum expectation	D9-CTL-06: Document and implement controls for command identity, state, range, sequence and replay validation; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D9-CTL-06: For an education system using this capability, record command identity, state, range, sequence and replay validation, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D9-CTL-06: System-specific design or configuration for command identity, state, range, sequence and replay validation; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D9-CTL-06: Treating command identity, state, range, sequence and replay validation as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D9-CTL-06: Inspect the live design and configuration, execute targeted tests of command identity, state, range, sequence and replay validation, sample records and verify that failed results trigger containment, correction or rollback.

D9-CTL-07 — Physical Incident Evidence Preservation

Field	Implementation
Applicability	Additional/conditional. Determine applicability in the Applicable-Control Baseline; mandatory where Physical AI or cyber-physical actuation is in scope.
Minimum expectation	D9-CTL-07: Document and implement controls for synchronised command, sensor, model, video and intervention evidence; define an accountable owner, measurable acceptance threshold, change trigger and evidence-retention rule.
Example	D9-CTL-07: For an education system using this capability, record synchronised command, sensor, model, video and intervention evidence, test the configured service before release, and block deployment when the approved threshold is not met.
Evidence	D9-CTL-07: System-specific design or configuration for synchronised command, sensor, model, video and intervention evidence; targeted test results; approval record; exceptions; change and incident history.
Failure modes	D9-CTL-07: Treating synchronised command, sensor, model, video and intervention evidence as a generic policy topic, relying solely on vendor assertions, or failing to link it to a named system, threshold and retained evidence.
Validation	D9-CTL-07: Inspect the live design and configuration, execute targeted tests of synchronised command, sensor, model, video and intervention evidence, sample records and verify that failed results trigger containment, correction or rollback.

8. Human Oversight and Academic Integrity

AI-authorship detection is not definitive proof. Institutions must use context-specific validation, corroborating evidence, accessible appeal and false-positive/subgroup review. Material record changes require records-owner approval, transaction limits, audit and rollback.

9. Safeguarding, Accessibility and Research

AI is not a substitute for qualified safeguarding, healthcare, counselling or emergency personnel. Accessibility-critical services require effective alternatives. Confidential, participant, export-controlled, dual-use and unpublished research requires specialist governance.

10. Vendor Due Diligence

The refined questionnaire contains 47 questions with tailored evidence and disqualifying conditions.

11. Metrics

Metric	Definition	Target	Escalation
Inventory coverage	Discovered systems recorded / total systems found	100% within 10 business days	Any unowned high-impact system
Unapproved high-impact systems	Count operating without approval	0	Any instance
Priority assessments completed	Approved assessments / systems due	100%	Any overdue critical system
Vendor evidence current	High-impact vendors with current assurance / total	100%	Expired critical evidence
Critical accessibility defects	Open critical defects without alternative	0 before release	Any essential-access defect
Safeguarding response timeliness	High-risk alerts reviewed within approved time / total	100%	Any missed emergency threshold
Assessment agreement	AI-assisted scores within tolerance / sampled scores	Approved subject threshold	Below threshold or subgroup gap
Appeal reversal rate	Corrected AI-related outcomes / decided appeals	Trend review	Spike or repeated cause
Deletion within deadline	Verified deletions on time / due	100%	Any overdue high-risk request
Material changes reviewed	Changes reviewed before release / total	100%	Any unreviewed production change
AI incident recurrence	Repeated incidents with same unresolved cause	0	Any repeated high-severity cause
Human override rationale	Overrides with rationale / sampled decisions	100%	Missing rationale
Training competence	Critical role holders meeting threshold / required	100%	Any untrained critical reviewer
Evidence currency	Evidence reviewed on time / due	>=95%; 100% critical	Expired critical evidence
Suspended systems	Suspensions and days to disposition	Within SLA	Operation despite trigger

12. Roadmap

Stage	Period	Exit criterion
Stabilisation	0–30	High-impact systems identified, owners assigned, procurement freeze and restrictions operating.
Control Implementation	31–60	Assessments and specialist reviews complete; gaps owned.
Validation and Closure	61–90	Tests complete; gaps remediated or accepted.
Optimisation and Readiness	91–180	Automation and scoped readiness assessment complete.

13. AI Literacy and Student Policy

Training covers permitted use, data restrictions, verification, automation bias, override, incident reporting, academic integrity and research confidentiality. Student policy covers permitted/prohibited use, disclosure, monitoring limits, alternatives and appeals.

14. Differentiated Scenarios

EDU-THR-001 — Student data uploaded to public model

Failure path: Unauthorised disclosure through prompts/files. Prevention: EDU-THR-001: Restrict approved services and fields; apply DLP, minimisation, access control, retention and contractual no-training terms. Detection: EDU-THR-001: Alert on protected-data patterns, unusual uploads, unapproved domains, cross-tenant access and deletion failures. Response: EDU-THR-001: Disable the route, preserve prompts/files/logs, determine affected data and recipients, complete privacy/legal review, notify where required and verify deletion. Evidence: EDU-THR-001: Data-flow map; DLP/configuration; access logs; provider terms; deletion confirmation; incident decision.

EDU-THR-002 — Model trained on student work without approval

Failure path: Secondary use beyond stated purpose. Prevention: EDU-THR-002: Assign ownership, approved purpose, risk classification, supplier controls, human authority and periodic review. Detection: EDU-THR-002: Use inventory reconciliation,

outcome sampling, complaints, audit and change monitoring. Response: EDU-THR-002: Restrict operation, preserve evidence, conduct human review and close governance gaps before return to service. Evidence: EDU-THR-002: Inventory; risk assessment; approval; audit trail; exception, incident and remediation records.

EDU-THR-003 — Prompt injection reveals assessment guidance

Failure path: Malicious instructions override system controls. Prevention: EDU-THR-003: Require representative validation, qualified human decision authority, reasons, subgroup review and accessible appeal. Detection: EDU-THR-003: Monitor agreement with qualified reviewers, overrides, subgroup errors, complaints and appeal reversals. Response: EDU-THR-003: Suspend adverse use, preserve decision traces, conduct human reassessment, correct records and communicate appeal rights. Evidence: EDU-THR-003: Validation dataset; reviewer records; score/rationale trace; subgroup metrics; appeals and corrections.

EDU-THR-004 — AI detector produces false misconduct allegation

Failure path: Probabilistic score treated as proof. Prevention: EDU-THR-004: Require representative validation, qualified human decision authority, reasons, subgroup review and accessible appeal. Detection: EDU-THR-004: Monitor agreement with qualified reviewers, overrides, subgroup errors, complaints and appeal reversals. Response: EDU-THR-004: Suspend adverse use, preserve decision traces, conduct human reassessment, correct records and communicate appeal rights. Evidence: EDU-THR-004: Validation dataset; reviewer records; score/rationale trace; subgroup metrics; appeals and corrections.

EDU-THR-005 — Automated grading systematically mis-scores a group

Failure path: Biased features or drift. Prevention: EDU-THR-005: Require representative validation, qualified human decision authority, reasons, subgroup review and accessible appeal. Detection: EDU-THR-005: Monitor agreement with qualified reviewers, overrides, subgroup errors, complaints and appeal reversals. Response: EDU-THR-005: Suspend adverse use, preserve decision traces, conduct human reassessment, correct records and communicate appeal rights. Evidence: EDU-THR-005: Validation dataset; reviewer records; score/rationale trace; subgroup metrics; appeals and corrections.

EDU-THR-006 — Examination content leakage

Failure path: Weak access control or logging. Prevention: EDU-THR-006: Assign ownership, approved purpose, risk classification, supplier controls, human authority and periodic review. Detection: EDU-THR-006: Use inventory reconciliation, outcome sampling, complaints, audit and change monitoring. Response: EDU-THR-006: Restrict operation, preserve evidence, conduct human review and close governance gaps before return to service. Evidence: EDU-THR-006: Inventory; risk assessment; approval; audit trail; exception, incident and remediation records.

EDU-THR-007 — Deepfake impersonation of administrator

Failure path: Synthetic media induces payment or disclosure. Prevention: EDU-THR-007: Require out-of-band verification, strong identity proofing, transaction controls and role-specific fraud training. Detection: EDU-THR-007: Monitor anomalous requests, payment changes, identity mismatches and high-risk communication patterns. Response: EDU-THR-007: Place holds, verify identity through a trusted channel, preserve media/metadata and issue corrective communication. Evidence: EDU-THR-007: Verification records; training drills; fraud alerts; media evidence; correction log.

EDU-THR-008 — Account takeover alters student records

Failure path: Credential theft and excessive privilege. Prevention: EDU-THR-008: Assign ownership, approved purpose, risk classification, supplier controls, human authority and periodic review. Detection: EDU-THR-008: Use inventory reconciliation, outcome sampling, complaints, audit and change monitoring. Response: EDU-THR-008: Restrict operation, preserve evidence, conduct human review and close governance gaps before return to service. Evidence: EDU-THR-008: Inventory; risk assessment; approval; audit trail; exception, incident and remediation records.

EDU-THR-009 — Wellbeing chatbot gives unsafe advice

Failure path: Hallucination or missed escalation. Prevention: EDU-THR-009: Define prohibited advice, trained responders, routing rules, response times, confidentiality boundaries and tested high-risk disclosures. Detection: EDU-THR-009: Monitor escalation timeliness, missed-case reviews, false alerts, complaints and vendor/model changes. Response: EDU-THR-009: Route to trained personnel, preserve minimum necessary evidence, provide human support and review affected cases. Evidence: EDU-THR-009: Safeguarding assessment; test transcripts; routing logs; responder actions; error review.

EDU-THR-010 — Safeguarding alert is missed

Failure path: False negative or routing failure. Prevention: EDU-THR-010: Define prohibited advice, trained responders, routing rules, response times, confidentiality boundaries and tested high-risk disclosures. Detection: EDU-THR-010: Monitor escalation timeliness, missed-case reviews, false alerts, complaints and vendor/model changes. Response: EDU-THR-010: Route to trained personnel, preserve minimum necessary evidence, provide human support and review affected cases. Evidence: EDU-THR-010: Safeguarding assessment; test transcripts; routing logs; responder actions; error review.

EDU-THR-011 — Safeguarding alert causes unjustified intervention

Failure path: False positive and automation bias. Prevention: EDU-THR-011: Define prohibited advice, trained responders, routing rules, response times, confidentiality boundaries and tested high-risk disclosures. Detection: EDU-THR-011: Monitor escalation timeliness, missed-case reviews, false alerts, complaints and vendor/model changes. Response: EDU-THR-011: Route to trained personnel, preserve minimum necessary evidence, provide human support and review affected cases. Evidence: EDU-THR-011: Safeguarding assessment; test transcripts; routing logs; responder actions; error review.

EDU-THR-012 — Biometric attendance data reused

Failure path: Purpose creep and weak retention. Prevention: EDU-THR-012: Restrict approved services and fields; apply DLP, minimisation, access control, retention and contractual no-training terms. Detection: EDU-THR-012: Alert on protected-data patterns, unusual uploads, unapproved domains, cross-tenant access and deletion failures. Response: EDU-THR-012: Disable the route, preserve prompts/files/logs, determine affected data and recipients, complete privacy/legal review, notify where required and verify deletion. Evidence: EDU-THR-012: Data-flow map; DLP/configuration; access logs; provider terms; deletion confirmation; incident decision.

15. Evidence Catalogue

Evidence type	Minimum content	Verification
Policy	Purpose, scope, authority, requirements, exceptions, enforcement and review cycle.	Verify approval, version, system linkage and exceptions.
AI system inventory	System/model/vendor, owner, purpose, users, data, integrations, risk, approval and lifecycle dates.	Reconcile samples from procurement, network and user discovery.
Risk or impact assessment	Use case, affected persons, harms, likelihood, controls, residual risk, owner and decision.	Verify ratings and implemented treatments.
Privacy assessment	Roles, lawful basis, data map, minimisation, retention, rights, transfers and residual risk.	Trace sampled data and rights requests.
Safeguarding review	Scope, harmful scenarios, escalation, response time, confidentiality and error testing.	Test high-risk disclosures and routing.
Accessibility test	Standard, assistive technologies, tasks, defects, alternatives and remediation.	Repeat tests after change and validate with users.
Vendor due diligence	Responses, evidence, gaps, risk decision, contract actions and sign-off.	Confirm scope, date and product match.
Contract	Data use, security, incidents, change, audit, accessibility, continuity, liability and exit.	Compare terms with configuration and practice.
Architecture/data-flow diagram	Trust boundaries, identities, stores, tools, vendors, logs, regions and controls.	Walk a transaction against live configuration.
Model card	Purpose, populations, data, limitations, evaluations, prohibited uses and oversight.	Compare with institution testing and live behaviour.
Penetration/adversarial test	Scope, environment, methods, cases, evidence, severity, limitations and retest.	Verify independence and closure.
Quality/bias evaluation	Dataset, population, metrics, subgroup results, thresholds and uncertainty.	Recalculate samples and assess representativeness.
Log/decision trace	Timestamp, identity, model/version, input reference, output/action and human decision.	Reconstruct a sampled decision.
Incident record	Detection, severity, affected persons, containment, evidence, notification and correction.	Verify timeline and remediation.
Appeal/correction record	Original outcome, grounds, evidence, reviewer, decision and correction.	Sample for independence and timeliness.
Training record	Audience, role-specific content, completion, assessment and refresh.	Verify competence, not attendance alone.
Committee minutes	Attendees, conflicts, materials, decisions, dissent and actions.	Confirm authority and closure.
Change/release record	Classification, risk, tests, approvals, deployment, monitoring and rollback.	Trace a material change.
Retirement/deletion record	Dependencies, export, notices, retention, deletion and verification.	Verify absence of access and deletion evidence.

Implementation-stage boundary. These education capability stages are a sector self-assessment aid only. They do not replace, rename or modify GAISSF Foundational, Operational or Optimized conformance tiers.

16. Education Implementation Stages

The workbook contains 70 observable criteria across 14 dimensions and five levels. Level 3 requires consistent implementation and evidence, not merely reproducing controls.

17. Notably Absent

Area	Not identified or excluded	Implication
Evidence scope	No institution-specific frequency, loss estimate or universal error rate.	Prevents unsupported quantitative claims.
Jurisdiction	No determination for every jurisdiction or institution type.	Local legal review required.
Specialised institutions	Military, clinical teaching and highly regulated research profiles.	Additional specialist profiles may be needed.
Dual-use research	Detailed biological, radiological and offensive cyber scenarios.	Dedicated research-security analysis required.
Product endorsement	No vendor or model approval or safety claim.	Guidance is not product assurance.
Named incidents	No scenario is presented as a named institutional incident.	Separates scenarios from incident evidence.

18. Glossary

Term	Definition
GAISSF	Global AI Safety and Security Framework.
RAG	Red-Amber-Green-Blue status.
Competent reviewer	Trained and authorised reviewer with evidence access and override authority.
Independent assessor	Assessor without operational responsibility for the system.
MCP	Model Context Protocol for connecting AI to tools or data.
LoRA	Low-Rank Adaptation, an adapter-based model modification method.

19. Quality Assurance

Check	Result	Evidence
Control completeness	Pass	59 unique IDs and titles.
Differentiation	Pass	Controls, threats, use cases and VDD differentiated.
Education implementation stages	Pass	70 unique criteria.
Roadmap	Pass	Separate 0–90 and 0–180 artefacts.
Publication readiness	Blocked	Specialist approvals remain.

Substantive refinement is complete. Final publication remains blocked until legal, safeguarding, accessibility and editorial approvals are recorded.