GAISSF
LEGAL SECTOR
IMPLEMENTATION GUIDE

SEC-048 | Version 1.0

Controlled publication candidate - release blocked pending external legal/professional review and final publication approvals

ODA3 Institute

Document Control

Document ID	SEC-048
Version	1.0
Status	Controlled publication candidate - release blocked pending external legal/professional review and final publication approvals
Publisher	ODA3 Institute
Legal entity	ODA3 Pvt Ltd (legal notices only)
Classification	Informative sector implementation guidance
Authoritative baseline	GAISSF-NOR-001 v1.0; GAISSF-NOR-004 v1.0; 59 controls across D1-D9
Publication date	Not assigned - release blocked
Licence / trademark	Final approved wording required

Important Notice

This document provides non-normative sector implementation guidance. It does not amend, replace or override GAISSF-NOR-001 or applicable law, court or tribunal rules, professional duties, client instructions or contractual commitments. It is not legal, professional-conduct, regulatory, litigation, evidentiary or privilege advice. Jurisdiction-, forum-, client-, matter- and use-case-specific verification remains required.

1. Executive Summary

Legal AI risk concentrates around authority integrity, confidentiality, privilege, client and matter separation, evidence reliability, professional responsibility, provider dependence and agentic action. This revision provides 59 control-specific legal interpretations and machine-readable traceability to the authoritative GAISSF baseline.

Five Leadership Decisions

- Authorised, restricted and prohibited AI uses
- Permitted information classes by deployment model
- Required competent human review by legal consequence
- Client, matter, forum and professional restrictions that must be technically enforced
- Evidence retained for accountable and reproducible operation

Notably Absent

- No claim that AI use is inherently incompatible with legal professional duties.
- No claim that use of an AI provider automatically preserves or waives privilege.
- No claim that GAISSF conformance establishes legal, ethical, regulatory or evidentiary compliance.
- No claim that AI-generated legal work is inherently admissible or inadmissible.
- No claim that human review eliminates all AI-related risk.
- No claim that private deployment automatically ensures confidentiality.
- No claim that provider certification establishes operating effectiveness for a specific legal use.
- No claim that citation functionality proves authority validity.
- No claim that SEC-048 determines whether a court, tribunal or regulator requires disclosure of AI use.
- No claim that technical safeguards alone preserve privilege, work product or professional secrecy.
- No jurisdiction-specific conclusion on unauthorised practice, attorney-client relationship formation, fee ethics or disclosure duties.

2. Risk Taxonomy and Scoring

1. Confidentiality and professional secrecy
2. Privilege and work-product protection
3. Authority and citation integrity
4. Evidence integrity, provenance and procedural reliability
5. Professional responsibility, competence and supervision
6. Client and matter separation
7. Privacy and data protection
8. Bias, fairness and rights impact
9. Intellectual property and licence risk
10. Cybersecurity and adversarial manipulation
11. Operational quality and reproducibility
12. Financial, reputational and liability exposure
13. Unauthorised legal-service delivery and relationship formation
14. Judicial, tribunal and regulatory disclosure compliance
15. Billing and fee integrity

Score = Likelihood x Impact. 1-4 Low; 5-9 Moderate; 10-16 High; 17-25 Critical. Local governance may impose stricter thresholds.

High and Critical residual risks require named accountable approval; Critical risks ordinarily require avoidance or reduction before use.

3. Use-Case Catalogue

ID	Use case	Primary risk	Domains	Priority controls
UC-01	Legal research and authority identification	Authority and citation integrity	D1, D5, D6, D8	D1-CTL-03, D5-CTL-01, D6-CTL-01, D6-CTL-02
UC-02	Case-law and legislation summarisation	Authority and citation integrity	D1, D5, D6	D1-CTL-03, D5-CTL-01, D6-CTL-01
UC-03	Drafting legal documents	Professional responsibility, competence and supervision	D5, D6, D8	D5-CTL-01, D6-CTL-01, D6-CTL-02
UC-04	Contract review and deviation analysis	Operational quality and reproducibility	D1, D5, D6	D1-CTL-03, D5-CTL-01, D6-CTL-01
UC-05	Due diligence	Confidentiality and professional secrecy	D1, D4, D5, D6	D4-CTL-05, D5-CTL-02, D6-CTL-02
UC-06	E-discovery and technology-assisted review	Evidence integrity, provenance and procedural reliability	D1, D2, D5, D6	D1-CTL-01, D2-CTL-02, D6-CTL-02
UC-07	Litigation and arbitration preparation	Professional responsibility, competence and supervision	D5, D6, D8	D5-CTL-01, D6-CTL-01, D8-CTL-03
UC-08	Investigations and regulatory response	Bias, fairness and rights impact	D5, D6, D7, D8	D5-CTL-02, D6-CTL-01, D6-CTL-02
UC-09	Transcription and	Evidence integrity,	D1, D5, D6	D1-CTL-03, D6-

ID	Use case	Primary risk	Domains	Priority controls
	translation	provenance and procedural reliability		CTL-01, D6-CTL-02
UC-10	Evidence analysis	Evidence integrity, provenance and procedural reliability	D1, D2, D5, D6	D1-CTL-01, D2-CTL-04, D6-CTL-02
UC-11	Matter intake and triage	Unauthorised legal-service delivery and relationship formation	D3, D5, D6, D8	D3-CTL-01, D6-CTL-01, D6-CTL-02
UC-12	Conflicts screening	Client and matter separation	D1, D5, D6	D1-CTL-03, D5-CTL-02, D6-CTL-02
UC-13	Legal holds and records management	Evidence integrity, provenance and procedural reliability	D3, D6, D8	D3-CTL-07, D6-CTL-02, D6-CTL-05
UC-14	Knowledge management and precedent retrieval	Client and matter separation	D1, D2, D5, D6	D1-CTL-05, D2-CTL-06, D5-CTL-02
UC-15	Client-facing assistants	Unauthorised legal-service delivery and relationship formation	D3, D5, D6, D8	D3-CTL-01, D5-CTL-01, D6-CTL-01
UC-16	Billing and pricing analysis	Billing and fee integrity	D5, D6, D8	D6-CTL-01, D6-CTL-02, D8-CTL-03
UC-17	Outcome prediction	Bias, fairness and rights impact	D1, D5, D6, D8	D1-CTL-03, D6-CTL-01, D6-CTL-03
UC-18	Privilege and confidentiality review	Privilege and work-product protection	D4, D5, D6	D4-CTL-05, D5-CTL-02, D6-CTL-02
UC-19	Redaction and de-identification	Privacy and data protection	D5, D6	D5-CTL-02, D5-CTL-05, D6-CTL-02
UC-20	Regulatory horizon scanning	Judicial, tribunal and regulatory disclosure compliance	D6, D8	D6-CTL-02, D8-CTL-01, D8-CTL-03
UC-21	Court, tribunal and regulator disclosure tracking	Judicial, tribunal and regulatory disclosure compliance	D6, D8	D6-CTL-02, D8-CTL-03
UC-22	Agentic external communication and filing workflows	Professional responsibility, competence and supervision	D2, D3, D6, D8	D2-CTL-05, D3-CTL-01, D6-CTL-01

4. Foundational Implementation Principles

- Verify primary authorities before material reliance.
- Match human review depth to legal consequence.
- Technically enforce client and matter boundaries.
- Do not infer privilege from a system label.
- Track forum-specific AI disclosure and certification duties.

- Document system-specific reviewer competence.
- Separate AI processing, professional time and client billing rules.
- Treat provider certifications as supporting evidence, not proof.

Agentic System Restrictions

- File or lodge a document
- Submit material to a court, tribunal or regulator
- Send formal external legal communications
- Accept or vary contract terms or settlement offers
- Waive rights, privilege or confidentiality
- Make payments or financial commitments
- Modify client/matter access or ethical walls
- Delete or alter preserved records
- Create, release or modify a legal hold
- Represent itself as an authorised lawyer or provide personalised advice outside an approved supervised workflow

5. Threat Scenarios

TS-01 - Fabricated legal authorities in a filed document

Primary risk: Authority and citation integrity
Relevant controls: D5-CTL-01, D6-CTL-01, D6-CTL-02
Prevent: Primary-authority verification gate before filing
Detect: Citation exception analytics and filing checklist
Respond: Withdraw or correct as required; preserve prompt, output and verification record
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-02 - Reliance on overruled or obsolete authority

Primary risk: Authority and citation integrity
Relevant controls: D1-CTL-03, D6-CTL-01
Prevent: Current-status validation against authoritative source
Detect: Shepardisation/citator discrepancy review
Respond: Reassess advice or filing and document corrective action
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-03 - Privileged material submitted to an unauthorised public AI service

Primary risk: Privilege and work-product protection
Relevant controls: D4-CTL-06, D5-CTL-02, D6-CTL-04
Prevent: Approved-tool allowlist, DLP and matter-data classification
Detect: DLP alerts and shadow-AI monitoring
Respond: Contain disclosure, preserve facts, conduct jurisdiction-specific privilege analysis
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-04 - Provider retention of confidential client data

Primary risk: Confidentiality and professional secrecy
Relevant controls: D4-CTL-05, D6-CTL-06
Prevent: Contract and configuration review of retention, training and deletion
Detect: Provider assurance and configuration drift review
Respond: Suspend transfer, invoke deletion/return rights and assess notice duties
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-05 - Cross-client retrieval through a shared knowledge base

Primary risk: Client and matter separation
Relevant controls: D1-CTL-05, D2-CTL-06, D5-CTL-02
Prevent: Matter-scoped indexes, ACLs and retrieval filters
Detect: Cross-matter canary tests and access-log review
Respond: Disable affected index and investigate all potentially exposed matters
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-06 - Prompt injection in discovery material exfiltrates matter information

Primary risk: Cybersecurity and adversarial manipulation
Relevant controls: D2-CTL-02, D2-CTL-05, D6-CTL-04
Prevent: Treat retrieved documents as untrusted; isolate tools and secrets
Detect: Injection test corpus, blocked-call logs and exfiltration alerts
Respond: Contain system, preserve malicious source and assess disclosure impact
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-07 - Malicious document manipulates an AI review agent

Primary risk: Cybersecurity and adversarial manipulation
Relevant controls: D2-CTL-02, D2-CTL-04, D3-CTL-03
Prevent: Sanitise and sandbox multimodal inputs; constrain agent actions
Detect: Agent trace review and anomalous-action detection
Respond: Quarantine document and revoke agent credentials or tools
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-08 - AI omits a critical contract schedule or annex

Primary risk: Operational quality and reproducibility
Relevant controls: D1-CTL-03, D6-CTL-01, D6-CTL-02
Prevent: Completeness checks against source document manifest
Detect: Page/section reconciliation and exception reporting
Respond: Repeat review from verified source package and notify accountable lawyer
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-09 - AI misidentifies parties or governing law

Primary risk: Professional responsibility, competence and supervision
Relevant controls: D5-CTL-01, D6-CTL-01

Prevent: Structured party, jurisdiction and governing-law verification
Detect: Entity mismatch and jurisdiction conflict checks
Respond: Correct work product and reassess affected advice or obligations
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-10 - Automated redaction fails

Primary risk: Privacy and data protection
Relevant controls: D5-CTL-02, D5-CTL-05, D6-CTL-01
Prevent: Dual-pass redaction with visual and text-layer verification
Detect: Residual-text and metadata scan
Respond: Recall or replace document and assess notification requirements
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-11 - Transcription materially alters testimony

Primary risk: Evidence integrity, provenance and procedural reliability
Relevant controls: D1-CTL-03, D6-CTL-01, D6-CTL-02
Prevent: Retain source recording and require material-statement verification
Detect: Confidence threshold and random sample comparison
Respond: Correct transcript while retaining original and amendment history
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-12 - Synthetic evidence accepted without provenance checks

Primary risk: Evidence integrity, provenance and procedural reliability
Relevant controls: D5-CTL-04, D7-CTL-H02, D6-CTL-02
Prevent: Provenance and authenticity checks; preserve original media
Detect: Deepfake indicators and metadata anomaly review
Respond: Escalate to forensic review and preserve chain of custody
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-13 - Automated deletion defeats a legal hold

Primary risk: Evidence integrity, provenance and procedural reliability
Relevant controls: D3-CTL-07, D6-CTL-05, D6-CTL-02
Prevent: Hold-aware deletion overrides and immutable preservation rules
Detect: Deletion-event reconciliation against hold register
Respond: Stop deletion, restore available copies and document preservation gap
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-14 - Agent sends an unauthorised communication

Primary risk: Professional responsibility, competence and supervision
Relevant controls: D2-CTL-05, D3-CTL-01, D6-CTL-01
Prevent: External-send gate requiring named human approval
Detect: Outbound communication and tool-call monitoring
Respond: Recall where possible; notify responsible counsel and preserve full trace

Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-15 - Agent accepts terms or creates a binding obligation

Primary risk: Financial, reputational and liability exposure
Relevant controls: D3-CTL-01, D3-CTL-05, D6-CTL-01
Prevent: Deny contractual acceptance and payment tools by default
Detect: Binding-action alerting and transaction reconciliation
Respond: Revoke authority, contest action where appropriate and preserve evidence
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-16 - AI-assisted billing creates duplicate or unreasonable entries

Primary risk: Billing and fee integrity
Relevant controls: D6-CTL-01, D6-CTL-02, D8-CTL-03
Prevent: Separate AI processing from professional time and require invoice review
Detect: Duplicate, anomaly and client-rule checks
Respond: Correct invoice, assess client disclosure and retain review record
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-17 - Client-facing chatbot provides unauthorised personalised advice

Primary risk: Unauthorised legal-service delivery and relationship formation
Relevant controls: D3-CTL-01, D5-CTL-01, D6-CTL-01
Prevent: Limit to approved information; identity, jurisdiction and relationship disclaimers; escalation
Detect: Conversation sampling and personalised-advice detection
Respond: Suspend workflow, route affected user to authorised professional and assess consequences
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-18 - Litigation prediction introduces unexamined bias

Primary risk: Bias, fairness and rights impact
Relevant controls: D1-CTL-03, D6-CTL-01, D6-CTL-03
Prevent: Validate populations, intended use and prohibited reliance
Detect: Outcome disparity and drift monitoring
Respond: Withdraw model from decision support and reassess affected recommendations
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-19 - Provider silently changes the model

Primary risk: Operational quality and reproducibility
Relevant controls: D1-CTL-03, D4-CTL-05, D6-CTL-06
Prevent: Contractual notice, version pinning and change gates
Detect: Fingerprint, benchmark and output-drift monitoring
Respond: Suspend high-risk use pending revalidation
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-20 - Relied-on output cannot be reproduced

Primary risk: Operational quality and reproducibility
Relevant controls: D6-CTL-02, D6-CTL-03
Prevent: Record model/version, prompt, sources, parameters and reviewer
Detect: Reproduction sampling and missing-record alerts
Respond: Reconstruct where possible and qualify evidentiary reliance
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-21 - Translation changes legal effect

Primary risk: Professional responsibility, competence and supervision
Relevant controls: D5-CTL-01, D6-CTL-01
Prevent: Qualified bilingual review for legally operative text
Detect: Terminology and clause-difference checks
Respond: Correct translation and reassess dependent action
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-22 - Conflicts screening fails through entity-resolution error

Primary risk: Client and matter separation
Relevant controls: D1-CTL-03, D6-CTL-01, D6-CTL-02
Prevent: Authoritative entity sources, aliases and human conflict clearance
Detect: False-negative sampling and later-match alerts
Respond: Freeze engagement decisions and perform manual conflicts analysis
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-23 - Investigation tool wrongly attributes misconduct

Primary risk: Bias, fairness and rights impact
Relevant controls: D5-CTL-01, D6-CTL-01, D6-CTL-02
Prevent: Corroboration requirement; prohibit sole-source adverse decisions
Detect: Attribution-confidence and contradictory-evidence review
Respond: Correct record, suspend adverse action and preserve investigation trail
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-24 - Logs disclose confidential strategy

Primary risk: Confidentiality and professional secrecy
Relevant controls: D5-CTL-02, D6-CTL-02
Prevent: Minimise and segregate logs; redact content and restrict access
Detect: Sensitive-term scanning and privileged-access review
Respond: Restrict logs, rotate access and conduct disclosure analysis
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-25 - Legal technology provider suffers a security breach

Primary risk: Confidentiality and professional secrecy
Relevant controls: D4-CTL-05, D6-CTL-04, D6-CTL-06

Prevent: Provider security due diligence, contractual notification and exit planning
Detect: Provider alerts, threat intelligence and independent assurance review
Respond: Activate incident plan, identify affected matters and assess notification
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

TS-26 - Court-required AI disclosure or certification is omitted

Primary risk: Judicial, tribunal and regulatory disclosure compliance
Relevant controls: D6-CTL-02, D8-CTL-03
Prevent: Forum-specific disclosure register and filing gate
Detect: Matter checklist and rule-change monitoring
Respond: Assess corrective filing and preserve verification record
Evidence: Source inputs, model/version, prompts, outputs, approvals, logs, notifications and corrective-action record
Limit: Requires organisation-, matter-, forum- and jurisdiction-specific validation

6. GAISSF Control Interpretation Profile

Normative wording is not reproduced. Each record references the authoritative control. GAISSF-NOR-001 and GAISSF-NOR-004 prevail in a conflict.

D1-CTL-01 - DATASET PROVENANCE & POISONING PREVENTION

Domain: D1 — MODEL INTEGRITY & ADVERSARIAL ROBUSTNESS

Normative reference: GAISSF-NOR-004 v1.0, control D1-CTL-01; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Verify the origin, authority, permitted use and integrity of legal corpora, precedent banks, discovery datasets and matter uploads before they influence legal analysis.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Dataset manifests, source licences, ingestion approvals, hash records and poisoning tests

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Unverified or adversarial source material can distort research, discovery prioritisation and legal conclusions.

D1-CTL-02 - MODEL EXTRACTION RESISTANCE

Domain: D1 — MODEL INTEGRITY & ADVERSARIAL ROBUSTNESS

Normative reference: GAISSF-NOR-004 v1.0, control D1-CTL-02; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Protect proprietary legal models, prompts, embeddings and client-derived tuning artefacts from extraction through repeated queries or exposed interfaces.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Rate-limit evidence, extraction tests, API access logs and incident records

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Controls must not obstruct lawful client access, discovery duties or authorised portability.

D1-CTL-03 - BEHAVIORAL DRIFT DETECTION

Domain: D1 — MODEL INTEGRITY & ADVERSARIAL ROBUSTNESS

Normative reference: GAISSF-NOR-004 v1.0, control D1-CTL-03; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Detect changes in citation reliability, jurisdictional interpretation, refusal behaviour, bias and completeness after model or provider updates.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Versioned benchmark results, drift thresholds, change approvals and rollback records

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Stable aggregate accuracy does not prove stable performance for a specific jurisdiction or practice area.

D1-CTL-04 - FEDERATED LEARNING POISONING PREVENTION

Domain: D1 — MODEL INTEGRITY & ADVERSARIAL ROBUSTNESS

Normative reference: GAISSF-NOR-004 v1.0, control D1-CTL-04; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Where legal organisations participate in federated learning, validate contributor identity, update quality and resistance to malicious or cross-client influence.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Participant register, signed updates, anomaly reports and aggregation tests

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Federated architecture does not by itself preserve confidentiality or privilege.

D1-CTL-05 - EMBEDDING SPACE ROBUSTNESS

Domain: D1 — MODEL INTEGRITY & ADVERSARIAL ROBUSTNESS

Normative reference: GAISSF-NOR-004 v1.0, control D1-CTL-05; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Test legal retrieval embeddings for cross-matter leakage, adversarial nearest-neighbour manipulation, entity confusion and omission of controlling authority.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Retrieval tests, cross-matter canaries, embedding-version records and exception logs

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Retrieval similarity is not legal relevance or authority.

D1-CTL-06 - POST-QUANTUM MODEL SIGNING & CRYPTO HARDENING

Domain: D1 — MODEL INTEGRITY & ADVERSARIAL ROBUSTNESS

Normative reference: GAISSF-NOR-004 v1.0, control D1-CTL-06; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Protect long-lived legal records, model artefacts and provenance signatures with a documented cryptographic migration strategy proportionate to confidentiality duration.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Cryptographic inventory, signing verification, key-rotation and migration plan

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: This control does not require premature deployment of unapproved cryptography.

D1-CTL-07 - LORA/ADAPTER INTEGRITY VERIFICATION

Domain: D1 — MODEL INTEGRITY & ADVERSARIAL ROBUSTNESS

Normative reference: GAISSF-NOR-004 v1.0, control D1-CTL-07; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Verify legal-domain adapters and fine-tunes before deployment, including source, authorisation, hash, test results and client-data contamination risk.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Adapter inventory, hashes, approvals, validation tests and rollback package

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: An adapter labelled for legal use is not evidence of legal accuracy.

D1-CTL-08 - MODEL MERGE ATTACK DETECTION

Domain: D1 — MODEL INTEGRITY & ADVERSARIAL ROBUSTNESS

Normative reference: GAISSF-NOR-004 v1.0, control D1-CTL-08; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Detect unauthorised or unsafe merging of legal-domain models that could introduce hidden behaviour, licensing conflicts or data leakage.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Merge provenance, component hashes, differential tests and approval records

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Merged model performance must be validated for the actual legal use case.

D1-CTL-09 - QUANTIZATION BACKDOOR SCREENING

Domain: D1 — MODEL INTEGRITY & ADVERSARIAL ROBUSTNESS

Normative reference: GAISSF-NOR-004 v1.0, control D1-CTL-09; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Test quantised legal models for changed refusal, citation, confidentiality and tool-use behaviour before replacing the validated model.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Pre/post quantisation benchmark, backdoor tests and release approval

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Size or cost optimisation does not justify reduced legal assurance.

D2-CTL-01 - DIRECT PROMPT INJECTION PREVENTION

Domain: D2 — RUNTIME SECURITY & ADVERSARIAL DEFENSE

Normative reference: GAISSF-NOR-004 v1.0, control D2-CTL-01; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Prevent users from overriding legal-system instructions, matter restrictions, disclosure limits or review gates through direct prompts.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Adversarial prompt tests, blocked requests and policy-enforcement logs

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Prompt filters alone are not sufficient for privileged or high-impact workflows.

D2-CTL-02 - INDIRECT PROMPT INJECTION PREVENTION

Domain: D2 — RUNTIME SECURITY & ADVERSARIAL DEFENSE

Normative reference: GAISSF-NOR-004 v1.0, control D2-CTL-02; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Treat contracts, emails, websites, discovery documents and knowledge-base content as untrusted instructions that must not control the legal AI workflow.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Document-borne injection corpus, sanitisation results and tool-call logs

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Content authenticity and instruction safety are separate determinations.

D2-CTL-03 - JAILBREAK RESISTANCE TESTING

Domain: D2 — RUNTIME SECURITY & ADVERSARIAL DEFENSE

Normative reference: GAISSF-NOR-004 v1.0, control D2-CTL-03; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Test whether users can bypass restrictions on personalised advice, confidential data disclosure, harmful content or unauthorised legal actions.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Jailbreak test plan, results, remediation and retest evidence

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: A passed test set does not establish universal resistance.

D2-CTL-04 - MULTI-MODAL INJECTION DEFENSE

Domain: D2 — RUNTIME SECURITY & ADVERSARIAL DEFENSE

Normative reference: GAISSF-NOR-004 v1.0, control D2-CTL-04; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Inspect text, images, audio, metadata and embedded objects in legal evidence and filings for hidden instructions or manipulative content.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Multimodal test corpus, content sanitisation logs and analyst review

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Sanitisation must preserve evidentiary originals and chain of custody.

D2-CTL-05 - FUNCTION CALL/TOOL CALL INJECTION PREVENTION

Domain: D2 — RUNTIME SECURITY & ADVERSARIAL DEFENSE

Normative reference: GAISSF-NOR-004 v1.0, control D2-CTL-05; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Validate every requested tool action against matter scope, user authority and an explicit allowlist before execution.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Tool policy, approval records, denied-call logs and transaction reconciliation

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: The model must not be treated as the authority to approve its own action.

D2-CTL-06 - CROSS-CONTEXT HIJACKING MITIGATION

Domain: D2 — RUNTIME SECURITY & ADVERSARIAL DEFENSE

Normative reference: GAISSF-NOR-004 v1.0, control D2-CTL-06; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Prevent one client, matter, conversation or retrieved source from influencing another context through shared memory, caches or prompts.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Context-isolation tests, cache controls, canary records and access logs

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Logical separation must be tested; naming conventions are insufficient.

D3-CTL-01 - LEAST AGENCY ENFORCEMENT

Domain: D3 — AGENTIC RISK & AUTONOMOUS SYSTEM SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D3-CTL-01; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Limit agents to the minimum legal tasks and permissions required. Deny filing, external communication, contract acceptance, settlement, payment, waiver, deletion and legal-hold modification unless a named human authorises the specific action.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Agent permission matrix, approval gates, denied-action logs and periodic recertification

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: This guidance does not determine whether a particular act is legally delegable.

D3-CTL-02 - INTER-AGENT COMMUNICATION SECURITY

Domain: D3 — AGENTIC RISK & AUTONOMOUS SYSTEM SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D3-CTL-02; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Authenticate and constrain messages between research, drafting, review and filing agents so instructions and matter data cannot be spoofed or crossed.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Agent identities, message signatures, routing policy and trace logs

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Authenticated messages may still contain incorrect legal content.

D3-CTL-03 - AGENTIC PROMPT CHAINING DETECTION

Domain: D3 — AGENTIC RISK & AUTONOMOUS SYSTEM SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D3-CTL-03; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Detect chained instructions that gradually bypass client restrictions, review gates or limits on personalised legal advice.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Chain traces, policy-violation alerts and adversarial tests

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Single-step review may miss cumulative effects.

D3-CTL-04 - EMBODIED AI SAFETY CONTROLS

Domain: D3 — AGENTIC RISK & AUTONOMOUS SYSTEM SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D3-CTL-04; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Where legal operations use robots or embodied systems for evidence, records or secure facilities, constrain physical actions and preserve accountable human control.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Physical-use inventory, safety tests and override evidence

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Often not applicable; applicability must be documented rather than assumed.

D3-CTL-05 - MULTI-AGENT TRUST CHAIN ATTESTATION

Domain: D3 — AGENTIC RISK & AUTONOMOUS SYSTEM SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D3-CTL-05; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Maintain verifiable identity, version, permissions and hand-off records across agents contributing to legal work product.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Agent attestation, chain-of-custody record and hand-off approvals

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: A complete chain does not prove substantive legal correctness.

D3-CTL-06 - PERSISTENT MEMORY EXFILTRATION PREVENTION

Domain: D3 — AGENTIC RISK & AUTONOMOUS SYSTEM SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D3-CTL-06; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Prevent long-term agent memory from exposing client strategy, privileged content or data from closed matters.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Memory access tests, DLP logs, isolation evidence and incident records

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Memory disablement must be verified, not inferred from interface settings.

D3-CTL-07 - SECURE MEMORY LIFECYCLE MANAGEMENT

Domain: D3 — AGENTIC RISK & AUTONOMOUS SYSTEM SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D3-CTL-07; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Define creation, classification, matter association, retention, legal-hold override, review and deletion for agent memory.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Memory register, retention schedule, hold overrides and deletion verification

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Deletion must not conflict with preservation duties.

D4-CTL-01 - AI BILL OF MATERIALS (AI BOM) MAINTENANCE

Domain: D4 — SUPPLY CHAIN & THIRD-PARTY AI SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D4-CTL-01; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Maintain a current inventory of models, datasets, adapters, retrieval stores, tools, APIs and subprocessors used in legal services.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: AI BOM, dependency versions, owner and change history

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: An inventory does not establish supplier assurance.

D4-CTL-02 - MODEL FILE & ARTIFACT SCANNING

Domain: D4 — SUPPLY CHAIN & THIRD-PARTY AI SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D4-CTL-02; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Scan model files, adapters, prompt packages and legal-technology artefacts for malware, unsafe serialization and unauthorised modifications.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Scan results, quarantine records and release approvals

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Scanning cannot detect every behavioural backdoor.

D4-CTL-03 - MODEL HUB & REGISTRY VETTING

Domain: D4 — SUPPLY CHAIN & THIRD-PARTY AI SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D4-CTL-03; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Permit models from approved registries only after provenance, licence, security and legal-use assessment.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Approved registry list, model intake record and licence review

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Popularity or download count is not assurance.

D4-CTL-04 - MCP SERVER BEHAVIORAL MONITORING

Domain: D4 — SUPPLY CHAIN & THIRD-PARTY AI SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D4-CTL-04; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Monitor Model Context Protocol or equivalent tool servers that access document stores, email, calendars, matter systems or filing services.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Server inventory, permission review, call logs and anomaly alerts

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: A server declaration is not evidence of actual behaviour.

D4-CTL-05 - THIRD-PARTY AI API SECURITY ASSESSMENT

Domain: D4 — SUPPLY CHAIN & THIRD-PARTY AI SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D4-CTL-05; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Assess API authentication, encryption, logging, retention, training use, data location, incident terms and matter isolation before transmitting legal data.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: API assessment, contract, configuration evidence and tests

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Provider certifications are supporting evidence, not legal-sector equivalence.

D4-CTL-06 - SHADOW AI DISCOVERY & GOVERNANCE

Domain: D4 — SUPPLY CHAIN & THIRD-PARTY AI SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D4-CTL-06; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Discover unapproved legal AI use, including browser features, embedded copilots and personal accounts, and route it into approval or containment.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Discovery scans, user attestations, DLP alerts and remediation

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Absence of detected use does not prove absence of shadow AI.

D4-CTL-07 - AI SOFTWARE COMPOSITION ANALYSIS (SCA)

Domain: D4 — SUPPLY CHAIN & THIRD-PARTY AI SECURITY

Normative reference: GAISSF-NOR-004 v1.0, control D4-CTL-07; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Track vulnerable and licensed components in legal AI applications, connectors and local inference stacks.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: SCA reports, remediation tickets and licence inventory

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: SCA does not assess model behaviour or legal accuracy.

D5-CTL-01 - HARMFUL CONTENT BLOCKING

Domain: D5 — CONTENT SAFETY & OUTPUT INTEGRITY

Normative reference: GAISSF-NOR-004 v1.0, control D5-CTL-01; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Prevent outputs that are unlawful, abusive, deceptive or unsuitable for the legal use while preserving legitimate legal analysis of harmful material.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Policy tests, override procedure and review logs

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Content restriction must not suppress relevant evidence or defence analysis without documented justification.

D5-CTL-02 - PII LEAKAGE PREVENTION

Domain: D5 — CONTENT SAFETY & OUTPUT INTEGRITY

Normative reference: GAISSF-NOR-004 v1.0, control D5-CTL-02; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Detect and prevent disclosure of personal, confidential, privileged or matter-restricted data in prompts, outputs, logs and external communications.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: DLP tests, redaction verification, leakage alerts and incident records

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: PII controls alone do not protect all confidential legal information.

D5-CTL-03 - COPYRIGHT DETECTION

Domain: D5 — CONTENT SAFETY & OUTPUT INTEGRITY

Normative reference: GAISSF-NOR-004 v1.0, control D5-CTL-03; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Identify potentially protected source material and output overlap before publishing, licensing or incorporating generated legal content.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Source records, similarity checks and licence review

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Detection does not determine infringement or ownership.

D5-CTL-04 - AI WATERMARKING ROBUSTNESS

Domain: D5 — CONTENT SAFETY & OUTPUT INTEGRITY

Normative reference: GAISSF-NOR-004 v1.0, control D5-CTL-04; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Where watermarking or provenance markers are used for legal media, test persistence, false positives and evidentiary limitations.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Watermark tests, provenance records and analyst guidance

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Watermarks are not conclusive proof of authenticity or AI generation.

D5-CTL-05 - PRIVACY-BY-DESIGN VERIFICATION

Domain: D5 — CONTENT SAFETY & OUTPUT INTEGRITY

Normative reference: GAISSF-NOR-004 v1.0, control D5-CTL-05; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Verify minimisation, purpose limitation, access, retention, rights handling and privacy impact controls in legal AI workflows.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Privacy assessment, data map, retention and access evidence

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: This guide does not determine the applicable lawful basis.

D5-CTL-06 - PRIVACY-PRESERVING ML VALIDATION

Domain: D5 — CONTENT SAFETY & OUTPUT INTEGRITY

Normative reference: GAISSF-NOR-004 v1.0, control D5-CTL-06; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Validate claimed privacy-preserving techniques against realistic legal datasets, membership risk and utility degradation.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Technique configuration, attack tests and validation report

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Privacy-enhancing technology does not automatically preserve privilege.

D6-CTL-01 - HUMAN-IN-THE-LOOP FOR HIGH-RISK ACTIONS

Domain: D6 — GOVERNANCE, ACCOUNTABILITY & HUMAN OVERSIGHT

Normative reference: GAISSF-NOR-004 v1.0, control D6-CTL-01; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Require a competent, named legal professional to review legally significant output and all external, binding or rights-affecting actions. Reviewer competence must include system-specific failure modes.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Reviewer qualification, approval record, source verification and escalation log

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Human presence without adequate time, authority or competence is not effective oversight.

D6-CTL-02 - AUDIT TRAIL COMPLETENESS

Domain: D6 — GOVERNANCE, ACCOUNTABILITY & HUMAN OVERSIGHT

Normative reference: GAISSF-NOR-004 v1.0, control D6-CTL-02; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Record client, matter, user, model, version, sources, prompt, output, tools, approvals, disclosures and changes needed to reconstruct material legal AI use.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Immutable or controlled logs, sampling results and access review

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Logging must be minimised and protected to avoid creating a confidentiality risk.

D6-CTL-03 - AI MODEL CARD COMPLETENESS

Domain: D6 — GOVERNANCE, ACCOUNTABILITY & HUMAN OVERSIGHT

Normative reference: GAISSF-NOR-004 v1.0, control D6-CTL-03; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Document intended legal uses, prohibited uses, jurisdictions, known failure modes, training limits, review requirements and change history.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Approved legal-use model card and reviewer sign-off

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: A model card is a disclosure record, not proof of performance.

D6-CTL-04 - AI INCIDENT RESPONSE READINESS

Domain: D6 — GOVERNANCE, ACCOUNTABILITY & HUMAN OVERSIGHT

Normative reference: GAISSF-NOR-004 v1.0, control D6-CTL-04; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Prepare for fabricated authority, confidential disclosure, prompt injection, provider breach, unauthorised action and evidence-integrity incidents.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Playbooks, exercises, contact matrix and retained incident evidence

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Notification and privilege implications require qualified analysis.

D6-CTL-05 - MODEL DEPRECATION & DECOMMISSIONING

Domain: D6 — GOVERNANCE, ACCOUNTABILITY & HUMAN OVERSIGHT

Normative reference: GAISSF-NOR-004 v1.0, control D6-CTL-05; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Retire legal models and integrations with preserved records, client/matter data return or deletion, hold compliance and replacement validation.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Decommission plan, data disposition, hold check and sign-off

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Deactivation does not prove deletion from provider systems.

D6-CTL-06 - THIRD-PARTY AI VENDOR GOVERNANCE

Domain: D6 — GOVERNANCE, ACCOUNTABILITY & HUMAN OVERSIGHT

Normative reference: GAISSF-NOR-004 v1.0, control D6-CTL-06; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Govern legal AI suppliers through due diligence, contract controls, assurance, change notice, incident response, audit, exit and subprocessor oversight.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Due-diligence checklist, contract, assurance review and exit test

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: SOC 2 or ISO certification does not establish privilege, accuracy or legal compliance.

D6-CTL-07 - AI RESILIENCE & BUSINESS CONTINUITY

Domain: D6 — GOVERNANCE, ACCOUNTABILITY & HUMAN OVERSIGHT

Normative reference: GAISSF-NOR-004 v1.0, control D6-CTL-07; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Ensure legal work can continue when AI services, retrieval stores or providers fail, change or become legally unusable.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Continuity plan, manual fallback, backup and exercise results

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Continuity objectives must reflect court, filing and client deadlines.

D7-CTL-H01 - AI-GENERATED PHISHING SIMULATION

Domain: D7 — HUMAN & SOCIETAL HARMS

Normative reference: GAISSF-NOR-004 v1.0, control D7-CTL-H01; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Test legal personnel against realistic AI-generated impersonation of clients, courts, partners and counterparties without exposing real confidential data.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Approved simulation plan, results and remediation

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Simulation metrics do not directly measure resistance to all attacks.

D7-CTL-H02 - DEEPFAKE DETECTION TRAINING

Domain: D7 — HUMAN & SOCIETAL HARMS

Normative reference: GAISSF-NOR-004 v1.0, control D7-CTL-H02; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Train personnel to challenge synthetic voice, video and documents used in payment, evidence, instruction or identity fraud.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Training records, exercises and escalation procedure

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Human detection is fallible and should be combined with verification.

D7-CTL-H03 - OUT-OF-BAND AUTHENTICATION

Domain: D7 — HUMAN & SOCIETAL HARMS

Normative reference: GAISSF-NOR-004 v1.0, control D7-CTL-H03; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Verify sensitive client instructions, payment changes, settlement authority and evidence transfers through an independent channel.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Verification logs, approved channels and exception review

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: The independent channel must itself be trusted and current.

D7-CTL-H04 - AI SOCIAL ENGINEERING IR

Domain: D7 — HUMAN & SOCIETAL HARMS

Normative reference: GAISSF-NOR-004 v1.0, control D7-CTL-H04; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Include AI-enabled impersonation, synthetic evidence and targeted phishing in legal incident response.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Playbook, exercises and preserved evidence

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Response must account for confidentiality and legal-hold requirements.

D7-CTL-H05 - AI-ENHANCED EXTERNAL ATTACK DEFENSE

Domain: D7 — HUMAN & SOCIETAL HARMS

Normative reference: GAISSF-NOR-004 v1.0, control D7-CTL-H05; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Monitor and defend legal systems against automated reconnaissance, credential attacks and data exfiltration targeting high-value matters.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Detection coverage, threat scenarios and response evidence

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Threat intelligence does not prove control effectiveness.

D8-CTL-01 - EU AI ACT RISK TIER MAPPING

Domain: D8 — REGULATORY ALIGNMENT & COMPLIANCE

Normative reference: GAISSF-NOR-004 v1.0, control D8-CTL-01; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Where relevant, record role, system classification, prohibited-use analysis, obligations and effective dates for each legal AI use.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Applicability assessment, legal review and update log

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: No universal classification is asserted; jurisdiction-specific verification is required.

D8-CTL-02 - ISO 42001 GAP ANALYSIS

Domain: D8 — REGULATORY ALIGNMENT & COMPLIANCE

Normative reference: GAISSF-NOR-004 v1.0, control D8-CTL-02; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Map legal AI governance processes to the applicable edition of ISO/IEC 42001 without claiming equivalence or certification.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Gap analysis, action plan and source edition

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: A gap analysis is not certification.

D8-CTL-03 - GPAI TECHNICAL DOCUMENTATION VERIFICATION

Domain: D8 — REGULATORY ALIGNMENT & COMPLIANCE

Normative reference: GAISSF-NOR-004 v1.0, control D8-CTL-03; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Where the organisation has relevant provider or downstream obligations, verify technical documentation, limitations and change information.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Documentation review and obligation register

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Applicability depends on role and jurisdiction.

D8-CTL-04 - DORA ICT INCIDENT REPORTING (FINANCIAL SECTOR)

Domain: D8 — REGULATORY ALIGNMENT & COMPLIANCE

Normative reference: GAISSF-NOR-004 v1.0, control D8-CTL-04; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: For legal functions within DORA-regulated entities or providers, integrate AI incidents into the entity-specific ICT reporting process.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Applicability rationale, reporting workflow and incident records

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Not generally applicable outside the regulated financial context.

D8-CTL-05 - NIST SP 800-218A COMPLIANCE CHECK

Domain: D8 — REGULATORY ALIGNMENT & COMPLIANCE

Normative reference: GAISSF-NOR-004 v1.0, control D8-CTL-05; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Use the applicable NIST secure-development profile as a technical reference for legal AI development without representing it as a universal legal obligation.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Secure-development assessment and remediation evidence

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Reference use is not government certification or legal compliance.

D9-CTL-01 - PHYSICAL HARM BOUNDARY ENFORCEMENT

Domain: D9 — PHYSICAL AI SAFETY

Normative reference: GAISSF-NOR-004 v1.0, control D9-CTL-01; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: For embodied legal-support systems, define physical zones, prohibited actions and safety boundaries for evidence rooms, archives and secure facilities.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Boundary configuration, tests and approvals

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Usually not applicable to software-only legal AI.

D9-CTL-02 - SAFE STATE AND GRACEFUL DEGRADATION

Domain: D9 — PHYSICAL AI SAFETY

Normative reference: GAISSF-NOR-004 v1.0, control D9-CTL-02; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Ensure embodied or cyber-physical legal systems enter a safe, non-destructive state when uncertain, disconnected or compromised.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Safe-state tests and recovery logs

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Safe state must preserve evidence and access requirements where possible.

D9-CTL-03 - HUMAN OVERRIDE AND EMERGENCY STOP

Domain: D9 — PHYSICAL AI SAFETY

Normative reference: GAISSF-NOR-004 v1.0, control D9-CTL-03; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Provide authorised personnel with tested means to stop physical or high-impact automated legal operations.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Override tests, access list and exercise records

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Override must not permit unauthorised destruction or disclosure.

D9-CTL-04 - CYBER-PHYSICAL ATTACK DETECTION

Domain: D9 — PHYSICAL AI SAFETY

Normative reference: GAISSF-NOR-004 v1.0, control D9-CTL-04; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Detect attacks on sensors, access systems, robots or devices handling evidence and legal records.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Detection tests, alerts and incident records

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Detection scope must reflect the actual physical deployment.

D9-CTL-05 - PHYSICAL ENVIRONMENT INTEGRITY MONITORING

Domain: D9 — PHYSICAL AI SAFETY

Normative reference: GAISSF-NOR-004 v1.0, control D9-CTL-05; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Monitor environmental and access conditions affecting secure evidence, archives and embodied AI operation.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Sensor logs, calibration and exception records

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Monitoring does not replace physical security governance.

D9-CTL-06 - ACTUATOR COMMAND VERIFICATION

Domain: D9 — PHYSICAL AI SAFETY

Normative reference: GAISSF-NOR-004 v1.0, control D9-CTL-06; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Verify that physical commands affecting evidence, records or secure areas are authorised, bounded and attributable.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Command allowlist, approvals and reconciliation

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Model-generated commands must never be self-authorising.

D9-CTL-07 - PHYSICAL INCIDENT EVIDENCE PRESERVATION

Domain: D9 — PHYSICAL AI SAFETY

Normative reference: GAISSF-NOR-004 v1.0, control D9-CTL-07; GAISSF-NOR-001 v1.0 conformance baseline

Legal-sector interpretation: Preserve telemetry, video, command history and physical evidence after an AI-related incident under controlled chain-of-custody procedures.

Minimum implementation: Assign an owner; document scope and applicability; implement the stated safeguards; test design and operation; retain evidence; approve exceptions and residual risk.

Human oversight: A competent named reviewer must approve material legal reliance or external/binding action affected by this control.

Evidence: Preservation package, hashes and custody log

Testing: Test representative legal use cases, adverse inputs, client/matter boundaries, provider changes and failure recovery.

Metrics: Coverage of in-scope systems; test pass rate; exception age; unresolved high-risk findings; change revalidation completion.

Common failure: Reliance on policy statements or provider claims without configuration evidence, testing, matter-level scope and accountable review.

Limitations: Preservation methods must be adapted to applicable evidentiary rules.

7. Provider Due Diligence Checklist

- PDD-01 - Does the provider use prompts, files, outputs or metadata for its own training, evaluation or product improvement? Evidence: Contract, settings and provider statement.
- PDD-02 - Can retention be configured and independently verified, including backups and logs? Evidence: Retention schedule, deletion evidence.
- PDD-03 - How are tenants, clients, matters, indexes and encryption keys segregated? Evidence: Architecture and test evidence.
- PDD-04 - Which subprocessors and hosting locations are used, and how are changes notified? Evidence: Subprocessor register and contract.
- PDD-05 - What provider personnel can access customer content, under what controls and logging? Evidence: Access model and audit evidence.
- PDD-06 - How are vulnerabilities, patches, model artefacts and dependencies managed? Evidence: SDLC, SCA and vulnerability evidence.
- PDD-07 - How are customers notified of model, feature, API, safety-policy or deprecation changes? Evidence: Change-notice terms and history.
- PDD-08 - What incident-notification, investigation-support and evidence-preservation commitments apply? Evidence: Contract and playbook.
- PDD-09 - Does the provider support legal holds, preservation, data export and verified deletion? Evidence: Functional test and terms.
- PDD-10 - What assurance reports exist, what is their scope, period and exclusions? Evidence: SOC/ISO or equivalent evidence.
- PDD-11 - Are output ownership, confidentiality, indemnity and liability terms acceptable for the use case? Evidence: Legal review.
- PDD-12 - Can tool calls, agent permissions and external actions be restricted and logged? Evidence: Configuration and test evidence.
- PDD-13 - What controls address prompt injection, data exfiltration and cross-context leakage? Evidence: Security test evidence.
- PDD-14 - What exit, portability, escrow or continuity arrangements apply? Evidence: Exit plan and test.

8. Disclosure Requirement Register

Maintain a controlled register of court, tribunal and regulator AI-use disclosure or certification requirements. Record jurisdiction, forum, matter, authoritative source, effective date, required action, reviewer, evidence and last verification. SEC-048 does not itself determine whether disclosure is legally required.

9. 90-Day Implementation Sequence

Days 0-30: Inventory systems and uses; stop unauthorised protected-data use; classify risk; establish owners and review rules.

Days 31-60: Implement matter isolation, provider due diligence, citation verification, agent restrictions, legal-hold integration, training and evidence capture.

Days 61-90: Test design and operation; sample outputs; verify isolation; exercise incident response; assess residual risk and obtain accountable approval.

10. Publication Readiness

RELEASE BLOCKED. External legal/professional review, named approvals, final licence and trademark wording, approved publication metadata and final accessibility/publication QA remain open.