ONEQUILL / ENGINEERING REFERENCE

AI-native governance framework

Laws, roles and engineering controls for enterprise AI, retrieval and autonomous agents.

7 control areas / 21 requirements / 5 implementation phases

Published and reviewed: 16 October 2026
OneQuill Research · Editorial adaptation with source attribution inside
Canonical page and interactive requirements explorer

Start with the system you are governing

AI-native governance connects the intended use of a system to accountable people, proportionate controls and evidence that those controls work. It covers foundation models, retrieval-augmented generation (RAG), fine-tuning, third-party APIs and agents that act through tools. A trustworthy model alone does not establish a trustworthy application: retrieval permissions, downstream actions, supplier updates and human decisions also shape the outcome.

Use this framework to create a deployment-specific review. First record where the system operates, who it affects, which data it uses and what it is allowed to do. Then establish the applicable law and contractual commitments, select safeguards, assign owners and retain evidence. Revisit the assessment when the purpose, model, data, geography or level of autonomy changes.

The requirements explorer below is an engineering interpretation organised into seven control areas. Its suggested owners are a starting point for a local responsibility matrix. A control is not automatically a legal duty in every deployment, and implementing this reference does not establish certification or compliance.

Editorial provenance. OneQuill adapted this reference from the supplied manuscript "AI Governance Frameworks Overview: Laws, Standards, Roles, and Controls Behind Responsible Enterprise AI", credited in that document to Dr. Arpit Yadav (GenAI Expert, Aigenthix) / Aigenthix Technologies. OneQuill is responsible for this adaptation, structure and qualifications; the attribution does not imply an endorsement or partnership. Selected official sources were checked on 16 October 2026. Consult the controlling legal text and current guidance when making an applicability decision.

Laws and regulatory scope

European Union: classify the use and the operator

The EU AI Act distinguishes prohibited practices, high-risk systems, transparency obligations and general-purpose AI model obligations. Start with the intended purpose and the relevant actor. Article 5 addresses prohibited practices; Articles 9-15 address requirements for high-risk systems, including risk management, data governance, documentation, records, transparency, oversight and technical performance. Article 50 addresses defined transparency duties. The obligations and exceptions vary by use and actor.

High-risk classification can arise through regulated product categories in Annex I or uses listed in Annex III. A gateway, model API or private deployment does not receive a blanket classification merely because it uses AI. Document the actual

system boundary and decision context.

Article 25 change review. Rebranding an existing high-risk system, making a substantial modification, or changing the intended purpose so that a system becomes high-risk can bring provider responsibilities to a downstream operator. Assess the specific statutory conditions. Retain the original system documentation, modification record and contractual allocation of cooperation duties. Avoid treating every fine-tune or RAG integration as automatic reclassification.

Timing and transparency. Application dates and transition rules differ across provisions. The European Commission's current implementation guidance should be checked alongside the legal text before scheduling a release. For generated content, determine the marking or disclosure obligation that applies; do not assume a single universal cryptographic-watermark requirement.

Other jurisdictions and international instruments

Instrument	Type and scope	What to establish in a deployment review
South Korea AI Basic Act	National law with obligations for relevant high-impact and generative AI uses	Determine whether the business, service and users are in scope; establish applicable transparency, safety and documentation duties. The Act took effect on 22 January 2026. Consult the official law and government implementation notice.
Japan AI governance	AI promotion legislation alongside business guidance	Distinguish the statute from voluntary implementation guidance. Use the developer, provider and business-user perspectives in METI/MIC AI Guidelines for Business, version 1.2 to organise responsibilities.
China generative AI service measures	Service-specific regulatory rules	Assess whether the service is offered to the public in China, which content and data rules apply, and whether a security assessment or algorithm filing is required for the specific service. Begin with the CAC's published measures.
Council of Europe AI Framework Convention	International treaty framework on human rights, democracy and the rule of law	Check ratification, entry into force and domestic implementation for the relevant party. Signature alone does not establish every domestic duty. Consult the Council of Europe convention resource.

Privacy, sector rules and resilience

AI governance sits alongside existing rules. Record the relevant sector and territory rather than applying a single global compliance checklist.

Context	Reference points	Questions for the accountable team
Personal data	GDPR, California CCPA/CPRA, India DPDPA and Brazil LGPD	What is the lawful basis? Are notice, access, deletion or automated-decision rights engaged? Does processing require a DPIA? How do retention and deletion reach vectors, caches and backups? Check local commencement and implementing rules.
Financial services	Model-risk supervision, EBA expectations and OSFI E-23	Which regulated entity and model are in scope? Who independently validates assumptions, performance and limitations? Confirm the current version and effective date of supervisory guidance.
Employment	NYC Local Law 144 / DCWP and applicable employment/discrimination law	Is the tool a covered automated employment decision tool? Establish the independent audit, public summary and notice process. Assess discrimination separately; an audit metric alone does not settle legality.

Context	Reference points	Questions for the accountable team
Health and medical devices	EU MDR/IVDR and FDA AI-enabled device guidance	Does the intended medical use make this a regulated device? What validation, change control and clinical monitoring are required? A general AI assistant is not automatically a medical device.
Consumer-facing AI	Consumer protection, unfair commercial practices and FTC AI guidance	Can performance and safety claims be substantiated? Are users misled about automation, limitations, pricing or outcomes?
Operational resilience	NIS2, DORA and the Cyber Resilience Act where applicable	Which entity or product rules apply? Map reporting triggers, recipients and deadlines separately. Do not apply one incident deadline across every regime.

For a broader geographical index, continue with AI governance rules for gateways and local inference.

Standards and responsible-AI frameworks

Use standards and voluntary frameworks to organise repeatable work. Record the edition used and the scope of any certification claim. A management-system certification does not certify the legal compliance or safety of every model output.

Reference	Purpose	Practical application
ISO/IEC 42001:2023	AI management system requirements	Establish policy, assigned responsibilities, risk treatment, supplier processes, internal review and improvement. Certification concerns a defined organisational scope.
ISO/IEC 23894:2023	AI risk-management guidance	Identify risks, assess consequences and likelihood, choose treatments and monitor residual risk across the lifecycle.
ISO/IEC 42005:2025	AI system impact-assessment guidance	Evaluate effects on people and society and integrate the assessment with risk and management-system processes.
ISO/IEC 22989:2022	AI concepts and terminology	Use consistent definitions across procurement, engineering, assurance and legal records.
NIST AI RMF	Voluntary risk-management framework	Organise work through Govern, Map, Measure and Manage, with responsibilities and evidence attached to each activity.
NIST AI 600-1	Generative AI profile	Adapt risk assessment to generative systems, including confabulation, privacy, information integrity, security and human interaction.
OECD AI Principles	Intergovernmental principles	Connect organisational policies to human rights, transparency, robustness and accountability.
UNESCO Recommendation on the Ethics of AI	Global ethical-policy guidance	Include social impacts, inclusion, environmental considerations and meaningful oversight in policy and assessment.
G7 Hiroshima AI Process	International policy framework and voluntary code	Review advanced-model development and use against the relevant commitments and disclosure expectations.

Security frameworks and threat coverage

Model behaviour is one part of the threat surface. Evaluate untrusted data, tool boundaries, credentials, memory, code execution and downstream systems. Record the edition and date of every taxonomy used; identifier mappings below describe the supplied manuscript's **OWASP LLM Top 10 2025** baseline. Check the current OWASP release when planning a new assessment.

OWASP 2025 baseline	Failure to test	Engineering response
LLM01 · Prompt injection	A retrieved page or message redirects the assistant	Segregate untrusted content, enforce permissions outside the model and test indirect injection.
LLM02 · Sensitive information disclosure	Outputs reveal private records or secrets	Minimise context, enforce retrieval access and test output screening.
LLM03 · Supply chain	A model, adapter or dependency is compromised	Verify provenance, pin versions and keep an AI/software bill of materials.
LLM04 · Data and model poisoning	Modified training or retrieval data changes behaviour	Protect write access, track lineage and compare with trusted baselines.
LLM05 · Improper output handling	Model text becomes executable SQL, HTML or commands	Validate and encode for the destination; use parameterised operations.
LLM06 · Excessive agency	Broad permissions enable unintended actions	Reduce available tools, scope credentials and require approval at consequential boundaries.
LLM07 · System prompt leakage	Hidden instructions or embedded secrets are exposed	Keep secrets out of prompts and enforce business rules outside model context.
LLM08 · Vector and embedding weaknesses	Retrieval leaks across tenants or permissions	Apply access checks at retrieval time and test cross-tenant isolation.
LLM09 · Misinformation	Fabricated claims are accepted as authoritative	Verify evidence, use task-specific evaluations and offer escalation.
LLM10 · Unbounded consumption	Recursive calls exhaust tokens or compute	Limit iterations, spend, concurrency and execution time; provide cancellation.

The OWASP Agentic Top 10 2026 extends review to goal hijacking (ASI01), tool misuse (ASI02), identity and privilege abuse (ASI03), agentic supply chain (ASI04), unexpected code execution (ASI05), memory/context poisoning (ASI06), insecure inter-agent communication (ASI07), cascading failures (ASI08), human-agent trust exploitation (ASI09) and rogue agents (ASI10). Use these as scenario prompts, then verify the actual permissions and consequences of the application.

MITRE ATLAS provides an evolving knowledge base of adversary tactics and techniques against AI systems. Map realistic attack paths to relevant techniques rather than relying on a fixed count of tactics or treating taxonomy coverage as proof of security.

Organisational roles and accountability

An AI governance board or equivalent executive forum sets risk tolerance, approves exceptional uses and receives material incident and assurance reports. A chief AI officer or designated governance lead coordinates the programme. The model owner carries day-to-day accountability for a named application. Security owns defensive architecture; legal and privacy specialists determine applicable duties and data-protection requirements. Internal audit independently assesses whether the programme and selected controls operate as intended.

The DPO's advisory and monitoring role must remain independent where that role is legally required. Business management retains responsibility for processing decisions. Organisations can combine operational roles when appropriate, but should preserve the independence required for assurance and statutory oversight.

Example RACI for the AI lifecycle

This allocation is illustrative and must be agreed locally. A = accountable decision owner; R = performs the work; C = consulted; I = informed. Titles do not determine legal actor status.

Activity	Governance lead	Model owner	Security	Privacy / legal	Internal audit
Use-case registration	C	A/R	C	C	I
Classification and risk acceptance	A	R	C	R	I
Data sourcing and architecture	I	A/R	R	C	I
Threat modelling and hardening	I	R	A/R	C	I
Evals and release recommendation	C	A	R	C	I
Applicable conformity/sign-off process	A	R	C	R	I
Production monitoring and response	I	A/R	R	C	I
Independent programme assurance	I	C	C	C	A/R

Board reporting should connect metrics to decisions: unregistered AI uses, unresolved material findings, evaluation failures, harmful outcomes, supplier changes, incident impact and time to contain. Include the denominator, period, severity and owner; raw prompt-injection counts or aggregate sentiment scores can be misleading without context.

Requirements explorer

GOV-01: Register every AI use case

Topic: Inventory. Phase: Scope. Suggested owner: Model owner.

Record the business purpose, affected people, decision context, deployment locations, accountable owner and permitted uses. Reconcile the register with procurement, API usage and discovered shadow AI.

Evidence to keep: Approved use-case record with owner, scope, risk tier, review date and retirement criteria.

NIST AI Risk Management Framework

GOV-02: Maintain the model and supplier catalogue

Topic: Inventory. Phase: Design. Suggested owner: Model owner.

Track base models, fine-tuning adapters, hosting providers, retrieval stores, tool integrations and versions. Record licences, update policies, data processing terms and supplier dependencies.

Evidence to keep: Versioned model cards, supplier assessments, contracts and dependency inventory.

ISO/IEC 42001 · AI management · OWASP LLM application security

GOV-03: Create an AI bill of materials

Topic: Inventory. Phase: Design. Suggested owner: Security.

Use a machine-readable inventory such as CycloneDX to connect models, datasets and software components. Record provenance and trusted artifact hashes; a hash detects a change against a trusted baseline but does not establish that a dataset is safe.

Evidence to keep: AIBOM linked to release identifiers, artifact verification and a vulnerability response owner.

CycloneDX machine learning bill of materials

GOV-04: Classify the deployment and legal role

Topic: Risk assessment. Phase: Scope. Suggested owner: Legal & privacy.

Assess jurisdiction, intended purpose, affected people and provider/deployer responsibilities. Review EU AI Act Article 25 when rebranding, substantially modifying or changing the purpose of a system; ordinary integration is not automatically provider reclassification.

Evidence to keep: Dated applicability memo, actor mapping, risk rationale and change-review triggers.

EU AI Act · official text

GOV-05: Assess impacts before committing to architecture

Topic: Risk assessment. Phase: Design. Suggested owner: Legal & privacy.

Evaluate foreseeable impacts on people, safety, rights and the environment. Coordinate an AI impact assessment with a DPIA or fundamental-rights assessment where the legal conditions apply. ISO/IEC 42005 is guidance, not a universal statutory mandate.

Evidence to keep: Impact assessment, stakeholder input, mitigations and a residual-risk decision.

ISO/IEC 42005 · impact assessment · GDPR · official text

GOV-06: Set acceptance criteria and residual-risk authority

Topic: Risk assessment. Phase: Scope. Suggested owner: Governance lead.

Define unacceptable uses, measurable quality and safety targets, exception handling and who can approve remaining risk. Reassess when the model, data, autonomy, geography or intended use changes.

Evidence to keep: Approved risk policy, release thresholds, exception expiry dates and review history.

NIST AI Risk Management Framework · ISO/IEC 42001 · AI management

GOV-07: Run task-specific evaluations

Topic: Model assurance. Phase: Verify. Suggested owner: Model owner.

Use representative, versioned test sets to measure task accuracy, groundedness, refusal behaviour and performance for relevant user groups. Separate evaluation data from training material and repeat affected evaluations after meaningful changes.

Evidence to keep: Evaluation dataset provenance, reproducible results, failures and signed release decision.

NIST Generative AI Profile

GOV-08: Test adversarial paths across the whole system

Topic: Model assurance. Phase: Verify. Suggested owner: Security.

Exercise direct and indirect prompt injection, poisoned retrieval, cross-tenant access, unsafe output handling and tool misuse. Map scenarios to the selected OWASP edition and MITRE ATLAS; test actual permissions and downstream effects.

Evidence to keep: Threat model, red-team scenarios, remediation records and retest results.

OWASP LLM application security · OWASP Agentic Top 10 · 2026 · MITRE ATLAS

GOV-09: Evaluate fairness and employment-tool obligations

Topic: Model assurance. Phase: Verify. Suggested owner: Legal & privacy.

Assess relevant groups, error differences and decision impacts. For covered NYC automated employment decision tools, establish independent bias-audit, publication and notice requirements. An impact ratio is a diagnostic metric; a 0.8 threshold alone is not a universal legal pass/fail test.

Evidence to keep: Audit scope, selection/scoring rates, impact ratios, limitations and required notices.

NYC DCWP · automated employment decision tools

GOV-10: Place human approval before consequential actions

Topic: Human oversight. Phase: Design. Suggested owner: Model owner.

Define actions that need human-in-the-loop approval. Show the proposed action, affected records, material evidence, uncertainty and change preview. Give reviewers time and authority to refuse; do not substitute hidden model reasoning for verifiable evidence.

Evidence to keep: Approval policy, reviewer training, permission checks and recorded decisions.

EU AI Act · official text · OWASP Agentic Top 10 · 2026

GOV-11: Make intervention and shutdown effective

Topic: Human oversight. Phase: Operate. Suggested owner: Security.

Provide human-on-the-loop supervision, revocable credentials, circuit breakers and an operational stop mechanism. Test cancellation of in-flight tool actions, queued jobs and background tasks as well as the chat interface.

Evidence to keep: Shutdown drill, escalation contacts, recovery procedure and bounded stop latency.

OWASP Agentic Top 10 · 2026

GOV-12: Offer contestability and specialist escalation

Topic: Human oversight. Phase: Operate. Suggested owner: Governance lead.

Make it possible to challenge significant automated outcomes, request human review where applicable and route difficult cases to qualified people. Track whether overrides actually change the downstream decision.

Evidence to keep: Accessible review channel, service targets, appeal records and resolution analysis.

GDPR · official text · Council of Europe · AI Framework Convention

GOV-13: Give agents distinct identities and minimal authority

Topic: Agent security. Phase: Design. Suggested owner: Security.

Bind each agent to a scoped machine identity. Use short-lived credentials, explicit tenant boundaries and the least tools and data required. Avoid shared privileged credentials and verify authorisation again at the destination API.

Evidence to keep: Identity inventory, tool permission matrix, expiry policy and access-denial tests.

OWASP Agentic Top 10 · 2026

GOV-14: Enforce policy before tool execution

Topic: Agent security. Phase: Verify. Suggested owner: Security.

Treat proposed tool calls and arguments as untrusted. Validate schemas, authorisation, destinations and transaction limits in deterministic enforcement code. Add per-task budgets, rate limits, loop limits and explicit approval for consequential mutations.

Evidence to keep: Policy enforcement rules and tests for blocked destinations, invalid arguments and budget exhaustion.

OWASP LLM application security · OWASP Agentic Top 10 · 2026

GOV-15: Isolate execution and inter-agent communication

Topic: Agent security. Phase: Design. Suggested owner: Security.

Run generated code and untrusted tools in isolated environments with filesystem, network and resource limits. Verify MCP servers and plugins, authenticate agent messages and validate message schemas before acting on delegated work.

Evidence to keep: Sandbox configuration, egress rules, trusted-tool provenance and communication integrity tests.

OWASP Agentic Top 10 · 2026 · MITRE ATLAS

GOV-16: Trace data provenance and retrieval permissions

Topic: Data governance. Phase: Design. Suggested owner: Legal & privacy.

Record origin, transformations, permissions, licensing and retention for training, fine-tuning, retrieval and inference data. Enforce tenant and document permissions at retrieval time; embedding namespaces alone do not prove isolation.

Evidence to keep: Data-flow map, lineage register, access tests and documented legal basis where required.

GDPR · official text · OWASP LLM application security

GOV-17: Reduce sensitive data at ingress and egress

Topic: Data governance. Phase: Verify. Suggested owner: Security.

Minimise data before external model calls. Apply contextual detection and redaction to prompts, retrieved passages and outputs, with tests for missed secrets and false positives. Treat DLP as one layer alongside access control and endpoint policy.

Evidence to keep: Redaction rules, leakage tests, approved routing and retention settings.

OWASP LLM application security · NIST Generative AI Profile

GOV-18: Design retention, deletion and privacy measures

Topic: Data governance. Phase: Operate. Suggested owner: Legal & privacy.

Propagate approved retention and deletion policies across source records, vectors, caches, logs and backups. Assess differential privacy and model unlearning for the specific use case; neither is a universal requirement or a guaranteed substitute for retraining.

Evidence to keep: Retention schedule, deletion workflow, exception record and verified deletion outcomes.

GDPR · official text · NIST Generative AI Profile

GOV-19: Keep useful, protected audit evidence

Topic: Monitoring & response. Phase: Operate. Suggested owner: Security.

Record model and policy versions, access decisions, tool actions, approvals and outcomes with integrity protection and restricted access. Minimise sensitive payloads and define retention. Logging every raw prompt and response can create additional privacy and security risk.

Evidence to keep: Event schema, integrity verification, access audit and retention controls.

EU AI Act · official text · GDPR · official text

GOV-20: Monitor quality, drift, cost and operational limits

Topic: Monitoring & response. Phase: Operate. Suggested owner: Model owner.

Track task failure rates, relevant fairness measures, unsafe actions, latency, token use and spend. Establish baselines and actionable alerts; embedding distance or sentiment changes alone do not demonstrate poisoning or a breach.

Evidence to keep: Monitoring dashboard, thresholds, alert owners and periodic governance reports.

NIST Generative AI Profile

GOV-21: Rehearse incident response and independent assurance

Topic: Monitoring & response. Phase: Operate. Suggested owner: Internal audit.

Prepare runbooks for injection, memory poisoning, data disclosure and runaway tool loops. Contain access, preserve proportionate evidence, assess affected people and identify applicable notification rules with legal counsel. Independently review whether controls work and close tracked findings.

Evidence to keep: Tabletop exercise, incident register, notification decision log and independent review report.

OWASP Agentic Top 10 · 2026 · GDPR · official text · EU AI Act · official text

Five-phase implementation blueprint

Treat each phase as a review with an owner, evidence and a decision. Re-enter the appropriate phase after a material change. Gate criteria should reflect the risk of the actual use case.

Phase	Engineering deliverables	Governance decision and evidence	Suggested lead
1 · Inception and scoping	Purpose, users, system boundary, initial model and tool choices	Register the use case; assess role and risk; define prohibited uses and acceptance criteria. GOV-01, GOV-04, GOV-06.	Model owner with governance and legal
2 · Data and architecture	Data flows, retrieval permissions, dependencies, supplier catalogue and AIBOM	Review impacts, data permissions, human escalation and isolation. GOV-02, GOV-03, GOV-05, GOV-10, GOV-13, GOV-15, GOV-16.	Engineering with security and privacy
3 · Model and agent hardening	Tool policy, scoped identities, DLP, output validation and sandbox limits	Demonstrate that unauthorised actions and sensitive data flows are blocked. GOV-14, GOV-17.	Security and backend engineering
4 · Verification and assurance	Task evaluations, adversarial scenarios, fairness analysis and approval workflow	Review failures, required audits, residual risk and release authority. GOV-07, GOV-08, GOV-09.	Model owner with independent reviewers
5 · Production and continuous review	Monitoring, protected logs, deletion workflow, stop controls and incident runbooks	Test intervention, review operational risk, resolve incidents and obtain independent assurance. GOV-11, GOV-12, GOV-18-GOV-21.	Operations, security and internal audit

Evidence record template

For each selected requirement, record: system and version; applicable source and edition; accountable owner; implementation/configuration; evidence location; test or review date; result and limitations; residual risk; approver; and next review trigger. Record a justified “not applicable” decision rather than silently omitting a requirement.

Source notes and careful interpretation

This adaptation keeps the supplied manuscript's legal, standards, organisational and engineering structure while qualifying broad claims that could mislead an implementation team:

- Legal duties depend on jurisdiction, actor, use and commencement. Treaty signature, guidance publication and certification are different events.
- The EU AI Act's prohibited practices are addressed in Article 5. A provider change under Article 25 requires assessing the statutory conditions; it is not an automatic consequence of using an upstream model.
- An employment audit's impact ratio is not a universal compliance verdict. Review the applicable rule and the evidence behind the audit.
- A dataset hash establishes integrity against a baseline, not freedom from poisoning. Differential privacy and unlearning require their own suitability and effectiveness assessment.
- Security controls and protected audit trails should be proportionate. Collecting every raw prompt, retrieved record and response can conflict with minimisation and increase exposure.

- Incident notification is regime-specific. GDPR Article 33 uses a conditional 72-hour supervisory-notification rule; EU AI Act and NIS2 provisions have their own triggers and schedules.
- Security taxonomies evolve. This page identifies the manuscript's OWASP 2025 LLM baseline and separately links the current release and Agentic 2026 framework.

Primary sources are linked beside the relevant discussion and each explorer requirement. This page is an educational engineering reference. The responsible organisation should document its own legal applicability, risk decisions and implementation evidence.

Downloads and related reading

- Download the complete framework as a PDF.
- Read or download the full Markdown reference.
- Download the structured requirements as JSON.
- Global AI governance rules.
- AI gateway security guide.
- AI gateway privacy guide.