AGCS

AGENT GOVERNANCE CERTIFICATION STANDARD

Autonomous Agent Governance, Audit, and Accountability

A Certifiable Technical Standard for Enterprise Autonomous Agent Deployment

VERSION 0.9 DRAFT FOR PUBLIC COMMENT · AGENT GOVERNANCE CERTIFICATION COUNCIL · 2025

This document is a pre-publication draft circulated for expert review and comment. It does not represent a final published standard.

Preamble

This document sets out the Agent Governance Certification Standard (AGCS), Version 0.9, issued as a public comment draft by the Agent Governance Certification Council (the Council). The Council invites technical, legal, and operational comment from enterprises, vendors, academic institutions, and regulatory bodies prior to publication of the final Version 1.0 standard.

AGCS is a certifiable technical standard defining the minimum controls required for the responsible deployment of autonomous AI agents in enterprise environments. It specifies requirements across three progressively rigorous certification tiers, addressing policy governance, audit trail integrity, cryptographic non-repudiation, and hardware-attested supervisor independence.

AGCS is designed to be technology-neutral and vendor-neutral. No specific product, platform, or vendor is required to satisfy any control in this standard. The standard specifies what must be achieved, not how it must be implemented. Assessors are required to evaluate evidence of control satisfaction independently of the technology used to achieve it.

DRAFT STATUS

Version 0.9 is a pre-publication draft. Control identifiers, tier boundaries, and normative language are subject to revision based on public comment. Organisations implementing controls against this draft should monitor the Council registry for Version 1.0 publication.

1 Scope and Applicability

1.1 Subject Matter

This standard applies to the deployment of autonomous AI agents—software systems capable of taking actions in the real world without continuous human instruction, including but not limited to: executing API calls, sending communications, accessing or modifying data records, initiating financial transactions, and interacting with external services on behalf of an organisation or its customers.

The standard does not apply to: AI systems that operate purely in read-only or analytical modes with no capacity for external action; human-supervised copilot systems where every action requires explicit individual human approval; or AI systems operating exclusively within an isolated sandbox environment with no external connectivity.

1.2 Applicable Organisations

AGCS certification is applicable to any organisation that deploys autonomous AI agents in a production environment where those agents take actions affecting external systems, third parties, regulated data, or financial obligations. This includes but is not limited to:

- Financial services organisations deploying agents for customer communication, transaction processing, CRM management, or compliance reporting.

- Healthcare organisations deploying agents for administrative workflows, patient record management, or clinical decision support.
- Legal and professional services organisations deploying agents for document processing, client communication, or regulatory filing.
- Technology companies deploying agents in customer-facing workflows or on behalf of regulated clients.
- Any organisation subject to the EU AI Act Articles 9–15 (high-risk AI system requirements) where the AI system meets the autonomous agent definition in Section 1.1.

1.3 Relationship to Existing Standards

AGCS is designed to complement, not replace, existing information security and AI governance standards. The following relationships apply:

Standard	Relationship to AGCS
ISO 27001	AGCS Tier 1 controls are compatible with and extend ISO 27001 Annex A controls for AI-specific contexts. ISO 27001 certification does not satisfy AGCS requirements.
SOC 2 Type II	AGCS audit trail and integrity controls extend SOC 2 Availability, Integrity, and Confidentiality criteria for agentic AI contexts. SOC 2 certification does not satisfy AGCS requirements.
ISO 42001	AGCS provides the technical implementation specification for AI management system requirements under ISO 42001. The two standards are complementary.
NIST AI RMF	AGCS Tier 2–3 controls operationalise the GOVERN, MAP, MEASURE, and MANAGE functions of the NIST AI Risk Management Framework with specific technical requirements.
EU AI Act	AGCS Tier 2–3 certification is designed to provide evidence of conformity with EU AI Act Article 9 (risk management), Article 12 (record-keeping), and Article 14 (human oversight) obligations for high-risk AI systems.
GDPR	AGCS Tier 2 Control AG-2.4 (GDPR/retention reconciliation) directly addresses the conflict between Article 17 right-to-erasure obligations and immutable audit trail requirements.

2 Definitions

The following definitions apply throughout this standard. Where a term is used in a regulatory instrument referenced in Section 1.3, the definition in this standard is intended to be consistent with but more specific than the regulatory definition.

Term	Definition
Autonomous AI Agent	A software system that, given a goal or task, autonomously selects and executes sequences of actions using available tools, without requiring human approval for each individual action. Distinguished from AI assistants by its capacity for unsupervised multi-step execution.
Agent Action	Any discrete interaction between an autonomous AI agent and an external system, including but not limited to: API calls, database reads/writes, email or messaging transmissions, file operations, and financial transaction initiations.
Agent Supervisor	An infrastructure component that sits in the execution path of agent actions and is capable of evaluating, permitting, blocking, or logging those actions independently of the agent and its host platform.
Policy Bundle	A versioned, cryptographically hashable collection of machine-evaluable rules defining the authorised operating parameters of an agent. In implementations using Open Policy Agent, the policy bundle is the OPA bundle artifact.
Bundle Revision Hash	The cryptographic hash (SHA-256 or stronger) of a specific policy bundle version, used to link audit records to the exact policy that governed them.
Audit Record	A structured record of an agent action, including: action identifier, agent identifier, action type, parameters, timestamp, policy verdict, bundle revision hash, and cryptographic link to the preceding record in the audit chain.
Hash Chain	A sequence of audit records where each record contains the cryptographic hash of the preceding record, such that modification of any record invalidates all subsequent records.
Merkle Tree	A tree data structure in which every leaf node is the hash of an audit record and every non-leaf node is the hash of its children, enabling efficient cryptographic proof of record inclusion and log consistency.
Merkle Root	The root hash of a Merkle tree constructed over a set of audit records, representing a single cryptographic commitment to all records in that set.
External Anchor	A Merkle root committed to a public, permissionless blockchain, providing a timestamped cryptographic commitment to the audit record set that is independent of the certifying organisation.
Crypto-shredding	A technique for satisfying data erasure obligations against an immutable log by encrypting sensitive data fields with a per-data-subject key and destroying the key rather than the ciphertext. The resulting ciphertext is considered computationally unrecoverable.
Hardware Security Module (HSM)	A dedicated hardware device or cloud service providing tamper-resistant cryptographic key management. Used in this standard for crypto-shredding key operations and audit record signing.

Term	Definition
Trusted Execution Environment (TEE)	A hardware-isolated execution environment providing confidentiality and integrity guarantees for code and data running within it, even against a privileged attacker on the host system. Examples include AWS Nitro Enclaves, Intel TDX, and AMD SEV-SNP.
TPM Remote Attestation	A process by which a Trusted Platform Module (TPM) chip produces a cryptographically signed statement of the software currently running on a host, enabling a remote party to verify the integrity of that software without physical access.
Policy Replay	The process of reproducing a historical policy decision by evaluating the original input document against the policy bundle version identified by the bundle revision hash recorded in the audit record. Used for independent verification of disputed decisions.
Certifying Organisation	An organisation that has obtained AGCS certification at one or more tiers.
Accredited Assessor	A third-party organisation accredited by the Council to conduct AGCS certification assessments.

3 Certification Tiers

AGCS defines three certification tiers representing progressively rigorous levels of agent governance assurance. Each tier is independently certifiable. Higher tiers include all requirements of lower tiers. An organisation may certify at any tier independently of other tiers, and may certify different agent deployments at different tiers.

TIER 1
Supervised Agent Baseline

Foundational controls establishing non-bypassable supervision, complete audit logging, and human escalation for high-risk actions. Achievable by most organisations within 60–90 days of initiating a governance programme. Required for all subsequent tier certifications.

TIER 2
Cryptographic Accountability

Adds cryptographic integrity controls ensuring audit records are independently verifiable and policy decisions are mathematically linked to and reproducible from the governing policy version. Required for regulated industries where third-party audit is a compliance obligation.

TIER 3
Forensic-Grade Non-Repudiation

The highest assurance level, providing hardware-rooted integrity guarantees, external cryptographic anchoring, and full GDPR erasure certification. Required for deployments where agent actions constitute legally significant events subject to regulatory examination or litigation.

Control Area	Tier 1	Tier 2	Tier 3
Policy definition & versioning	Required	Required	Required
Audit completeness	Required	Required	Required
Audit structural independence	Required	Required	Required
Human escalation workflows	Required	Required	Required
Data classification	Required	Required	Required
Incident response	Required	Required	Required
Hash-chained audit integrity	—	Required	Required
Policy-action cryptographic linkage	—	Required	Required
Decision replayability	—	Required	Required
GDPR/retention reconciliation	—	Required	Required
Supervisor independence attestation	—	Required	Required
External blockchain anchoring	—	—	Required
HSM-backed key operations	—	—	Required
Hardware-attested supervisor integrity	—	—	Required
Erasure certificates	—	—	Required
Cross-border jurisdiction mapping	—	—	Required
Adversarial resilience testing	—	—	Required

4 Tier 1 Controls — Supervised Agent Baseline

All controls in this section are REQUIRED for Tier 1 certification. Controls are expressed using the following normative language: MUST indicates an absolute requirement; MUST NOT indicates an absolute prohibition; SHOULD indicates a strong recommendation; MAY indicates an option.

4.1 Policy Definition and Versioning

AG-1.1 Machine-Evaluable Policy

AG-1.1 **Machine-evaluable policy format.** The operating parameters of each deployed agent MUST be defined in a machine-evaluable, structured format that is capable of producing a deterministic permit or deny decision for any given agent action without human interpretation at evaluation time. Natural language system prompts alone do not satisfy this requirement.

AG-1.2 **Version control.** All policy definitions MUST be maintained in a version control system that records authorship, timestamps, and change history for every modification. Policy MUST NOT be modified in production without creating a new, attributable version record.

AG-1.3 **Separation of duties.** Policy authoring and policy activation in production MUST be performed by different individuals or through a documented approval workflow. A single individual MUST NOT have unilateral authority to both author and activate policy changes.

AG-1.4 **Policy scope documentation.** Each policy definition MUST include a documented scope statement identifying: the agent or agent class it governs; the data categories the agent is authorised to access; the tools the agent is authorised to invoke; and the human escalation triggers that apply.

4.2 Audit Completeness

AG-1.5 Universal action logging

AG-1.5 **Universal action logging.** Every agent action, as defined in Section 2, MUST be logged before or at the time of execution. The logging mechanism MUST be non-bypassable by the agent process. Logging MUST NOT be dependent on the agent's own instrumentation or cooperation.

AG-1.6 **Mandatory log fields.** Each audit record MUST contain at minimum: a unique action identifier; the agent identifier and version; the action type and destination system; the action parameters (subject to data classification controls in AG-1.11); the UTC timestamp of the action; the policy verdict (permit/deny/escalate) and reason; and the identity of the policy version that produced the verdict.

AG-1.7 **Log gap detection.** The audit system MUST provide a mechanism to detect and alert on gaps in the action log. Any period during which the supervision infrastructure was unavailable and agent actions may have proceeded unsupervised MUST be recorded as a supervision gap event in the audit log.

AG-1.8 **Log retention.** Audit records MUST be retained for a minimum of 12 months in immediately accessible storage and 36 months in total. Organisations subject to longer regulatory retention requirements (e.g., MiFID II 7-year requirement) MUST comply with the longer period.

4.3 Audit Structural Independence

AG-1.9 Separate authority

AG-1.9 **Separate authority.** The audit logging system MUST be operated by an entity that is not the same legal entity as the agent platform vendor, or MUST be architecturally isolated such that the agent platform operator cannot modify, delete, or selectively withhold audit records without this action being detectable.

AG-1.10 **Write-once audit store.** The audit store MUST be configured such that existing records cannot be modified or deleted through normal operational access. Append-only configuration MUST be documented and independently verifiable.

4.4 Human Escalation

AG-1.11 Escalation trigger definition

AG-1.11 **Escalation trigger definition.** Organisations MUST formally define and document the categories of agent action that require human approval before execution. Trigger definitions MUST be machine-evaluable and version-controlled under AG-1.2.

AG-1.12 **Escalation path documentation.** For each escalation trigger category, organisations MUST document: the approval authority; the maximum time permitted for approval before the action is rejected; and the action taken when approval is not received within the specified time.

AG-1.13 **Escalation audit trail.** Human approval and rejection decisions MUST be recorded in the audit log with the approver identity, timestamp, and the action approved or rejected. Approvals MUST NOT be recorded in a system controlled solely by the approver.

4.5 Data Classification

AG-1.14 PII identification

AG-1.14 **PII identification.** Organisations MUST maintain and apply a data classification schema that identifies personal data, sensitive personal data, and other regulated data categories. The classification schema MUST be applied to agent action parameters before the action is logged or forwarded.

AG-1.15 **Classification-based controls.** The policy definition under AG-1.1 MUST include controls specific to each data classification level, including: restrictions on which agents may process which data categories; restrictions on the external destinations to which classified data may be transmitted; and escalation triggers for actions involving sensitive personal data.

4.6 Incident Response

AG-1.16 Policy violation response

AG-1.16 **Policy violation response procedure.** Organisations MUST maintain a documented incident response procedure specifically for agent policy violations. The procedure MUST include: immediate containment steps (including agent suspension capability); investigation steps using audit records; root cause analysis; and policy remediation workflow.

AG-1.17 **Violation response testing.** The policy violation response procedure MUST be tested at least annually through a structured exercise or simulation. Test results and any identified gaps MUST be documented and remediated.

5 Tier 2 Controls — Cryptographic Accountability

All Tier 1 controls MUST be satisfied before Tier 2 certification may be granted. All controls in this section are REQUIRED for Tier 2 certification.

5.1 Hash-Chained Audit Integrity

AG-2.1 Cryptographic hash chaining

AG-2.1 **Cryptographic hash chaining.** Audit records MUST be linked in a hash chain where each record contains the cryptographic hash of its predecessor, computed using SHA-256 or a stronger approved algorithm. The chain MUST be initialised with a documented genesis record. Any gap, insertion, deletion, or modification in the chain MUST be detectable through chain verification.

AG-2.2 **Merkle tree construction.** Audit records MUST be periodically aggregated into a Merkle tree structure. The Merkle root MUST be computed and recorded at intervals not exceeding one hour. Merkle roots MUST themselves be hash-chained to provide consistency proof across periods.

AG-2.3 **Inclusion proof capability.** The audit system MUST be capable of producing a Merkle inclusion proof for any individual audit record on request, enabling a third party to verify that the record exists in the log without accessing all records. The proof MUST be computable in O(log n) time relative to log size.

5.2 Policy-Action Cryptographic Linkage

AG-2.4 Bundle revision hash recording

AG-2.4 **Bundle revision hash in audit records.** Every audit record MUST include the bundle revision hash of the policy bundle version that was loaded at the time the policy verdict was produced. This hash MUST be recorded at the time of evaluation and MUST NOT be retrospectively modified.

AG-2.5 **Policy bundle archival.** Every policy bundle version that has been used in a production policy evaluation MUST be archived in immutable storage for the full audit record retention period. The bundle MUST be retrievable by its revision hash.

AG-2.6 **Retroactive policy reconstruction prohibition.** Organisations MUST implement technical controls preventing the modification of a policy bundle without generating a new revision hash. The ability to produce a modified bundle with the same hash as a previously published bundle MUST be treated as a critical security incident.

5.3 Decision Replayability

AG-2.7 Deterministic evaluation requirement

AG-2.7 **Deterministic evaluation requirement.** The policy evaluation mechanism MUST be deterministic: given the same input document and the same policy bundle, the evaluation MUST always produce the same output. Non-deterministic evaluation mechanisms do not satisfy this requirement.

AG-2.8 **Replay capability.** Organisations MUST be capable of reproducing any historical policy decision using only: the input document recorded at evaluation time; and the policy bundle identified by the bundle revision hash in the audit record. Replay MUST be performable by a party with no access to the live production system.

AG-2.9 **Input document archival.** The input document provided to the policy evaluation engine for each action MUST be archived alongside the audit record for the full retention period, in a form that enables the replay required by AG-2.8.

5.4 GDPR/Retention Reconciliation

AG-2.10 Erasure mechanism

AG-2.10 **Erasure mechanism for immutable logs.** Organisations processing personal data of EU data subjects MUST implement a mechanism for satisfying Article 17 right-to-erasure requests against immutable audit records. The accepted technical approach is crypto-shredding: personal data fields are encrypted with a per-data-subject key before ingestion into the audit chain, and erasure is effected by destroying the key in a manner that renders the ciphertext computationally unrecoverable.

AG-2.11 **Erasure event documentation.** Each erasure event MUST be documented with: the data subject identifier; the timestamp of key destruction; the fields affected; and an attestation by the key management system that the key has been destroyed. This documentation MUST be appended to the audit chain.

AG-2.12 **Residual data assessment.** Organisations MUST assess and document whether any residual personal data exists in the audit chain outside of encrypted fields following a right-to-erasure request, including in: action parameter logs; metadata fields; and any derived or aggregated records.

5.5 Supervisor Independence Attestation

AG-2.13 Annual independence attestation

AG-2.13 **Annual independence attestation.** Organisations MUST obtain and maintain an annual written attestation from the operator of their agent supervision infrastructure confirming that: the supervisor is operated independently of the supervised agent platform; the supervisor operator has no commercial arrangement that creates an incentive to modify or withhold audit records; and no agent platform operator has write access to the audit store.

AG-2.14 **Conflict of interest disclosure.** Any commercial relationship between the supervision infrastructure operator and the agent platform vendor MUST be disclosed to the Accredited Assessor at the time of certification assessment. The assessor will determine whether the relationship creates a material threat to audit independence.

6 Tier 3 Controls — Forensic-Grade Non-Repudiation

All Tier 1 and Tier 2 controls MUST be satisfied before Tier 3 certification may be granted. All controls in this section are REQUIRED for Tier 3 certification.

6.1 External Blockchain Anchoring

AG-3.1 Public ledger commitment

AG-3.1 **Public ledger commitment.** Merkle roots MUST be committed to a public, permissionless blockchain at intervals not exceeding 24 hours. The blockchain MUST be one on which no single entity—including the certifying organisation, the supervision infrastructure operator, or any of their affiliates—has the ability to modify or reverse committed transactions. The specific blockchain network and smart contract address MUST be documented and published in the Council registry.

AG-3.2 **Anchor record in audit chain.** Each blockchain commitment MUST be recorded in the audit chain with: the transaction hash; the block number and timestamp; the blockchain network identifier; and the Merkle root committed. This record creates a bi-directional link between the audit chain and the public ledger.

AG-3.3 **Anchoring continuity.** Gaps in anchoring of more than 48 hours MUST be documented as anchoring gap events in the audit chain with the reason for the gap. Anchoring gaps do not invalidate preceding records but MUST be disclosed to the Accredited Assessor.

6.2 HSM-Backed Key Operations

AG-3.4 Hardware Security Module requirement

AG-3.4 **HSM requirement for key operations.** All cryptographic key operations supporting crypto-shredding erasure and audit record signing MUST be performed within a Hardware Security Module (HSM) rated to FIPS 140-2 Level 3 or equivalent. Software-based key management does not satisfy this requirement for Tier 3 certification.

AG-3.5 **Key custody and access controls.** HSM key access MUST be subject to documented custody controls including: multi-party authorisation for key destruction operations; access logging for all key operations; and annual key custody audit. Key destruction events MUST be logged by the HSM and the log MUST be independently verifiable.

6.3 Hardware-Attested Supervisor Integrity

AG-3.6 TEE or TPM attestation

AG-3.6 **Hardware attestation of supervisor.** The agent supervisor MUST produce cryptographic evidence of its own integrity using either: a Trusted Execution Environment (TEE) attestation document (e.g., AWS Nitro attestation, Intel TDX quote, AMD SEV-SNP attestation report); or a TPM remote attestation quote demonstrating that the specific, unmodified supervisor software is running on the attesting host. Software-only integrity claims do not satisfy this requirement.

AG-3.7 **Attestation verification.** The attestation produced under AG-3.6 MUST be verified by an independent party—either the Council registry or a designated third-party verifier—at intervals not exceeding 24 hours. Failed attestation verification MUST trigger an alert and MUST be recorded in the audit chain.

AG-3.8 **Attestation anchoring.** Attestation verification results MUST be incorporated into the Merkle tree and therefore into the blockchain anchor, such that the audit record includes cryptographic proof of supervisor integrity for the period covered by each anchor.

6.4 Erasure Certificates

AG-3.9 Cryptographic erasure certificates

AG-3.9 **Cryptographic erasure certificate issuance.** Upon completion of a right-to-erasure request, the HSM MUST issue a signed erasure certificate containing: the data subject identifier (or pseudonym); the timestamp of key destruction; the set of audit record identifiers in which the subject's data appeared in encrypted form; and the HSM attestation of key destruction. This certificate MUST be incorporated into the Merkle tree and therefore anchored to the public blockchain.

AG-3.10 **Erasure certificate registry.** Organisations MUST maintain a registry of issued erasure certificates, indexed by data subject identifier (or pseudonym), accessible to the data subject upon request as evidence of erasure. The registry MUST itself be protected against modification.

6.5 Cross-Border Jurisdiction Mapping

AG-3.11 Jurisdiction metadata

AG-3.11 **Jurisdiction metadata in audit records.** Every audit record MUST include metadata identifying: the jurisdiction of the data subject(s) whose data was processed in the action; the physical location of the data processing; and the jurisdiction of the destination system. This metadata MUST be sufficient to enable reconstruction of cross-border data flows for regulatory examination.

AG-3.12 **Jurisdiction-based policy controls.** The policy definition under AG-1.1 MUST include jurisdiction-specific controls covering at minimum: restrictions on cross-border transfer of personal data; differential escalation triggers based on the data subject jurisdiction; and documentation of the legal transfer mechanism relied upon for each cross-border data flow category.

6.6 Adversarial Resilience

AG-3.13 Adversarial testing

AG-3.13 **Adversarial testing programme.** Organisations MUST conduct adversarial testing of their agent supervision infrastructure at least annually, covering at minimum: prompt injection attacks targeting the agent's decision-making; agent-as-proxy attacks attempting to coerce the agent into transmitting adversarial content through supervised channels; and policy bypass attempts including routing around the supervisor at the network layer.

AG-3.14 **Hybrid monitoring.** The agent supervisor MUST evaluate both the agent's internal reasoning (Chain-of-Thought, where accessible) and its tool call parameters and responses. Monitoring that evaluates only one of these layers does not satisfy this requirement.

AG-3.15 **Adversarial test documentation.** All adversarial testing MUST be documented with: test scenarios; observed supervisor behaviour; identified gaps; and remediation actions taken. This documentation MUST be available to the Accredited Assessor.

7 Certification Assessment Process

7.1 Assessor Accreditation

Certification assessments MUST be conducted by an Accredited Assessor. Accredited Assessors are organisations or individual practitioners who have demonstrated the following competencies to the Council's satisfaction and hold a current accreditation:

- Demonstrated technical competency in AI security, policy-as-code evaluation, and cryptographic audit trail verification.
- Familiarity with the regulatory frameworks cited in Section 1.3 as applicable to the certification tier being assessed.
- Independence from the organisation being assessed and from any agent platform vendor deployed by the assessed organisation.
- Completion of the Council's AGCS Assessor Training and Examination programme.

Assessor accreditation is granted for a period of two years and is subject to annual continuing education requirements. A current list of Accredited Assessors is maintained in the Council public registry.

7.2 Assessment Scope and Evidence

The Accredited Assessor determines the scope of the certification assessment in consultation with the certifying organisation. The assessment scope MUST cover all autonomous AI agents deployed by the organisation in production environments at the time of assessment. The assessment is conducted against the controls applicable to the requested certification tier.

The following evidence types are accepted as control evidence. The Accredited Assessor has discretion to require additional evidence where submitted evidence is insufficient to demonstrate control satisfaction:

Evidence Type	Applicable Controls	Notes
Policy bundle archive with revision history	AG-1.1, AG-1.2, AG-2.4, AG-2.5	Version control export showing full change history.
Audit log sample with gap analysis	AG-1.5, AG-1.6, AG-1.7	Minimum 30-day sample; assessor may request longer period.
Hash chain verification report	AG-2.1, AG-2.2, AG-2.3	Produced by audit system; assessor independently verifies.
Policy replay demonstration	AG-2.7, AG-2.8, AG-2.9	Live demonstration using a historical dispute scenario.
Blockchain anchor transaction records	AG-3.1, AG-3.2, AG-3.3	On-chain verification; assessor independently queries blockchain.
HSM key custody audit	AG-3.4, AG-3.5	Provided by HSM operator; assessor reviews access logs.

Evidence Type	Applicable Controls	Notes
TEE or TPM attestation verification log	AG-3.6, AG-3.7, AG-3.8	Produced by attestation verifier; assessor independently verifies selected samples.
Erasure certificate samples	AG-3.9, AG-3.10	Assessor verifies certificate incorporation in Merkle tree.
Adversarial test reports	AG-3.13, AG-3.14, AG-3.15	Third-party test reports preferred; internal reports accepted with assessor review of methodology.
Independence attestation	AG-2.13, AG-2.14	Written attestation from supervisor operator; assessor reviews for conflicts.

7.3 Certification Outcomes

Following the assessment, the Accredited Assessor issues one of the following outcomes:

- **Certified:** All controls for the requested tier are satisfied. The assessor submits a certification report to the Council registry. The Council issues a certificate valid for 24 months from the assessment date.
- **Certified with Observations:** All controls are satisfied but the assessor has identified areas of weakness that, while not constituting control failures, represent material risk. Observations are recorded in the registry and must be addressed at the next annual review.
- **Not Certified — Remediation Required:** One or more controls are not satisfied. The assessor documents the specific gaps. The organisation may submit evidence of remediation within 90 days for re-assessment without a full re-assessment fee.
- **Not Certified — Assessment Incomplete:** The organisation was unable to provide sufficient evidence for assessment to be completed. A new full assessment is required.

7.4 Ongoing Compliance

Certification is valid for 24 months from the assessment date. Organisations MUST notify the Council within 30 days of any material change to their agent supervision infrastructure, policy governance process, or audit trail architecture. The Council may require an interim assessment in response to material change notifications. Annual self-attestation of continued compliance is required in the 12-month period between full assessments.

8 The Agent Governance Certification Council

8.1 Council Purpose and Independence

The Agent Governance Certification Council is the governing body responsible for: maintaining and publishing the AGCS standard; operating the Accredited Assessor programme; maintaining the public certification registry; and evolving the standard in response to technical developments, regulatory changes, and stakeholder input.

The Council is constituted as an independent entity. No single member organisation exercises unilateral control over the standard text or the certification registry. The Council's independence from any single vendor is a condition of its authority to issue certifications that are recognised by third parties, regulators, and counterparties.

8.2 Public Registry

The Council maintains a public registry containing:

- All currently certified organisations, their certification tier, the agent deployment scope covered, and the certification expiry date.
- All Accredited Assessors, their accreditation scope (tier coverage), and accreditation expiry date.
- All blockchain anchor addresses registered by certified organisations, enabling independent on-chain verification.
- The complete version history of the AGCS standard, including superseded versions.
- Public comment submissions and Council responses for each standard version.

8.3 Standard Evolution

The AGCS standard will be reviewed and updated on an 18-month cycle, or more frequently in response to material changes in the threat landscape, regulatory environment, or underlying technology. Version updates follow a structured process:

1. Draft preparation by the Council Technical Secretariat, incorporating feedback from the Assessor community, certified organisations, and the Technical Advisory Board.
2. Public comment period of not less than 60 days. Comments accepted from any party.
3. Council review and resolution of public comments, with published responses.
4. Publication of the new version with a transition period of not less than 12 months during which either version may be used for certification.

9 Legal Status and Limitations

AGCS certification is a voluntary third-party conformity assessment. It does not constitute legal advice and does not guarantee compliance with any specific law or regulation. Organisations are responsible for determining the applicability of specific laws and regulations to their activities and for obtaining appropriate legal advice.

The Council makes no warranty, express or implied, that AGCS certification satisfies the requirements of any specific regulatory instrument, including but not limited to the EU AI Act, GDPR, MiFID II, HIPAA, or PCI DSS. The relationship between AGCS controls and specific regulatory requirements described in Section 1.3 represents the Council's good-faith technical assessment and does not constitute a legal determination.

AGCS certification does not constitute evidence of security against all possible attacks or misuse scenarios. The adversarial resilience controls in Section 6.6 represent a baseline testing programme, not a guarantee of immunity to novel or sophisticated attacks.

LEGAL NOTE

This draft standard should be reviewed by qualified legal counsel before use in any regulatory filing, procurement requirement, or contractual obligation. The Council recommends that organisations intending to incorporate AGCS certification requirements into contracts or procurement frameworks obtain independent legal review of the applicable controls and definitions.

10 Public Comment Process

Version 0.9 of the AGCS standard is published for public comment prior to finalisation of Version 1.0. The Council welcomes technical, legal, and operational comment from all stakeholders, including enterprises deploying autonomous agents, security and audit practitioners, AI vendors, academics, and regulatory bodies.

10.1 Comment Submission

Comments should be submitted in writing to the Council at the address published on the Council website. Comments should identify: the specific section, control identifier, or definition being addressed; the nature of the comment (technical error, ambiguity, missing requirement, over-specification, or general feedback); and, where applicable, a proposed alternative formulation.

10.2 Priority Review Areas

The Council specifically invites comment on the following areas where the draft may be under-specified or where alternative approaches merit consideration:

- AG-1.9 (Separate authority): The boundary between acceptable architectural independence and requirement for separate legal entity operation. The Council recognises this control may be difficult to satisfy for organisations using vertically integrated agent platforms.
- AG-2.10 (Erasure mechanism): Whether crypto-shredding is the sole acceptable approach or whether alternative erasure mechanisms should be recognised, and what evidentiary standard should apply.
- AG-3.1 (Blockchain network selection): Whether the standard should specify approved blockchain networks or maintain technology neutrality, and the implications of network deprecation or compromise.
- AG-3.6 (Hardware attestation): Whether TPM attestation on commodity hardware provides sufficient assurance equivalence to TEE-based attestation, or whether the two should be distinguished in the standard.
- Tier boundaries: Whether the current allocation of controls across tiers appropriately reflects the governance maturity journey of a typical enterprise.

Annex A — Illustrative Control Evidence Mapping

This annex provides illustrative guidance on the types of evidence that may satisfy specific controls. This guidance is non-normative. The Accredited Assessor has discretion to accept equivalent evidence not listed here and to require additional evidence where listed types are insufficient.

Control ID	Control Title	Illustrative Satisfying Evidence
AG-1.1	Machine-evaluable policy	OPA/Rego policy bundle with documented schema; Sentinel policy files; AWS SCP JSON policies with documented evaluation engine.
AG-1.2	Version control	Git repository with full commit history; policy change approval workflow records (pull requests, code review approvals).
AG-1.5	Universal action logging	Demonstration that a test action executed by a zero-instrumentation agent is captured in the audit log; network tap or proxy logs independent of agent SDK.
AG-1.9	Separate authority	Architecture diagram demonstrating audit store network isolation from agent platform; separate cloud account for audit infrastructure; contractual separation with independent audit operator.
AG-2.1	Hash chaining	Audit system documentation demonstrating hash chain algorithm; live verification of chain integrity across a 30-day audit window; demonstration of tamper detection on a modified test record.
AG-2.4	Bundle revision hash	Sample audit records showing bundle revision hash field; demonstration that the same hash resolves to the correct bundle in the archive.
AG-2.8	Replay capability	Live replay demonstration: assessor selects a historical action at random; organisation reproduces the policy decision using only the archived input document and bundle.
AG-3.1	Blockchain anchoring	On-chain transaction records; smart contract address; demonstration that the on-chain Merkle root matches the root computed from the audit records for the relevant period.
AG-3.6	Hardware attestation	AWS Nitro attestation document with PCR values; TPM attestation quote with IMA measurement log; independent verification service log showing attestation verification results.
AG-3.9	Erasure certificates	Sample erasure certificate showing HSM signature; demonstration that the certificate is incorporated into the Merkle tree; on-chain anchor for the relevant period.

Annex B — Regulatory Alignment Reference

This annex provides a non-normative reference mapping between AGCS controls and specific regulatory requirements. This mapping represents the Council's technical assessment and does not constitute legal advice. Organisations should obtain independent legal review of regulatory obligations applicable to their specific activities and jurisdictions.

Regulation / Article	Obligation	AGCS Controls
EU AI Act — Article 9	Risk management system for high-risk AI	AG-1.1, AG-1.4, AG-1.11, AG-1.16, AG-3.13
EU AI Act — Article 12	Automatic logging for high-risk AI systems	AG-1.5, AG-1.6, AG-1.7, AG-1.8, AG-2.1
EU AI Act — Article 14	Human oversight measures	AG-1.11, AG-1.12, AG-1.13
GDPR — Article 17	Right to erasure	AG-2.10, AG-2.11, AG-2.12, AG-3.9, AG-3.10
GDPR — Article 30	Records of processing activities	AG-1.6, AG-3.11, AG-3.12
MiFID II — Article 16(7)	Record-keeping of communications and orders	AG-1.8, AG-2.1, AG-2.4, AG-3.1
HIPAA — 45 CFR §164.312(b)	Audit controls for electronic PHI	AG-1.5, AG-1.6, AG-1.9, AG-2.1
NIST AI RMF — GOVERN 1.2	Policies for AI risk	AG-1.1, AG-1.2, AG-1.3, AG-1.4
NIST AI RMF — MANAGE 2.2	Mechanisms for addressing AI incidents	AG-1.16, AG-1.17
SOC 2 — Integrity CC7	System monitoring	AG-1.5, AG-1.7, AG-2.1, AG-2.3