

AI Security Policy

Template 4 from the free ISO/IEC 27090 course at shahzadasghar.com. Blue italic text is a placeholder to replace.

Document control	
Document owner	<i>[Chief Information Officer]</i>
Approved by	<i>[Executive committee / risk committee]</i>
Version	0.1 draft
Approval date	<i>[date]</i>
Next review	<i>[date, at most 12 months after approval]</i>
Reference standards	ISO/IEC 27090:2026; ISO/IEC 27001:2022; ISO/IEC 27002:2022; ISO/IEC 42001:2023; ISO/IEC 22989:2022

1. Purpose, scope and definitions

This policy sets the minimum security requirements for artificial intelligence (AI) systems that [the organisation] develops, procures or uses. It applies ISO/IEC 27090:2026 guidance within the existing information security management system.

Scope: every AI system and AI feature listed in the AI system inventory (Template 1), including AI capabilities inside vendor software, cloud AI services and internally built models, across the full life cycle from inception to retirement.

Definitions follow ISO/IEC 22989:2022. Key terms: AI system, model, training data, inference, provider, deployer, life cycle stage.

2. Roles and responsibilities

- Chief Information Officer: accountable for the AI inventory, life-cycle ownership and this policy.
- Security lead: accountable for AI threat and risk assessment, controls, monitoring and incident response.
- AI product owner: responsible for each system's inventory record, acceptance and operation.
- Data owner: responsible for provenance, classification and integrity of training and retrieval data.
- Legal and privacy: consulted on personal data, regulatory and contractual matters.

Provider duties (model robustness, training data integrity, secure development, documentation of limits) and deployer duties (integration security, access, monitoring, acceptance) are assigned per system in the inventory. The responsibility matrix in Annex A applies.

3. AI system inventory and risk tiering

- Every AI system or feature is registered in the inventory before go-live and reviewed at least annually.
- Each entry records purpose, AI type, source, data used, classification, exposure, owners and risk tier.
- Risk tier rule: High where personal or confidential data, external exposure, or decisions about people or money are involved; Medium where internal with confidential data; Low where public data and no decisions.

4. Threat and risk assessment

- A threat and risk assessment (Template 2) is completed for every High-tier system before go-live and for Medium-tier systems within one quarter of go-live.

- The assessment covers every threat in the ISO/IEC 27090 catalogue: data poisoning, evasion, membership inference, model inversion, model exfiltration, direct model poisoning, sensitive model output, prompt injection, output containing injection attacks, and the Annex B conventional attacks.
- Scores of 15 or more are treated before go-live; 8 to 14 within the quarter; below 8 may be accepted with a recorded decision.
- Assessments are repeated on retraining, vendor version change, new data source, new tool or integration, after an incident, and at least every 12 months.

5. Secure development, data governance and supply chain

- Training data, test data, pipeline code, notebooks, configurations, model artefacts and retrieval stores are protected assets under ISO/IEC 27002 controls 8.25, 8.31 and 8.4.
- Data provenance and integrity are verified before data is used for training, fine-tuning or retrieval. Data snapshots are retained for forensics.
- Supplied models, data sets, libraries and MLOps tools are treated as supply-chain components: supplier assessment, contract clauses, signed and versioned artefacts, and an AI bill of materials per system.
- Retrieval data in production receives the same protection as training data.

6. Deployment, access control and rate limiting

- Inference endpoints require authentication, per-client rate limits and usage baselines.
- Confidence and probability detail in outputs is limited to what the use case needs.
- AI agents and tools run with least privilege; actions with financial, access or safety impact require human approval.
- Model output is treated as untrusted input by every consuming system.

7. Logging, monitoring and revalidation after change

- Inputs, outputs, user or client identity, time, model version, prompt template version, tool calls and retrieval sources are logged in line with controls 8.15 and 8.16, with retention and protection appropriate to their sensitivity.
- Systems are monitored for accuracy drift, anomalous query patterns, injection attempts and sensitive output.
- Any change (retraining, vendor update, new data or tool) triggers revalidation of security controls before return to service.

8. Incident response for AI-specific incidents

- Playbooks exist for suspected poisoning, data leakage through outputs, and injection-driven unauthorised actions.
- The security lead may remove a model from service pending investigation. Criteria are defined in advance.
- AI incidents are reported, recorded and reviewed through the existing incident management process (controls 5.24 to 5.28).

9. Testing, red teaming and acceptance criteria

- High-tier systems undergo adversarial testing and red teaming, including indirect prompt injection, before go-live and at least annually.
- Controls are evaluated together, not one at a time, because mitigations can interact and weaken each other.
- Acceptance requires a completed assessment, treated high scores, logging in place and documented residual risk.

10. Review cycle and links to other standards

- This policy is reviewed every 12 months or when ISO/IEC 27090, ISO/IEC 27001 or applicable regulation changes.
- AI risk is reported quarterly to the risk committee.
- This policy operates within the ISMS (ISO/IEC 27001) and, where adopted, the AI management system (ISO/IEC 42001).

Annex A: Responsibility matrix (RACI)

R responsible, A accountable, C consulted, I informed. One A per row.

Activity	CIO	Security lead	AI product owner	Data owner	Legal / privacy
Maintain AI inventory	A	C	R	C	I
Threat and risk assessment	I	A/R	C	C	C
Data provenance and integrity	I	C	C	A/R	C
Supplier and model assurance	A	R	C	I	C
Logging, monitoring, detection	I	A/R	C	I	C
Testing and revalidation	A	R	R	I	I
Incident response	I	A/R	C	C	C
Acceptance and go-live decision	A	C	R	C	C
Policy review	A/R	C	I	I	C

Annex B: Related templates

- Template 1: AI system inventory register (Excel)
- Template 2: AI threat and risk assessment, one workbook per system (Excel)
- Template 3: Control mapping and gap analysis (Excel)