GAMUT AI

End User & Assessor Manual

A complete guide to operating Gamut, assessing AI systems, governing AI agents and producing defensible evidence.

Edition 2026-07-21
Comprehensive customer edition

About this manual

This manual is generated from Gamut's maintained customer guidance after verification against current product behaviour. It is written for end users, assessors, reviewers, evidence owners, workspace administrators and runtime policy approvers.

The manual explains the workflows, responsibilities and decision points that users need. Only customer-facing operating information is included. The live workspace remains authoritative for the features, plan entitlements and permissions available to a particular tenant and user.

> Framework references support structured governance work. They do not replace licensed standards, legal advice, certification audits or regulator guidance. AI Assist is advisory and never approves evidence, determines legal compliance, confirms scope or replaces accountable human judgement.

Contents

Part 1: Orientation and first assessment

This part compiles the current public operating guidance for the workflows in scope.

1. What is Gamut AI

Gamut AI is an AI governance lifecycle platform that helps organisations discover AI, classify risk, assess against multiple frameworks, and produce audit-ready evidence.

Gamut AI is an **AI governance lifecycle platform**. It gives organisations one place to discover the AI they use, understand and classify the risk it carries, assess it against the frameworks that matter, and produce the evidence that boards, buyers, auditors and regulators ask for.

It is built for organisations that need to demonstrate structured governance over their AI systems, banks, financial institutions, regulated enterprises and any organisation whose AI adoption has outgrown spreadsheets and fragmented point tools.

The problem Gamut solves

AI adoption usually runs ahead of governance. Teams buy AI tools, embed models in products, and build agentic workflows faster than risk, compliance and security functions can track them. The result is familiar:

- No reliable inventory of where AI is used, who owns it, or what data it touches.
- Risk decisions made informally and inconsistently, with no traceable rationale.
- Policy that exists on paper but cannot be evidenced when someone asks for proof.
- Evidence scattered across documents, email threads and screenshots, gathered in a panic at audit time.

Gamut replaces that with a single, governed operating layer that takes an AI system from first discovery through to board-level reporting and continuous improvement.

What Gamut gives you

- **A standing AI inventory**: a register of AI systems, use cases, owners, suppliers, data context and lifecycle status.
- **Consistent risk classification**: deterministic risk tiering driven by a configurable governance weighting profile of purpose, impact, human oversight, data exposure and operating context.
- **Multi-framework assessment**: assess a system against GTSAF, the EU AI Act, NIST AI RMF, ISO/IEC 42001 and 42005, NAGF, ACRS, the Agentic Trust Framework (ATF) and MAESTRO.
- **Evidence as a first-class object**: evidence requests, artefacts, quality tracking, findings and remediation, captured as governance work happens.
- **Audit-ready outputs**: workpaper-grade exports and board reporting that trace from requirement to evidence to finding.
- **AI assistance, governed**: an AI Consultant grounded in your own assessment records, plus policy drafting, all proxied server-side and metered by plan.
- **Governed agentic AI**: Agentic CISO, Gamut Gateway and Gamut Claw extend governance to AI that takes action, not just AI that answers questions.

Design principles

Gamut is built around a small number of deliberate choices:

- **Evidence, not just policy.** Governance is only credible when it can be proven. Gamut is organised around producing defensible evidence, not generating documents.
- **Traceability end to end.** Every record connects: AI system -> risk decision -> control expectation -> evidence -> finding -> remediation.
- **Governance is a lifecycle, not an event.** Gamut keeps governance current as systems and obligations change, rather than producing a point-in-time snapshot.
- **Trust by construction.** AI analysis is proxied server-side so model keys are never exposed to the browser, tenants are isolated, and access is controlled by role.

How this documentation is organised

- **Getting started**: the fastest path from sign-in to your first assessed AI system.
- **Platform**: the modules that make up Gamut and how they work.
- **Governance frameworks**: the frameworks Gamut supports and how routing works.
- **Agentic stack**: Agentic CISO, Gateway and Claw.
- **Administration**: workspaces, users, SSO, plans and audit.
- **API reference**: integrate Gamut with the rest of your stack.

> **New to AI governance?:** Start with The governance lifecycle for the mental model, then Core concepts & glossary for the vocabulary used throughout these docs.

2. Who Gamut is for

Gamut serves organisations running AI governance and the auditors who review it, across risk, compliance, security, procurement and the board.

Gamut is built for two audiences that have to trust the same records: the **organisation running AI governance** and the **auditor reviewing it**. The platform is designed so that both work from one source of truth.

For organisations running AI governance

Gamut gives AI governance, risk, compliance, security, procurement and business teams a shared operating layer for responsible AI adoption.

- One place to see AI systems, owners, risks and evidence.
- A practical route from AI literacy to board reporting.
- A governance lifecycle that supports EU AI Act readiness and wider assurance needs.
- Clear outputs for leadership, risk, compliance, security and procurement teams.

Roles that use Gamut

Role	What they get from Gamut
Chief Risk / Compliance Officer	A defensible view of AI risk across the organisation and board-ready reporting.
AI governance / responsible AI lead	A repeatable operating model across the lifecycle and across frameworks.
CISO / security	Control expectations, evidence and the agentic stack for AI that takes action.
Procurement / vendor risk	Structured assessment of vendor and third-party AI before and during use.
System / model owners	A clear record of their AI system, its risk tier and the controls expected of it.

For auditors reviewing AI governance

Gamut helps auditors and reviewers assess governance activity through structured records, evidence quality, findings and remediation history.

- Structured evidence instead of scattered files and informal explanations.
- Traceability from AI system record to risk decision, control expectation, evidence and finding.
- Consistent review packs for internal audit, client assurance and external review.
- Clearer follow-up on findings, exceptions, remediation and residual risk.

Where Gamut fits

Gamut is most valuable when AI adoption has outgrown informal tracking but governance still has to satisfy external scrutiny, buyers running vendor due diligence, boards asking for assurance, internal audit testing controls, and regulators assessing readiness against the EU AI Act and equivalent regimes.

Tip: If you are evaluating Gamut for a specific obligation, for example EU AI Act readiness, start with the relevant framework page, which explains what Gamut captures and produces for that regime.

3. Core concepts & glossary

The vocabulary used throughout Gamut, AI systems, intake, risk tiers, frameworks, controls, evidence, findings, model cards and the agentic stack.

This page defines the core objects and terms used across Gamut and throughout this documentation. They are deliberately consistent: the same words mean the same thing in the product, the API and these docs.

Core objects

AI system : A registered unit of AI use, an internal application, an embedded model, a vendor AI tool or an agentic workflow. The AI system is the anchor that every other record connects back to. See Registry & Discovery.

Use case : The business purpose an AI system serves. One AI system can support multiple use cases, each with its own context and impact.

Intake : The structured capture of context for an AI system or use case, purpose, users, data, oversight, suppliers and geography, used to determine the level of governance required. See Intake & risk tiering.

Risk tier : The classification of an AI system's risk, derived from intake. Risk tiers drive which controls and frameworks apply.

Framework : A structured set of controls or requirements, for example GTSAF, the EU AI Act, NIST AI RMF or ISO/IEC 42001. See Frameworks overview.

Domain : A grouping of related controls within a framework. GTSAF, for example, organises its controls across 17 domains.

Control : A specific governance expectation that can be assessed and evidenced. Controls belong to domains within a framework.

Assessment : The act of scoring an AI system against a framework's controls, recording the rationale and current state. See Assessments & control testing.

Control test : Evidence that a control is operating effectively, beyond simply being designed. Control testing produces the proof that backs an assessment.

Evidence : An artefact that supports a governance claim, a document, export, screenshot, log or record. Evidence is requested, captured, quality-checked and linked to controls and findings. See Evidence & findings.

Finding : A recorded deficiency, gap or exception identified during assessment or audit, tracked through to remediation and closure.

Remediation : The work that closes a finding, with its progress visible to the right stakeholders.

Model card : A document of a model's technical and ethical characteristics, intended use, data, limitations, performance and oversight. See Model cards.

Policy : An AI governance policy, which Gamut can help draft with AI assistance. See Policy generation.

The agentic stack

Agentic AI : AI that takes action, calling tools, invoking APIs and running workflows, rather than only producing text. Agentic AI needs runtime governance, not just design-time assessment.

Agentic CISO : Gamut's capability for governing agentic AI, the agent register, ATF assessment, tool and data governance, approvals and runtime evidence. See Agentic CISO.

Gamut Gateway : The policy decision and enforcement layer that applies governance policy to agent actions at runtime. See Gamut Gateway.

Gamut Claw : The secure agent execution layer that requests and runs work only through Gateway-controlled paths. See Gamut Claw.

ATF (Agentic Trust Framework) : The framework that defines trust, control and evidence expectations for agentic AI, which Gateway and Claw implement at runtime. See ATF.

ATF level : An agent's autonomy ceiling, L1 Intern, L2 Junior, L3 Senior or L4 Principal. Gateway enforces a different action boundary at each level, from read-only at L1 to strategic autonomy at L4.

ACRS (Agentic Capability Risk Score) : A score of an agent's capability risk across dependency, action, access and harm dimensions, producing a band that sets how tightly to govern. See ACRS.

Connector : A governed adapter through which Gateway performs a tool, model or data call. The connector holds the credential and endpoint policy; the agent never does. See Connector catalog.

Tool permission : The business authorisation in Agentic CISO that an agent may use a given tool. It must align with a registered connector before Gateway allows the call.

Approval gate : A configured requirement that a named human approves a sensitive or mutating action before Gateway will allow it.

Gateway decision : The verdict Gateway returns for a requested action, allow, require approval, degrade or block, with a full decision path and a signed, short-lived authorisation token on allow.

Claw task : A governed unit of agent work of a defined type (for example governed reasoning, evidence gap analysis, incident triage), run under a lease with bounded steps and redacted output. See Gamut Claw.

Trajectory event : A journaled record of what an agent did at runtime, hash-chained and fed back to Agentic CISO as runtime evidence.

Platform terms

Workspace / tenant : An isolated environment for one organisation. Each workspace keeps its own users, systems, assessments and evidence separate from every other. See Workspaces & tenancy.

Role : The set of permissions a user holds within a workspace, controlling what they can see and do. See Users & roles.

Entitlement / plan : The feature tier available to a workspace, which gates frameworks and capabilities. See Plans & entitlements.

Audit log : The immutable record of state-changing actions in a workspace, used for accountability and review. See Audit log.

4. Quickstart

Get from sign-in to your first assessed AI system in Gamut, with the shortest path through the governance lifecycle.

This quickstart takes you from signing in to your first assessed AI system. It is the fastest useful path through Gamut; later pages go deeper on each step.

> **Access:** Gamut is available at **run.gamutassure.com**. If you do not yet have a workspace, ask your administrator to invite you, or start a trial from the sign-in page.

Before you start

You will move faster if you have one real AI system in mind to register, an internal tool, an embedded model or a vendor AI product. Have a rough idea of:

- What it does and who uses it.
- What data it touches.
- Who owns it.
- Which obligation you care about most (for example EU AI Act readiness).

Step 1: Sign in to your workspace

Sign in at run.gamutassure.com. You land in your **workspace**, an environment isolated to your organisation. Everything you create stays within it. If you belong to more than one workspace, confirm you are in the right one.

Step 2: Register an AI system

Open **AI System Records** and add your AI system. Give it a name, owner and a short description of what it does. This creates the anchor record that everything else connects to.

-> Full walkthrough: Register your first AI system.

Step 3: Complete intake and set a risk tier

Run **AI Use Case Intake & Approval** for the system: capture purpose, users, data exposure, human oversight, supplier involvement and geography. Gamut uses this, through the **Risk Tiering Engine** and **ACRS**, to set a **risk tier** and route the system to the controls and frameworks that apply.

-> More detail: Intake & risk tiering.

Step 4: Run an assessment

Open the framework routing suggested for the system, for example GTSAF or EU AI Act readiness, and score its controls on each framework's answer scale, recording your rationale as you go.

-> Full walkthrough: Run your first assessment.

Step 5: Capture evidence and findings

As you assess, use the **Evidence Tracker** to raise evidence requests and attach artefacts, and the **Findings Register** to record any gaps. Evidence and findings link back to the controls they relate to, so your assessment is defensible rather than just complete.

-> More detail: Evidence & findings.

Step 6: Produce a report

Open the **Board Dashboard** or generate a **Workpaper Pack** to see your AI inventory, risk picture, evidence quality and open findings in one place, suitable for leadership and review.

-> More detail: Reporting & exports.

Next steps

5. Register your first AI system

Walk through capturing an AI system in AI System Records, the anchor record that intake, assessment, evidence and every finding connects back to.

The **AI system** is the anchor of everything in Gamut. It is registered in **AI System Records**, and registering one well makes every later step, intake, risk tiering, assessment, evidence and reporting, faster and more consistent. Nothing else can attach to a system until the record exists: the registry is what intake and assessment feed from.

Work inside a workspace

Everything you create belongs to a **workspace** (also called an assessment), the container that scopes your systems, intake, evidence and reports. Make sure you have created or opened a workspace before you begin; AI System Records lists the systems within the workspace you are in.

What counts as an AI system

Register anything that uses AI and carries governance relevance:

- Internal applications and services that embed a model.
- Vendor or third-party AI tools your teams use.
- GenAI assistants and copilots in active use.
- Agentic workflows that take action through tools and APIs.

If in doubt, register it. An under-governed AI system is a bigger risk than an extra record. If a system was surfaced by the AI Discovery Inbox, you can promote it into AI System Records rather than entering it from scratch.

Fields to capture

A system record is structured. You do not need every field on day one, but the more you capture, the better intake and routing work later:

Group	Fields
Identity	Name, system ID, version, lifecycle stage (development through retired).
Ownership	Owner and owner email, department, responsible-AI contact.
Nature	AI technique, model type, autonomy level, human-oversight model, vendor, deployment type and environment.
Data & exposure	Data classification, personal-data involvement, data sources, users, affected persons, geographies, regulatory exposure.
Governance	Risk tier, ACRS score and review dates (these are set or confirmed later, in intake and risk tiering).

Steps

1. Open **AI System Records** in your workspace.
2. Add a new system record (or promote one from the AI Discovery Inbox).
3. Enter the **identity** and **ownership** fields: name, owner, owner email, department.
4. Describe the **nature** of the system: what technique and model type it uses, its autonomy level and how humans stay in oversight.
5. Record the **data and exposure** context: data classification, personal data, who is affected, and where it operates.
6. Save. The system now appears in your inventory and is ready for intake.

Good practice

- **One owner, always.** Ownership is what makes governance actionable.
- **Describe purpose, not implementation.** Reviewers care about what the system does and why.
- **Register early.** Capturing a system at proposal stage is far cheaper than reconstructing its history later.
- **Let discovery feed the registry.** A registry that misses systems produces governance that only looks complete.

Next

With the system registered, complete AI Use Case Intake and set a risk tier, then run your first assessment.

6. Run your first assessment

Ground a system through intake and risk tiering, route it to the right frameworks, then score its controls with rationale, evidence and findings.

An **assessment** scores a registered AI system against the controls of a framework and records why. This is where governance becomes defensible: not just a decision, but a decision with a documented rationale and evidence behind it. In Gamut, a good assessment is the end of a short chain that grounds the system first.

Ground the system before you score it

A framework score is only as good as the context behind it. Before assessing, work through the chain that feeds it:

1. **AI System Records** establishes what the system is.
2. **AI Use Case Intake & Approval** explains what it does, who it affects, what data and decisions are involved, and records an accountable approval. See Intake & risk tiering.
3. **Risk Tiering Engine** and **ACRS Risk Assessment** set the depth of governance expected, confirming the risk tier and capability band.
4. **Assessment Plan & Assurance Routing** routes the system to the frameworks it actually needs.

By the time you open a framework, the system is grounded and routed, so you are scoring against the right controls rather than guessing which apply.

Choosing a framework

Routing will suggest the frameworks that fit. Common starting points:

- **EU AI Act readiness**: to demonstrate readiness for the EU AI Act.
- **GTSAF**: for depth, 358 controls across 17 domains.
- **NIST AI RMF** or **ISO/IEC 42001**: if you align to those standards.
- **ATF**: for systems that take action as agents.

You can assess the same system against more than one framework; Gamut keeps each distinct while sharing the underlying system, evidence and findings. See Frameworks overview.

Scoring controls

Open the framework and work through its controls, domain by domain. Each framework uses the answer scale that fits it:

Framework	How controls are scored
GTSAF, ATF	Each control is answered **Yes / No / N/A**. ATF adds a depth rating: how embedded a Yes is, or how far a No has progressed.
EU AI Act, NAGF	Each requirement is **Compliant / Non-compliant / N/A**, with a depth rating.
ISO/IEC 42005	Each requirement is scored on a **maturity scale**.
MAESTRO	Each threat is scored for **likelihood and impact**, which combine into a risk band.

For each control, whatever the scale:

- Record the **answer or score**.
- Capture the **rationale**, why you reached that conclusion.
- Attach or request evidence where relevant.
- If a control genuinely does not apply, mark it **N/A with a documented justification**. An unjustified N/A stays in scope and counts as not met.
- Raise a **finding** for any gap, deficiency or exception.

Scoring is **coverage-inclusive**: controls you have not assessed count as not-yet-met, so the score reflects real progress rather than only the parts you have looked at. See how scoring works for the model that applies across every framework.

Capture evidence, tests and findings

As you score, the supporting registers are one click away:

- **Evidence Tracker**: raise evidence requests and attach artefacts against controls.
- **Testing Centre**: record control tests with design and operating effectiveness.
- **Findings Register**: track gaps through root cause, remediation and validated closure.

After the assessment

7. Invite your team

Bring colleagues into Gamut and assign roles from the real catalogue, a tenant role for organisation-wide access plus additive assessment-level roles for the work they touch.

Gamut is a shared operating layer, so it works best with the right people in place, governance leads, risk and compliance, security, procurement and the owners of individual AI systems.

Inviting people

Administrators invite colleagues by email. Each invitation is tied to a **role** that determines what the person can see and do once they accept.

1. Open **Administration -> Users**.
2. Select **Invite user** and enter their email address.
3. Choose the **tenant role** appropriate to their responsibilities (see below).
4. Send the invitation. The recipient receives an email with a secure link to join.

Choosing a tenant role

The tenant role sets what someone can do across the organisation. Assign the least privilege needed for their job:

Role	Use it for
Administrator	People who manage users, SSO, plans and the admin console.
Advanced	Governance, risk and security staff who need all modules, AI analysis, the agentic stack and exports.
Standard	Contributors who need core frameworks and limited AI, without control testing, workpapers or the agentic stack.
Subscriber	Login only, no product access (useful while access is being arranged).

What a role can actually use is also capped by the workspace's plan: a user needs both the role permission and the plan entitlement.

Scoping access to specific assessments

Beyond the tenant role, you can add someone to an individual assessment with an **additive workspace role**, so they get broad read access organisation-wide but write access only where they work:

- **Lead Assessor**: full read/write on the assessment, can share, delete, approve and export.
- **Contributor**: create and update records, no delete or export.
- **Reviewer**: read plus notes and comments.
- **Viewer**: read-only.

This is the practical way to bring in a system owner, an auditor or a reviewer without granting them broad authority. See Users & roles for the full model.

A starting allocation

- **System / model owners**: Standard tenant role, added as Contributor on the assessments covering their systems. They also get read/update on objects they own automatically.
- **Governance, risk and compliance**: Advanced tenant role, Lead Assessor where they run the work.
- **Reviewers and auditors**: Standard or Subscriber tenant role, Reviewer or Viewer on the relevant assessments.
- **Administrators**: Administrator tenant role.

Single sign-on

For larger organisations, connect your identity provider so people sign in with corporate credentials and access follows your existing joiner / mover / leaver process. SSO governs authentication; roles still govern authorisation. See Single sign-on (SSO).

Accountability

Granting or revoking a role is an audited action, and every state-changing action in the workspace is recorded in the audit log. Suspending a user immediately revokes their access and ends their active sessions.

Next

Part 2: Core governance workflow

This part compiles the current public operating guidance for the workflows in scope.

8. Launch Center and Learning Hub

Use guided readiness and learning resources to move from account setup to repeatable governance operations.

The **Learning Hub** provides role-appropriate guidance. The **Launch Center** provides an operational readiness view for authorised platform operators. They support adoption but do not replace the governance records, approvals and evidence held elsewhere in Gamut.

Learning Hub

Use the Learning Hub when onboarding administrators, assessors, reviewers and agentic governance teams. Complete the recommended path for the work the person will perform, then validate learning through supervised use of a non-production or demonstration workspace.

Learning completion is not evidence that a control operates. Where competence is a control requirement, retain the organisation's approved training, role assignment and practical assurance evidence.

Launch Center

Launch Center is visible only to appropriately authorised platform operations users. It summarises readiness and directs operators to incomplete setup or control work. It should be treated as a navigation and readiness aid, not as a bypass around tenant administration, entitlement, MFA or approval requirements.

When an item is incomplete:

1. Open the identified area.
2. Confirm the accountable owner and required evidence.
3. Complete the underlying configuration or governance decision.
4. Return to verify that readiness reflects the saved state.

Do not publish screenshots or operational configuration from Launch Center into customer-facing evidence unless it is necessary, appropriately redacted and approved.

Recommended onboarding sequence

1. Understand tenancy, roles and plan entitlements.
2. Register a system and complete intake.
3. Review routing, ACRS and the assessment plan.
4. Perform a framework assessment with evidence.
5. Record risks, findings and remediation.
6. Generate a traceable report.
7. For agentic operation, complete the dedicated Agentic CISO and Gateway path.

Next

9. AI System Records

Create and maintain the authoritative inventory record that connects intake, routing, assessments, evidence, risks, model cards and reporting.

An **AI System Record** is the stable identity for the system Gamut governs. Intake describes a particular use and determines routing; the system record identifies the product, service, model or agentic workflow to which that work belongs. Keeping those responsibilities separate prevents an assessment result from becoming detached from its owner, version or operating context.

What to record

Complete enough detail for another reviewer to identify the same system without relying on local knowledge:

Area	Record
Identity	Name, unique identifier, version and lifecycle state.
Accountability	Business owner, technical owner, responsible-AI contact and operating department.
Purpose	Intended use, users, affected people and decisions or actions supported.
Technology	Model or AI type, supplier, deployment arrangement, environment and agentic capability.
Data	Sources, classifications, personal-data involvement and geographic exposure.
Operation	Autonomy, human oversight, integrations, access and dependency.
Governance	Registration status, risk information, review dates and linked assurance records.

Do not use a marketing product name alone where several deployments, versions or purposes have different risks. Create boundaries that match how change, ownership, access and evidence are managed.

Lifecycle

1. **Create** the record as soon as a proposed or discovered system enters governance.
2. **Complete** identity, ownership and operating context before relying on routing.
3. **Open or link intake** for the use case being approved.
4. **Assess** the system through the routed framework modules.
5. **Maintain** version, supplier, data, purpose, access and deployment changes.
6. **Reassess** when a material change invalidates prior scope or evidence.
7. **Retire** the record without deleting the history needed for accountability.

Relationship to intake

The system record and intake are connected but are not interchangeable. The record supplies durable facts such as system identity and ownership. Intake supplies the decision context used for risk, legal routing, framework applicability and assurance depth. Where the two disagree, investigate and correct the source record rather than selecting whichever result is more convenient.

Use **Sync system record** when the intake screen identifies that the linked record needs updating. Confirm the saved state before leaving the page. A system with incomplete intake can remain in the inventory, but its routing should be treated as provisional or incomplete.

Downstream effects

The selected system scopes framework assessments, AI-assisted assessment outputs, risks, model cards, evidence, findings, reports and reassessment history. Renaming a system should not create a new governance identity. A genuinely different product, materially different deployment or independently governed use may require a separate record.

Review checklist

- [ ] The record identifies one defensible system boundary.

- [ ] Named owners are current and able to act.
- [ ] Purpose, users and affected people are specific.
- [ ] Data, geography, supplier and deployment facts are current.
- [ ] Autonomy and human oversight describe actual operation.
- [ ] The correct intake and assessments are linked.
- [ ] Material changes have triggered reassessment.
- [ ] Retired systems retain the required audit history.

Next

10. Registry & Discovery

Maintain a standing inventory of AI systems and surface shadow AI across the organisation through a governed discovery pipeline of sources, rules, candidates and reconciled artifacts.

The **Registry** is your standing inventory of AI systems. **Discovery** is how you find AI that is in use but not yet registered, so the inventory stays honest.

The Registry

The Registry holds every AI system the organisation governs. Each record is structured, with the fields governance and audit actually need:

- **Identity and ownership**: name, system ID, version, owner, owner email, department, and a responsible-AI contact.
- **Nature of the system**: AI technique, model type, autonomy level, human-oversight model, vendor, deployment type and environment.
- **Data and exposure**: data classification, personal-data involvement and detail, data sources, users and affected persons, geographies and regulatory exposure.
- **Governance state**: risk tier, conformity status, registration status, ACRS score, and review dates (deployment, last review, next review).
- **Lifecycle stage**: from `development` through in-use to retired.

The Registry is the anchor of the platform: assessments, evidence, findings, model cards and reports all attach to a registry record. See Register your first AI system for the steps.

Discovery

The hardest part of AI governance is knowing what exists. Discovery is a governed pipeline for surfacing AI systems, GenAI tools and emerging agentic workflows that are in use but not yet registered. It is built from four connected object types:

Object	What it is
Sources	**Where signals come from:** connector-based collectors or manual imports, with a cadence, owner and run history.
Rules	Normalisation rules that map raw signals to a canonical app and vendor by domain or pattern, and tag sanction status and risk level.
Candidates	Suspected AI in use, each with a signal type, a detector, a confidence level, guessed owner/vendor/environment, and a review decision.
Artifacts	The underlying evidence (usage signals, code references, endpoints, model references) with a fingerprint and a reconciliation status.

A discovery **run** executes a source, produces candidates and artifacts, and records a summary and count. Candidates move through a review (`new` to a decision), and artifacts carry a **reconciliation status** so each piece of evidence is either tied to a known asset or flagged as unreconciled.

From discovery to registry

Reviewed candidates are promoted into the Registry, where they go through intake and risk tiering like any other system. Discovery finds the shadow AI; intake classifies and routes it; the Registry holds it. Nothing is governed until it is a registry record, and discovery is how things get there.

> **Note:** Discovery is an enterprise capability. Availability depends on your plan & entitlements.

Setting up discovery

1. Open **Discovery** and add a **discovery source** (for example a cloud or identity signal feed, or a manual import). Give it a name and type, and provide the import or connection detail it needs.
2. Optionally set a **schedule** (cadence) so the source is scanned automatically; otherwise run it on demand. A scheduled source shows its next and last run and overdue status.
3. Review the **candidates** the scan surfaces in the Discovery Inbox: each carries a source, confidence and rationale.
4. For a real, in-use system, **promote it to the Registry** (which starts the standard governance workflow); otherwise dismiss it with a logged reason.
5. In the **Registry**, complete each system record: identity, ownership, nature, data and exposure, and governance fields.

This is what feeds intake: a current registry makes intake and everything downstream complete.

Why this matters

An inventory that misses systems produces governance that looks complete but is not. Registry and Discovery together give you an inventory you can defend: comprehensive, owned, current, and with a documented trail showing how each system was found and brought under governance.

Next

11. Discovery operations

Operate AI Discovery from source coverage and rules through candidates, alerts, reconciliation, promotion and assurance reporting.

Discovery helps identify AI that may not yet be registered. It produces **signals requiring human review**, not automatic declarations that a person or team is using an unauthorised system.

Operating model

Stage	Question
Coverage	Which business areas, identities, cloud services, repositories or imports are observed?
Detection	Which built-in or approved custom rules interpret the signals?
Candidate review	Is the signal a real AI system, duplicate, permitted tool or false positive?
Reconciliation	Does it match an existing system, supplier, model or prior candidate?
Governance action	Should it be promoted, linked, dismissed, monitored or escalated?
Assurance	Can coverage, decisions and unresolved gaps be explained to management or audit?

Sources and credentials

Define the source owner, purpose, scope, cadence and lawful basis before collection. Use the smallest read scope that can produce the required signal. Store connection credentials through the approved tenant-scoped integration mechanism; do not paste credentials into candidate notes or imports.

Test a source before relying on its coverage. A healthy connection does not prove full coverage if important accounts, regions, repositories or environments are excluded.

Rules and scans

Built-in and custom rules normalise raw indicators into reviewable candidates. Document what a rule detects, expected false positives, risk labels and the owner who can change it. Preview an import or rule change before a live scan where that option is available.

Monitor last run, next run, errors, volume and unusual changes. A sudden zero may indicate a clean environment, but it can also indicate failed collection. A sudden spike may reflect a deployment, rule change or duplicate ingestion.

Candidate and alert decisions

For each candidate:

1. Inspect source, detector, confidence, rationale and supporting artifacts.
2. Search for an existing system or related candidate.
3. Confirm owner and intended use with an accountable person.
4. Promote a real system into the Registry or link it to the matching record.
5. Dismiss a false positive with a reason that another reviewer can understand.
6. Escalate material unknown use, prohibited tools or sensitive-data exposure.

Alerts should have a named disposition and owner. Bulk actions are appropriate only when the same decision basis genuinely applies to every selected item.

Reconciliation and playbooks

Reconciliation prevents duplicate inventory and preserves the chain from raw artifact to governed record. Suggested matches assist review but do not replace confirmation. Discovery playbooks can create or link follow-up risks and evidence work; confirm the target system and workspace before execution.

Assurance reporting

A defensible Discovery report explains sources, coverage, run health, rule set, candidates, decisions, unresolved artifacts, overdue alerts and limitations. It must not imply that "no candidates" means "no shadow AI" unless coverage and detector effectiveness support that conclusion.

Review checklist

- [ ] Sources have owners, purpose, scope and review dates.
- [ ] Credentials use least privilege and are not exposed in records.
- [ ] Rules are tested and false-positive assumptions documented.
- [ ] Failed and overdue scans are investigated.
- [ ] Candidates have attributable decisions.
- [ ] Promotions enter normal intake and routing.
- [ ] Unresolved artifacts and coverage gaps are reported.

Next

12. Intake & risk tiering

Capture the context of an AI system, classify its risk through structured signals and an ACRS capability score, and route the right frameworks and reviews to it.

Intake captures the context an AI system needs to be governed. **Risk tiering** turns that context into a structured classification that helps decide how much governance the system needs. Authoritative routing combines Intake with the linked System Record to determine which frameworks need assessment or screening. Together they form the front of the governance lifecycle.

What intake captures

Intake is a structured record, not a free-text form. Alongside the descriptive context, purpose, users, owner, business unit, vendor, deployment type and environment, data sources and geographies, it captures a set of **risk signals** as explicit flags:

- **Personal data** and **special-category data** involvement.
- **High-risk** use and **automated decision-making**.
- **Public-facing** exposure.
- **Retrieval characteristics: RAG**, retrieval sources, **vector store**, **long context** and **external retrieval**.
- Context-specific signals such as **cultural significance**, **heritage relevance**, **local language significance**, **creative-sector** use and **community impact**.
- Cross-framework routing facts covering **decision impact**, **action capability**, **model origin**, **organisation roles**, material capabilities and **jurisdictions**.

These are routing inputs. The linked System Record supplies the standing system facts, while Intake supplies use-case and impact context. See Routing & applicability for the complete decision model.

From intake to risk tier

Intake produces a **risk tier** (a system carries `undetermined` until classified, then low through critical) and an **initial risk rating**. The risk tier is the pivot of the whole lifecycle: it prioritises attention and routes the system to the controls and frameworks that matter for its level of risk.

- Higher-risk systems attract deeper assessment and more demanding control expectations.
- Lower-risk systems are governed proportionately, without unnecessary overhead.

How signals become a tier is deterministic: a governance weighting profile of weighted dimensions and configurable thresholds does the mapping, so the same inputs always yield the same tier.

The ACRS score and confirmation

For systems with agentic or capability-driven risk, intake derives an **ACRS score** and a capability **band** from the signals captured. Each dimension is seeded by inference from the intake facts and can be refined with explicit assessor evidence, so an intake-anchored capability score exists straight away. Gamut reconciles the system's assigned risk tier against the ACRS band, so a mismatch between "how risky we called it" and "how capable it actually is" is surfaced rather than hidden.

You can open ACRS, complete the dimension assessment, and move on to the assessment plan without a mandatory gate. **Confirmation is an optional, audited sign-off**: it records an assessor-signed ACRS for the system and strengthens the basis for assurance depth, cadence and escalation. It is not required for GTSAF factual applicability: a complete, conflict-free authoritative Intake route is what makes the system selectable for a scoped GTSAF assessment.

If routing-critical facts change after confirmation, the route becomes **Reassessment required**. Review the changed facts, reconsider ACRS where necessary, and reconfirm only after the resulting framework routes and assessment scope have been reviewed.

Framework routing

Intake does not route in isolation. Gamut combines it with the linked System Record and produces:

- **Applicable frameworks**: frameworks positively triggered by current facts.
- **Screening-required frameworks**: frameworks that cannot safely be ruled out while facts remain unknown.
- **Organisational review**: a decision, such as ISO/IEC 42001 AIMS scope, that is not a simple system-level exclusion.
- **Required reviews**: the specific reviews it must undergo.
- A reviewable routing basis, quality gaps and any conflicts between the records.

This is what connects intake to the rest of the platform: a routed system arrives at assessment already pointed at the right frameworks.

Unknown material facts fail closed as **screening required**. They do not silently remove a framework. A confirmed route becomes **reassessment required** when its material basis changes.

Approval

An intake record carries an approval status (`draft` through approved) with a named approver. Risk classification is therefore an accountable decision, not an automatic one: a person signs off that the system is classified and routed correctly before it moves into deeper governance work.

Field groups and decision ownership

Group	Examples	Primary effect
Identity and purpose	System, use case, owner, business unit, intended and prohibited use	Establishes the assessed boundary and accountability.
People and decisions	Users, affected people, decision impact, appeal and human oversight	Risk, impact and legal routing.
Data and retrieval	Sources, classification, personal data, RAG, external retrieval and retention	Privacy, security, supplier and impact controls.
Technology and supply	Model origin, vendor, deployment, integrations and dependencies	Threat, supplier, model and resilience routes.
Action and access	Autonomy, tools, permissions, targets and reversibility	ACRS, agentic and assurance depth.
Geography and regulation	Deployment, affected-person and market geography, regulated roles and sector	Jurisdictional and legal screening.
Lifecycle	Status, dates, review triggers and approver	Approval, reassessment and reporting.

The person completing intake should obtain facts from system owners, technical teams, data/privacy, security, legal and affected business functions as needed. "Unknown" is valid while investigating; it is not a convenient negative answer.

Saving and synchronising

The screen shows save or autosave status. Wait for a successful saved state before navigating away. If save fails, preserve the entered information, correct the displayed validation or service issue and retry. **Sync system record** updates the connected inventory record where supported; verify the result rather than assuming the two records now agree.

When authoritative facts are incomplete

Incomplete material facts produce a routing-incomplete or screening-required state. Gamut may show conservative starting values, but those are not a confirmed factual route. Complete the named facts, resolve contradictions, review ACRS and then confirm the route. Other descriptive intake fields remain valuable governance records even when they do not independently determine applicability.

Worked routing example

A marketing-content generator and a fraud-monitoring system are both registered AI systems, so both receive baseline governance. The marketing system may trigger supplier, retrieval and content controls while leaving many decision, privileged-access and severe-harm controls out of scope. The fraud monitor may process personal data, influence consequential decisions, depend on live systems and require stronger evidence, testing and review. ACRS and the Governance Weighting Profile can increase depth and escalation; they do not invent or remove the factual triggers.

Intake quality checklist

- [ ] System boundary and owner confirmed.
- [ ] Purpose, users, affected people and decision impact are specific.
- [ ] Data, retrieval, supplier, action and access facts reflect production reality.
- [ ] Geographic and regulated-role facts have an evidence basis.
- [ ] Unknowns and contradictions are resolved or explicitly escalated.
- [ ] Save and system-record synchronisation succeeded.
- [ ] ACRS, route, applicability and review streams were checked after changes.
- [ ] Approval and reassessment triggers are recorded.

Why consistency matters

The value of risk tiering is consistency. When every system is classified against the same signals and the same routing logic, risk decisions become comparable across the organisation and defensible to reviewers, who can trace each classification back to the intake that produced it. This is also the foundation for EU AI Act readiness, which is itself a risk-tiered regime.

> **Note:** Intake, risk and assessment capabilities vary by plan, volume and role. The current workspace plan and effective entitlements determine availability.

Next

13. Assessments & control testing

Score AI systems against framework controls and evidence that those controls are designed and operating effectively, with rationale, samples and traceable results.

An **assessment** evaluates its defined subject against a framework. Most subjects are AI systems; ISO/IEC 42001 uses an organisation-level AIMS scope and ATF uses a registered agent. **Control testing** goes a step further: it evidences that controls are not just designed, but operating effectively.

Assessments

An assessment connects an AI system to a framework and works through its controls, domain by domain. For each control you record:

- The **current state**, how well the control is met.
- The **rationale**, why you reached that conclusion.
- **Evidence** links, the proof that supports the conclusion.

Because an assessment records rationale and evidence per control, the result is defensible rather than just a score. See Run your first assessment for the steps.

Assessing against multiple frameworks

The same AI system can be assessed against more than one framework, for example GTSAF for depth and the EU AI Act for regulatory readiness. Gamut keeps each assessment distinct while sharing the underlying system, evidence and findings, so work done once is reused across frameworks rather than repeated. See Frameworks overview for how routing selects frameworks.

Running an assessment

1. Make sure the AI system is registered and has been through intake and risk tiering, so it is routed to the right frameworks.
2. Open the routed framework (for example GTSAF or the EU AI Act) for that system.
3. Work through the controls, scoring each on the framework's answer scale and recording your rationale. Each control's auditor advisory explains what to assess and how to score it.
4. Attach or request evidence, run control tests, and raise findings for gaps.
5. Roll the result up with reporting.

How scoring works

Gamut applies a consistent scoring discipline across every framework, so an assessment score is honest about both quality and completeness.

- **Each framework uses its own answer scale.** GTSAF uses Yes / No / N/A; ATF uses Fully met / Partially met / Not addressed; the EU AI Act and NAGF use Compliant / Non-compliant / N/A with a depth rating; ISO/IEC 42001 separates conformity from assurance; ISO/IEC 42005 uses maturity plus assurance; MAESTRO scores likelihood and impact. See Frameworks overview.
- **Scope-correct, with rollup.** System assessments remain bound to one system, ATF to one agent and ISO/IEC 42001 to the approved AIMS scope. Portfolio views do not transfer evidence between those subjects.
- **Applicability is separate from depth.** For GTSAF, system facts determine which conditional controls apply; ACRS and the governance weighting profile determine proportionate assessment rigour. A higher applicable count is not automatically a higher-risk conclusion.

- **Coverage-inclusive.** Controls you have not assessed count as not-yet-met, so a partly-finished assessment cannot report a high score.
- **Justified N/A only.** Marking an applicable control N/A requires a documented justification; a justified N/A leaves scope, while an unjustified one stays in scope and counts as not met.
- **Auditor advisory on every control.** Each control carries built-in guidance on what to assess, what evidence to expect, how to test it and how to score it, so assessors apply the scale consistently.

Control testing

Designing a control is not the same as operating it. A control test in Gamut captures the full testing record auditors expect:

- **Test objective** and **test procedure**, what was tested and how.
- **Sample method** and **sample size**, the basis for the conclusion.
- A **design effectiveness** score and an **operating effectiveness** score, kept separate, because a well-designed control can still fail in operation.
- A **test result**: `pass`, `partial`, `fail` or `not_tested`.
- **Exceptions**: a count and a summary of what failed.
- **Tester** and **reviewer**, test date and next test date, and **evidence references**.

This turns an assertion ("we review model outputs") into proof ("here are the reviews, sampled this way, on this cadence, by these people, with these exceptions"). Control tests produce the artefacts that back an assessment and that auditors rely on when they test your governance.

> **Note:** Control testing and workpaper-grade outputs are advanced capabilities. Availability depends on your plan & entitlements.

Findings

Where an assessment or control test reveals a gap, raise a **finding**. A finding carries a severity, a root cause, a recommendation and a management response, and it links back to the control, the test and the system it relates to, then tracks through to remediation and closure. See Evidence & findings.

AI assistance

Gamut can assist assessment work with AI, for example by helping analyse context or draft narrative. All AI analysis is proxied server-side, so model provider keys are never exposed to the browser. See AI assistance & data handling.

Next

14. Routing & applicability

How Gamut combines System Records and Use Case Intake into an authoritative, reviewable cross-framework route, handles unknowns and conflicts, and triggers reassessment after change.

Routing answers a narrower question than assessment:

> Given the approved facts about this AI system, which assessment frameworks require work, which require a scope review, and which can presently be treated as outside the route?

It does **not** decide that a control is satisfied, that a legal obligation has been met, or that a system is safe. It creates the defensible starting scope for those decisions.

The two records used

Gamut combines:

- The **System Record**, which is the standing description of the system, its ownership, technology, lifecycle, deployment, data and operating context.
- The **Use Case Intake**, which records why and how the system is being used, the people and decisions it affects, its action capability, jurisdictions and assessment-specific risk facts.

The records are linked by the system reference. Similar names are not treated as proof that two records describe the same system.

Where the same standing fact appears in both records, the System Record is treated as the primary record and a disagreement is surfaced for review. A conflict is not silently resolved in favour of the more convenient route.

Authoritative cross-framework routing facts

The structured routing section in Intake captures facts that materially change applicability:

Fact group	What to record	Why it matters
Decision impact	Advisory, operational, consequential, or legally significant	Helps identify impact-assessment, legal and assurance depth
Action capability	Read-only, recommend, write, execute, or administer	Identifies agentic and privileged-action exposure
Model origin	Internal, third-party, open-source/open-weight, or hybrid	Informs supplier, provider and lifecycle responsibilities
Organisation roles	For example provider, deployer, importer or downstream provider	Legal and value-chain duties depend on role
Capability flags	Privileged access, training activity and supplier dependency	Routes controls for access, change, data and third parties
Human-impact flags	Vulnerable populations, biometrics and emotion recognition	Triggers enhanced legal, rights and impact screening
Service and content flags	Critical service, synthetic content and deepfakes	Routes resilience, transparency and content obligations
Jurisdictions	Markets, users, affected people and output destinations	Establishes territorial and regulatory nexus

Use **Unknown** when the answer has not been established. Do not select "No" merely because evidence has not yet been obtained.

Information needed for a reliable route

A reliable route normally requires:

- A stable link to the System Record.
- System name, purpose and use-case boundary.
- Model type or AI technique.
- Autonomy and human oversight.
- Deployment context.
- Users and affected people.
- Data sources and classification.
- Relevant geographies.
- Decision impact and action capability.

Missing information remains visible as a routing-quality gap.

Routing states

Each framework receives one of four states:

State	Meaning	Assessor action
Applicable	Current facts positively trigger the framework or it is a universal baseline	Include it in the assessment plan
Screening required	Applicability cannot safely be ruled out because a material fact is unknown or needs specialist review	Complete the missing facts or scope review; do not treat it as excluded
Not applicable	Current facts provide a documented basis for leaving the framework outside the system route	Retain the basis and review it after change

State	Meaning	Assessor action
Organisational review	Applicability is decided at organisation or management-system scope, not solely for one system	Resolve it through the appropriate organisational governance process

This is fail-closed scoping: uncertainty creates work; it does not create an exemption.

Overall routing status

The route itself carries a status:

Status	Meaning
Incomplete	Routing-critical information is missing
Conflict review	The linked records disagree on one or more material facts
Provisional	A route has been calculated but not confirmed by an accountable reviewer
Confirmed	The current basis has been reviewed and accepted
Reassessment required	Material facts changed after the previous confirmation

Confirmation applies only to the facts and route visible at that time.

What to do when reassessment is required

A route panel may remain green because the current facts are complete and conflict-free while also showing an amber reassessment warning. These messages answer different questions:

- **Green route panel:** the current route is calculable and has no missing-fact or conflict blocker.
- **Reassessment required:** the basis changed after the previous confirmation, so accountable review is required.

Review the changed System Record and Intake facts, reconsider the ACRS dimensions, inspect any changed framework routes and control scope, then approve the current ACRS position or adjust it where the evidence requires. Revisit affected assessments and reconfirm only when the current basis is understood.

How each framework is treated

Framework	Routing approach
GTSAF	Universal framework baseline; complete system facts determine per-control applicability, while ACRS and governance weighting adjust assurance depth
ACRS	Universal capability-risk baseline for the system; informs proportionate assurance depth
NIST AI RMF	Available as a universal risk-management baseline; context determines prioritisation
ISO/IEC 42001	Organisation-level AIMS scope decision informed by the AI portfolio
ISO/IEC 42005	Triggered by actual or foreseeable impacts, affected parties and sensitive uses; incomplete impact facts require screening
EU AI Act	Routed from territorial nexus, role, use and regulated characteristics; the module then performs authoritative legal classification
ATF	Routed when the system can act, use tools or resources, orchestrate work, or exercise delegated authority
NAGF	Routed when markets, users, affected people, outputs, operations or regulated roles establish a Nigerian nexus
MAESTRO	Routed when architecture, model, data, tools, orchestration, ecosystem or human interaction creates layer-specific exposure

Framework-level routing is deliberately separate from item-level applicability. Being routed to a framework does not mean every item within it applies.

Confirmation and change

Review the route before confirming it. Confirm that:

- The linked System Record is correct.

- Unknowns have been investigated or deliberately left for screening.
- Conflicts are resolved.
- The stated organisation roles are accurate.
- Jurisdictions cover deployment, users, affected people and outputs.
- Decision impact and action capability reflect real operation.
- The required assessment plan is proportionate.

When a routing-critical fact changes, Gamut marks the previous route for reassessment. Examples include:

- A system moves from recommendation to execution.
- A new market or affected population is introduced.
- A new supplier, model or training activity is added.
- Biometric, emotion-recognition or synthetic-content functionality is enabled.
- The system enters a critical service or consequential decision process.
- The organisation's legal role changes.

Reconfirm the route only after the changed basis and its assessment consequences have been reviewed.

Effect on assessment and scoring

Routing determines what work is presented and how it is prioritised. It does not award a favourable score.

- Unknown applicability is not counted as a pass.
- A route does not satisfy a requirement.
- A mapped control does not automatically satisfy another framework.
- GTSAF conditional controls are included or excluded from complete, conflict-free system facts; incomplete facts remain pending.
- ACRS and governance weighting may increase required assurance depth but cannot invent or remove factual control applicability.
- Legal modules perform their own detailed scope, role and requirement decisions after routing.

Understanding the GTSAF route

GTSAF distinguishes:

- **Applicability breadth:** the controls factually relevant to the selected system.
- **Assurance depth:** Baseline, Triggered or Enhanced rigour for each applicable control.
- **Assessment result:** what the answers, evidence and tests demonstrate.

The applicable count is not a risk ranking. A system using an external component may have more supplier controls than a higher-impact internal system, while the higher-impact system still requires much deeper assurance. GTSAF therefore also shows depth-weighted workload and normalised assurance intensity. See GTSAF system scope, applicability and assurance depth.

AI Assist and routing facts

In full-context mode, AI Assist may consider the selected system's current route, routing facts, quality gaps and conflicts together with the authorised assessment record.

In **Privacy Mode**, direct identifiers and free-text routing facts are excluded. The model receives only the minimum structured route, applicability signals, completeness state and assessment values needed for bounded assistance. Privacy Mode reduces disclosure; it does not change the route.

AI Assist cannot confirm a route, approve an exclusion, alter applicability or accept risk.

Reviewer checklist

- [ ] System Record and Intake are linked to the same system.

- [ ] Purpose and boundary are specific.
- [ ] Decision impact and action capability describe production reality.
- [ ] Organisation roles are complete.
- [ ] All four jurisdiction perspectives have been considered.
- [ ] Unknowns are not disguised as negative answers.
- [ ] Conflicts have been resolved.
- [ ] Screening-required frameworks have an owner and next action.
- [ ] Organisational decisions are not misrepresented as system-level exclusions.
- [ ] Route confirmation reflects the current basis.
- [ ] Material changes trigger reassessment.

Next

15. Evidence & findings

Capture governance evidence as work happens through request-based workflows, track its quality, and manage findings through root cause, remediation and validated closure.

Evidence and findings are where governance stops being a claim and becomes something you can prove. Gamut treats both as first-class objects, captured as work happens rather than gathered in a rush at audit time.

Evidence

Evidence is any artefact that supports a governance claim, a document, export, log, screenshot or record. In Gamut, evidence is managed as a request-based workflow rather than a folder of files. An **evidence request** carries:

- A **request title** and description, tied to a specific control.
- The **expected evidence type**, so the owner knows what will satisfy it.
- A named **owner** and **due date**.
- A **status** as it moves from `requested` to fulfilled.
- A **quality rating** and **review notes**, so weak or missing evidence is visible rather than assumed.

Linking evidence to controls is what creates traceability: a reviewer can move from a control to the evidence behind it without hunting through shared drives. And because requests are tied to controls and owners, the evidence base builds continuously as governance work happens, instead of being chased at the end of a cycle.

Findings

A **finding** records a gap, deficiency or exception identified during assessment, control testing or audit. Each finding is a structured record:

- A **title**, **description** and **finding category**.
- A **severity**: `low`, `medium`, `high` or `critical`.
- A **root cause** and a **recommendation**.
- A **management response**, a named **owner** and a **target date**.
- A **status** as it moves toward closure, with **closure notes**, a **validated by** and a **validated date** so closure is verified, not just asserted.
- Links to the **control test** and **evidence request** it arose from.

Recording findings honestly is more valuable than an unsupported pass: it gives leadership a true picture and gives reviewers confidence that governance is real.

Remediation

Remediation is the closure path for a finding. Keeping remediation visible, with an owner, a target date and a validated closure, turns findings from a list of problems into a managed improvement programme, and gives the Improve stage of the lifecycle something concrete to track over time.

Traceability

Evidence and findings complete the chain that makes Gamut defensible:

```
control -> control test -> evidence request -> finding -> remediation -> validated closure
```

Any conclusion in a report can be followed back through this chain to the underlying records, and every change along it is captured in the audit log.

Setting up evidence, tests and findings

1. From a scored control in an assessment, **request evidence**: set the control reference, what is needed, an owner and a due date. Requests track from open to satisfied.
2. **Attach evidence** to the request, or record it directly against the control.
3. **Run a control test** where you need to prove a control operates: capture the objective, procedure, sample method and size, the design and operating effectiveness scores, the result, and any exceptions.
4. Where there is a gap, **raise a finding**: severity, root cause, recommendation and management response, linked back to the control, test and system.
5. Convert a finding into a remediation item and track it to closure.

Owners and due dates make evidence and remediation overdue-trackable, and every step is captured in the audit log.

Next

16. System selection in assessments

Understand which systems appear in framework selectors, why a selection may be unavailable and how system-scoped results are kept separate.

Framework assessments are **system-scoped**. The system selector determines whose intake facts, routing decision, scores, evidence and AI Assist output the module loads. Selecting the wrong system can produce a well-formed but irrelevant assessment, so always verify the scope banner before scoring.

What selector states mean

State	Meaning	Action
Selectable	The system has the minimum confirmed records required by that module.	Select it and verify the displayed name and scope.
Greyed out	The system exists but a prerequisite, route or confirmation is incomplete.	Follow the displayed reason back to intake, ACRS or routing.
Not listed	The record is outside the current workspace, unavailable to the user or not eligible for the module.	Verify workspace, role, entitlement and system status.
All controls / framework overview	No system-specific record is loaded.	Use for catalogue review only; do not mistake it for a completed system assessment.

Some frameworks have additional legitimate prerequisites. ACRS confirmation can be required before GTSAF depth is final. ISO/IEC modules require an applicable licence confirmation. ATF is assessed per agent rather than as a generic system catalogue. Legal frameworks may require their authoritative scope gate before a conclusion can be confirmed.

Safe workflow

1. Complete and save the AI System Record.
2. Complete intake facts and resolve routing-incomplete prompts.
3. Review and, where required, approve the ACRS score.

4 Open the routed assessment from the plan or use the framework selector.

5 Confirm that the system name, scope and assessment status have changed.

6 Score, add evidence and run AI Assist only after that confirmation.

If the selector returns to the framework overview, do not continue entering system evidence. Refresh the routing plan and verify the prerequisite records. Results already saved for another system remain associated with that system; they should not be copied merely to make the new assessment look complete.

AI Assist and system scope

AI Assist output is stored against the selected system or agent and assessment item. Switching scope must not present the previous system's output as current. The output header should identify the analysed system. If it does not, stop and re-establish scope before relying on the analysis.

Troubleshooting checklist

- Correct client, workspace and system selected.
- Intake save shows a successful state.
- Required authoritative facts are resolved.
- ACRS is confirmed where the route requires it.
- The framework appears in the server-authoritative plan.
- The user's role and plan both permit the framework.
- The record is active and has not been retired.

See Routing & applicability for how prerequisites affect the assessment plan without changing subscription entitlements.

17. Reporting & exports

Produce board-level and workpaper-grade outputs from Gamut through eight governed pack types, each composed of the right sections and carrying run-level traceability metadata.

Reporting turns the records in Gamut into outputs people act on, a board view of AI governance maturity, and workpaper-grade exports for audit and assurance. Because reports are generated from the connected records, every conclusion can be traced back to the evidence behind it.

What reports show

A report draws the lifecycle together into a single, practical view:

- **AI inventory**: the systems in scope, their owners and lifecycle status.
- **Risk**: the classification picture across systems.
- **Framework scores**: per-system conformance, compliance or maturity, rolled up to the workspace using the same coverage-inclusive scoring as the source assessment at generation time.
- **Evidence quality**: how well governance claims are supported.
- **Findings**: open deficiencies and exceptions.
- **Remediation progress**: what is being fixed and how far along it is.
- **Readiness priorities**: where to focus next.

Pack types

Rather than one generic report, Gamut produces purpose-built **packs**, each composed of a different set of sections for a different audience:

Pack	Built for
Full assessment	**The complete picture:** every section, for a deep review.
Framework focus	A single framework in depth (for example GTSAF, EU AI Act, ATF).

Pack	Built for
Executive	**Leadership:** dashboard, estate, framework, evidence, findings, decisions, roadmap.
Board pack	**Board oversight:** adds estate, agentic and gateway posture to the executive view.
Board summary	A concise board-level summary.
Evidence pack	**Evidence-centred:** controls, evidence, findings and supporting workflow.
Control testing	The control-testing record for assurance work.
Agentic snapshot	**The agentic posture:** agents, ATF, gateway decisions and runtime evidence.

Each pack selects only the sections relevant to its purpose, so a board pack is not a 200-page audit file and an evidence pack is not a one-page summary.

Generating a report

1. Open **Reporting** in the workspace whose records you want to report on.
2. Choose a **pack type** for your audience (full assessment, framework focus, executive, board pack, evidence pack, control testing, or agentic snapshot).
3. Set the **framework scope** (all frameworks or a single one) and, optionally, enable the **scores-only AI narrative**.
4. **Generate** the pack. Each run gets a run ID and traceability metadata.
5. **Export** it (subject to the `export.report` / `export.workpaper` permissions) for boards, clients or audit.

Because reports are built from connected records, regenerate one whenever the underlying assessment, evidence or findings change rather than maintaining a separate document.

Two audiences, one source

Reporting serves two audiences at once, from the same records:

- **Boards and leadership** get a concise picture of AI governance maturity, exposure and the decisions required of them, without engineering detail.
- **Auditors and reviewers** get traceable, workpaper-grade outputs that connect each conclusion back to its underlying evidence.

Because both views draw on the same connected records, they can be reconciled to the same source state and generation time. Different pack purposes may summarise that state differently, and any pack becomes historical when the source records subsequently change.

Traceability metadata

Every generated report carries run-level metadata, a run ID, the generation timestamp, the assessment, the framework scope and the pack type, so any export can be located, reproduced and defended. A report is not an anonymous PDF; it is a dated, identified artefact tied to the records it was built from.

Optional AI narrative

A report can optionally include an AI-generated narrative summary. To protect confidentiality, that narrative is generated from **assessment statistics only**, scores, counts, framework scope and posture figures, and never from free-text governance content. No confidential record text is sent to the model, and the generation is proxied server-side so model-provider keys are never exposed. See AI assistance & data handling.

Exports

Reports are produced as governed HTML and can be exported for distribution and record-keeping, for board packs, client assurance, internal audit and external review, subject to the `export.report` and `export.workpaper` permissions. Because the underlying records are connected, an export reflects the traceable source state at its recorded generation time. Regenerate it after material assessment, evidence, finding or remediation changes.

From point-in-time to continuous

Traditional governance produces a report once a year. Because Gamut keeps records connected and current, reporting becomes something you run whenever you need it, supporting the Improve stage with a view of progress over time, not just a single snapshot.

Next

18. Policy generation

Draft AI governance policies, standards, procedures and guidelines in Gamut with AI assistance, grounded in your systems, risk and frameworks, under a draft-to-publish lifecycle and usage quotas.

Gamut can help you **draft AI governance documents** with AI assistance, turning the governance work you have already done into written policy, rather than starting from a blank page.

A library of document types

Policy generation is not a single template. It spans the document set a real AI governance programme needs, across four kinds of artefact:

- **Policies**: the overarching AI governance policy (purpose, scope, principles, roles, structure, risk approach, enforcement, review cadence), an AI acceptable-use policy (permitted and prohibited uses, bring-your-own-AI and shadow-AI controls), and jurisdiction-specific policies such as a NITDA-aligned Nigeria AI governance policy.
- **Standards**: data classification, privacy, risk and transparency standards.
- **Procedures**: risk and impact assessment, incident response, change, lifecycle, monitoring, bias, data handling, vendor and decommissioning procedures, and NITDA audit procedures.
- **Guidelines**: ethics, human-oversight and training guidelines.

Each is drafted against a structured outline so the output is consistent and complete, not a blank prompt.

Grounded in your governance

Generated documents are grounded in the context Gamut already holds, your AI systems, risk tiers and the frameworks you assess against, so policy reflects your actual estate rather than a generic template. The output is a **starting draft** to review and adapt, not a finished policy. Human ownership and sign-off remain essential; generation removes the blank-page problem and keeps policy consistent with the rest of your governance.

Draft to publish

Generated documents follow a lifecycle governed by RBAC:

- `policy.generate` produces a draft.
- `policy.publish` promotes a reviewed draft to a published document.
- `policy.delete` removes one.

So drafting, reviewing and publishing are separate, accountable steps, and a generated draft is never a published policy until a person with the right permission promotes it.

AI handling and quotas

As with all AI features in Gamut, policy generation is proxied server-side: model provider keys are never exposed to the browser (see Security & data handling). It is gated by the `policy_generation` entitlement and metered by daily and monthly generation quotas that depend on your plan & entitlements. When the entitlement is off, the quota is zero.

Generating a policy document

1. Confirm the workspace has the records to ground the draft (system, intake, risk and assessment work). The more complete these are, the more specific the draft.
2. Open **Policy generation** and pick a **document type** from the library.
3. **Generate** the draft. It is produced server-side from your governance records; provider keys are never exposed to the browser, and the call counts against the plan's AI quota.

4. **Review and own** the draft: it is a starting point for a human to finalise, never an auto-published artefact.
5. Publish or store the finished document through your normal process.

Generation requires the `ai_chat` entitlement and AI permission; if the entitlement is off, the quota is zero. See plans & entitlements.

Why generate policy inside Gamut

- **Consistency.** Generated documents reflect the same frameworks and risk model your assessments use, so policy and practice stay aligned.
- **Speed.** Drafting accelerates from days to minutes, leaving more time for review and tailoring.
- **Traceability.** Policy connects to the governance records it governs, rather than living as a detached document.

Next

19. Model cards

Document the technical and ethical characteristics of the models behind your AI systems, intended use, data, oversight, autonomy and risk, linked to the system that uses them.

A **model card** documents a model's technical and ethical characteristics in one place. Where an AI system record describes a use of AI, a model card describes the model itself, and the two are linked.

What a model card captures

A model card is created against a workspace and linked to a system. It records:

- **Model name** and the **linked system** it belongs to.
- **Model type** and **provider**.
- **Deployment environment**.
- **Purpose**: what the model is for.
- **Data sources**: the data it relies on, at an appropriate level.
- **Human oversight** model and **autonomy level**.
- **Risk tier**: the model's own risk classification.

Because a model card can be created directly from a system record, its fields are pre-filled from the system, the model type, provider, environment, purpose, data sources, oversight and autonomy flow through, so the card stays consistent with the system that uses it rather than drifting.

Creating a model card

1. Open **Model cards** and create a card, linking it to the AI system that uses the model.
2. Record the model identity and type, intended use and limitations, and the model-level characteristics relevant to impact.
3. Set the card's status and keep it current as the system's intake context changes.
4. Reference the card from ISO/IEC 42005 impact assessment and from reports.

Why model cards matter

Model cards give reviewers and decision-makers a consistent, comparable view of the models the organisation depends on. They support:

- **Assessment**: grounding control conclusions in documented model characteristics.
- **Reporting**: giving boards and auditors a clear view of model-level risk.
- **Procurement**: comparing vendor models on a like-for-like basis.

Relationship to assessments and frameworks

Model cards complement assessments: the assessment evaluates a system against controls, while the model card documents the model that system uses. Several frameworks, including ISO/IEC 42005 and the EU AI Act, expect model and system documentation of this kind, and model cards help you meet those expectations. Model-level detail can also be pulled into reports.

Next

20. Governance weighting profile

The deterministic policy profile that drives risk tiering in Gamut, six weighted governance dimensions, configurable thresholds and floor rules, owned and calibrated by accountable stakeholders.

The **governance weighting profile** is the policy that turns intake signals into a risk tier. It is what makes risk tiering **deterministic and explainable**: the same inputs always produce the same tier, and the reasoning can be traced to a profile you control rather than a black box.

The six dimensions

A profile scores a system across six weighted governance dimensions. The defaults are:

Dimension	Default weight
Business criticality	0.22
Autonomy level	0.14
Human impact	weighted
Regulatory sensitivity	weighted
Data sensitivity	weighted
External exposure	weighted

Each dimension carries a weight, and the weighted combination produces a governance score for the system. Weights are configurable, so an organisation can express its own risk appetite, for example weighting regulatory sensitivity more heavily in a regulated sector.

Thresholds and tiers

The score is mapped to a tier through configurable thresholds. The defaults are:

- **Moderate** at 45
- **High** at 65
- **Critical** at 80

Below the moderate threshold a system is treated as lower risk. Raising or lowering a threshold changes how readily systems are classified into each tier, again, a deliberate policy choice, not a hidden constant.

Floor rules

Profiles support **floor rules** (enabled by default): hard minimums that force a system to at least a given tier when a specific high-risk signal is present, regardless of the weighted score. This prevents a genuinely sensitive system from scoring its way under the radar because other dimensions are low.

A profile is policy, and is owned

A weighting profile is treated as governance policy in its own right. It carries a name, an owner, an approver, a review date and a status. The shipped default is explicitly marked as "current policy until formally approved or calibrated by accountable stakeholders", the intent is that an organisation reviews and signs off its own profile rather than silently inheriting the default.

Configuring the profile

The weighting profile is a workspace setting you own and calibrate:

1. Open the **governance weighting profile** settings for the workspace.

2. **Set the dimension weights** to reflect your risk appetite (for example weight regulatory sensitivity higher in a regulated sector). What matters is the relative emphasis.
3. **Set the tier thresholds**, the score at which a system becomes Moderate, High or Critical.
4. **Enable or adjust the floor rules** that force a minimum tier when a specific high-risk signal is present.
5. **Record the policy metadata**: a profile name, an accountable **owner** and **approver**, a **review date** and a **status**. Until your own profile is approved, the shipped default applies and is marked as interim policy.

   Because the profile is deterministic, recalibrating it re-tiers every system consistently from the same inputs. Save the profile, then re-run intake to apply it.

Effect on applicability and assurance depth

The weighting profile affects:

- Governance tier.
- Required assurance depth.
- Review cadence.
- Escalation and approval expectations.

It does **not** add or remove a control solely because a weighted score is higher or lower. Factual control applicability remains grounded in the System Record and Use Case Intake: architecture, data, suppliers, users, impacts, jurisdictions and operating characteristics.

This separation prevents policy calibration from rewriting the system's factual control boundary. In GTSAF, a higher tier may move an already applicable control to Triggered or Enhanced depth without making an unrelated conditional control applicable.

Why determinism matters

Because tiering runs through an owned, weighted, threshold-based profile:

- **Decisions are consistent.** Two systems with the same profile inputs get the same tier.
- **Decisions are defensible.** A reviewer can see exactly which dimensions and thresholds drove a classification.
- **Policy is adjustable in one place.** Changing risk appetite means recalibrating the profile, not re-judging systems case by case.

Next

21. Risk Register

Record, own, treat and monitor AI risks from intake, assessment, discovery, findings and runtime governance.

The **Risk Register** turns identified exposure into an owned management decision. A framework score describes assessment posture; a risk record describes an uncertain event or condition, its impact, the treatment decision and who is accountable for it.

Sources of risk

Create or link risks from:

- intake and risk-tiering decisions;
- framework gaps or failed control tests;
- findings and evidence deficiencies;
- Discovery alerts and unregistered AI;
- supplier, model, data or regulatory change;
- agentic simulations, runtime denials and incidents;
- manual management or audit review.

Avoid one generic "AI risk" covering unrelated causes. A useful statement connects a cause, event and consequence, for example: "Because customer prompts may contain special-category data, sending them to an unapproved provider could cause unlawful disclosure and customer harm."

Minimum defensible record

Field	Purpose
Title and statement	Identifies the risk and causal scenario.
Linked system	Establishes scope and ownership boundary.
Category and source	Explains where the risk arose.
Inherent exposure	Records exposure before the claimed treatment.
Existing controls	Identifies what is already reducing likelihood or impact.
Residual exposure	Records the post-control judgement and its basis.
Treatment	Avoid, reduce, transfer or accept, with actions and evidence.
Owner and due date	Assigns accountable action and review timing.
Status	Tracks open, treating, accepted, closed or superseded state.
Review trigger	Defines when the judgement must be revisited.

Operating workflow

1. Link the risk to the correct system and source record.
2. Write a specific cause-event-consequence statement.
3. Assess exposure using the organisation's approved method.
4. Identify controls and evidence actually operating, not planned controls.
5. Select treatment, assign an owner and set dates.
6. Link remediation work, tests and findings.
7. Obtain the required acceptance or escalation decision.
8. Review after material change, failed treatment, incident or due date.

Closing a finding does not automatically eliminate its underlying risk. Close the risk only when the treatment and residual decision are supported, or mark it superseded when a replacement record preserves the history.

Governance Weighting Profile

The Governance Weighting Profile can change how the organisation prioritises review, escalation and assurance. It does not rewrite the factual system scope or erase a risk. Document any management judgement applied above the calculated result.

Quality checks

- The risk is system-specific and understandable without oral explanation.
- Inherent and residual judgements are not confused.
- Controls have linked evidence or tests.
- Acceptance is within the approver's authority and time-bounded where appropriate.
- Overdue actions and high residual risks are escalated.
- Changes in model, data, supplier, autonomy, geography or law trigger review.

Next

22. Remediation Roadmap

Track the closure of findings as sequenced, owned remediation items with priorities, target dates and overdue tracking, the managed improvement programme behind the Improve stage.

The **Remediation Roadmap** is where findings become a managed programme of work rather than a list of problems. Each gap raised during assessment, control testing or audit becomes a tracked remediation item with an owner, a priority and a target date, and the roadmap shows what is in progress, what is done and what is overdue.

What a remediation item holds

Field	Purpose
Title	What is being fixed.
Owner	Who is accountable for the fix.
Control reference & framework	The control and framework the item relates to.
Priority	How urgent the work is.
Status	Where the work has reached (through to done).
Target date	When it is due.
Description & notes	The detail and running commentary.

Remediation items link back to the findings they close, completing the traceable chain from a control gap to the work that resolves it.

Overdue tracking

The roadmap watches target dates. Any item that is not done and is past its target date is flagged as **overdue**, and the count surfaces as a warning in the interface so slippage is visible rather than buried. The roadmap also shows progress at a glance as a done-over-total ratio.

Where it sits in the lifecycle

Remediation is the closure path for findings and the concrete substance of the Improve stage of the governance lifecycle:

```
control -> control test -> evidence -> finding -> remediation item -> done
```

Keeping remediation visible, owned and dated turns governance from a point-in-time assessment into a continuous improvement programme, and gives reporting a real view of progress over time, not just a snapshot of open gaps.

The AI Consultant can help plan it

The AI Consultant's **Remediation Planning** mode can draft sequenced remediation actions with owners, dependencies and closure tests, grounded in the findings already recorded. As always, the output is a drafting accelerator for a well-prepared record, not a substitute for accountable sign-off.

Next

23. AI Consultant

An in-platform advisory tool grounded in your live assessment records, with task-specific modes for governance, risk, evidence, policy, reporting, agentic review, board summary and remediation planning.

The **AI Consultant** is an in-platform advisory tool that analyses and drafts from your **live assessment records**, the system context, framework scores, evidence entries, risk items and findings already recorded. It draws on what is in the workspace, not on general knowledge alone, so its output is specific to the governance situation in front of it.

Grounded, not general

Every consultation reads the current assessment at the moment of the query. That grounding is both the Consultant's strength and its limit:

- When the record is detailed, scored controls with notes, evidence linked to controls, named risk owners, a specific system description, the Consultant produces directly relevant analysis.
- When the record is thin, the output is generic, because there is nothing specific to draw from.

The practical consequence: the Consultant cannot compensate for an incomplete assessment. The screen shows a **Workspace context included** manifest with the current workspace and record counts for systems, intake, risks, evidence, tests, findings and agentic governance. Use that manifest to check what can ground the answer. A zero count or unavailable module is a limitation, even when the answer reads fluently. Treat the output as a drafting accelerator for well-prepared records, not as authoritative governance and not as a substitute for the work that produces the record.

Task-specific modes

The Consultant offers focused modes, each oriented to a specific governance activity. Matching the mode to the question produces sharper output than a general query:

Mode	What it produces
Governance Review	Ownership, accountability, approval gaps and operating-model weaknesses.
Risk Review	Risk statements, exposure summaries and treatment options from the assessment data.
Evidence Review	What evidence is present, what is missing, and which requests would strengthen assurance.
Policy Support	Generates or improves policy wording from the governance record.
Report Support	Report-ready narrative.
Agentic AI Review	**Scoped to agentic governance:** agent controls, approval gates, ATF readiness and Gateway records.
Board Summary	The assessment distilled into executive priorities and decisions needed.
Remediation Planning	Sequenced actions with owners, dependencies and closure tests.
General Advice	Open-ended questions, broader output than the task-specific modes.

The selected **output type** controls the intended artefact, for example Advisory Note, Risk Summary, Evidence Gap Review, Agentic AI Control Review, Remediation Plan, Board Briefing, Policy Improvement Note, Report Narrative or Executive Summary. Mode controls the analytical lens; output type controls the form of the draft.

Context and privacy boundary

The Consultant uses permitted workspace records, notes and other free text so that it can answer workspace-specific questions. The framework **scores-only privacy mode does not narrow Consultant context**. If scores-only mode is enabled, the Consultant displays a warning explaining this difference.

Before sending a question:

1. Check the **Workspace context included** manifest.
2. Confirm that the selected workspace is the intended one.
3. Do not paste secrets, credentials or unnecessary personal data into the question.
4. Do not use the Consultant if the permitted workspace context should not be sent to the configured model provider; use the relevant framework's scores-only AI Assist instead.
5. Review the provider and organisational data-handling terms that govern your deployment.

How it is governed

The Consultant is held to the same controls as every other AI feature in Gamut:

- **Server-side only.** Prompts and responses are proxied server-side; model provider keys are never exposed to the browser. See AI assistance & data handling.
- **Entitlement and permission gated.** It requires both the `ai_chat` entitlement and the AI Consultant permission, an Advanced-tier capability. See Plans & entitlements and Users & roles.
- **Usage-metered.** AI calls count against the plan's daily and monthly quotas.

- **Tenant-scoped.** It only ever sees the records of the workspace you are in.
- **Visible context boundary.** The screen identifies the categories and record counts available to the consultation before the question is sent.

Get the most from it

- **Complete the assessment first.** A half-finished assessment leaves the Consultant blind to large parts of the picture, and the output will sound more complete than it is.
- **Pick the right mode.** "What are the ownership gaps here?" belongs in Governance Review, not General Advice.
- **Treat output as a draft.** Review and own everything it produces before it becomes a governance artefact.

Next

24. Audit and assurance workflows

Operate Evidence Tracker, Testing Centre, Findings Register, Workpaper Packs, Board Dashboard and the Audit Trail as one traceable assurance chain.

Gamut separates six assurance records so evidence, testing, deficiencies, reporting and system activity are not collapsed into a single score.

Area	Primary question
Evidence Tracker	What proof has been requested, received, reviewed and linked?
Testing Centre	Did the control operate as designed for the sampled period and scope?
Findings Register	What deficiency, exception or uncertainty remains?
Workpaper Packs	Can another reviewer reproduce the assurance conclusion?
Board Dashboard	What exposure, trend and decision requires leadership attention?
Audit Trail	Who changed which governed record, and when?

Evidence Tracker

Define the exact claim, framework item, system, owner, requested artifact and period. Record source, version, receipt date, reviewer and sufficiency decision. An attachment alone is not sufficient: explain what it proves, what it does not prove and whether it is current.

Reject or qualify evidence that is generic, self-authored without corroboration, outside the assessment period, scoped to another system or contradicted by tests.

Testing Centre

A control test should identify objective, population, sample, procedure, expected result, actual result, exceptions and reviewer. Distinguish design adequacy from operating effectiveness. A policy can demonstrate design but usually cannot prove repeated operation.

Failed or inconclusive tests should create or update a finding and, where material, a risk. Retests must preserve the original result and show the corrective action assessed.

Findings Register

State the condition, expected criterion, cause, consequence, affected scope and supporting evidence. Assign severity, owner, due date and status. Management response, remediation and closure evidence remain separate decisions. Closing a task is not the same as validating the finding.

Workpaper Packs

Select the correct system, framework, period and purpose. Before export, confirm scope, evidence links, reviewer sign-off, open limitations and generation metadata. Packs are point-in-time artifacts; regenerate after material source changes.

Board Dashboard

Use it to communicate exposure, trends, overdue high-severity work, material exceptions and decisions. Avoid presenting coverage or maturity as legal compliance. Drill into source records before acting on an aggregate.

Audit Trail

Use the Audit Trail to investigate state changes and decision history. It supports accountability but does not prove the substantive correctness of the decision. Restrict access, export only what is needed and preserve retention requirements.

End-to-end closure

1. Assessment identifies a claim or gap.
2. Evidence is requested and reviewed.
3. A test evaluates operation.
4. Exceptions become findings and risks where appropriate.
5. Remediation is completed and retested.
6. A reviewer closes or qualifies the finding.
7. Workpapers and board reporting preserve the relevant conclusion and limitations.

Next

25. Assessment history and delta tracking

Establish baselines, compare assessment states and explain what changed without overwriting the original evidence or conclusion.

Assessment History, **Assessment Baselines & Delta Tracking** and **Automated Gap Narratives** answer a different question from the live assessment: **what changed since the selected baseline, why did it change and does the earlier conclusion remain defensible?**

Assessment History

Use history to locate prior assessment states, dates, systems, frameworks and conclusions. Confirm that two records refer to the same system boundary and comparable framework version before drawing a trend. A higher score can reflect real improvement, increased assessment completion or changed applicability; these are not the same outcome.

Assessment Baselines & Delta Tracking

A baseline should have a clear purpose, such as pre-deployment approval, annual assurance, post-incident state or certification-readiness review. Record its scope, cutoff date, framework version, assessor and known limitations.

Do not move the baseline merely to improve the reported trend. Create a new baseline when a major system boundary or methodology change makes direct comparison misleading.

Delta interpretation

Review changes in:

- system scope, purpose, data, supplier, geography or autonomy;
- routing and control applicability;
- assessment status and assurance depth;
- evidence freshness and test results;
- findings, accepted risks and remediation;
- human conclusions and review dates.

Automated Gap Narratives

An Automated Gap Narrative can summarise recorded changes, but it cannot determine whether the business accepted the change, whether evidence is credible or whether a legal conclusion is valid. Require assessor review and retain material caveats.

Defensible comparison checklist

- [ ] Same system identity or an explained boundary change.
- [ ] Comparable framework and methodology versions.
- [ ] Applicability changes separated from control improvement.
- [ ] Completion changes separated from assurance changes.
- [ ] Evidence period and freshness considered.
- [ ] Material regressions have an owner and treatment.
- [ ] The new conclusion records the decision basis and reassessment trigger.

Next

26. Cross-framework analysis and mappings

Interpret Cross-Framework View and framework mapping screens as traceability aids rather than automatic equivalence or compliance.

Cross-framework analysis shows where one recorded control, outcome or evidence item can support several governance views. It reduces duplicate work, but it does not make unlike requirements equivalent.

Available views

Gamut provides **Cross-Framework View**, **EU AI Act Mapping**, **NIST RMF Mapping**, **ISO 42001 Mapping**, **ISO 42005 Mapping** and **NAGF Compliance** views. Each uses the records available to the current workspace and system.

Use a mapping to answer:

- Which existing controls may support this requirement?
- Which evidence can be reused without changing its meaning?
- Where does the target framework require additional scope, legal analysis or assurance?
- Which gaps are unique to the target framework?

What a mapping does not prove

A mapped relationship does not by itself prove implementation, operating effectiveness, legal applicability, conformity or certification readiness. Frameworks differ in subject, terminology, unit of assessment, required evidence and decision authority.

For example, a GTSAF security control may support an EU AI Act risk-management obligation, but the legal role, applicable article, quality-management context and conformity evidence still require separate assessment.

Review workflow

1. Select the correct system and confirm its authoritative route.
2. Review the source item's score, evidence and assurance depth.
3. Read the target requirement in its own framework context.
4. Decide whether the evidence is reusable, partially supportive or irrelevant.
5. Record any target-specific evidence, gap or conclusion.
6. Revisit the mapping after scope, framework version or evidence changes.

Good reporting language

Use "mapped support", "potentially reusable evidence" or "traceability coverage". Avoid "compliant through inheritance", "automatically satisfied" or "certified by mapping".

Next

27. Certification and assurance readiness

Use readiness views to prepare governance records without presenting them as certification, legal advice or an external assurance opinion.

Cert Readiness (certification and assurance readiness) brings together scope, assessment coverage, evidence, tests, findings and management conclusions to show what still needs attention before an external review.

Readiness is an internal preparation view. It is not a certificate, conformity decision, legal opinion or guarantee of audit outcome.

Readiness dimensions

- **Scope:** the management system, AI system, sites, functions and exclusions are clear.
- **Coverage:** required clauses, controls or outcomes have been assessed.
- **Implementation:** documented controls are operating in the assessed environment.
- **Evidence:** artifacts are current, attributable and linked to the claim.
- **Testing:** material controls have appropriate design and effectiveness testing.
- **Findings:** gaps have owners, dates, treatment and closure evidence.
- **Governance:** conclusions, approvals, risk acceptance and review triggers are recorded.

Workflow

1. Confirm the target standard, scheme, period and scope.
2. Review the applicable framework module and any licence requirements.
3. Resolve routing and applicability before interpreting completion.
4. Review evidence quality and tests, not scores alone.
5. Identify open major issues, exclusions and dependencies.
6. Generate a readiness pack for internal review.
7. Agree corrective work and retest before external assurance.

Interpreting the result

A high percentage with weak evidence is not strong readiness. A lower result caused by newly identified scope can represent better governance than a narrow, superficially complete assessment. Explain the denominator, assessment period, assurance depth and unresolved limitations.

For ISO/IEC work, use a licensed copy of the relevant standard and an appropriately competent certification or assurance provider. Gamut does not reproduce controlled standard text or make the certification decision.

Next

28. Global Search

Find governed records across the current workspace while preserving tenant, role and entitlement boundaries.

Global Search helps locate systems, assessments and related governance records in the current workspace. Search results are constrained by the signed-in user's tenant, workspace access, role and entitlements. Search does not grant access to the underlying record.

Search effectively

- Use a stable system ID, control ID, owner, supplier or distinctive title where available.
- Open the result and verify system, workspace, framework and record status.
- Treat snippets as navigation aids, not the complete evidence or decision.
- Refine ambiguous searches rather than assuming the first result is correct.

Security and privacy

Do not use search to enumerate records beyond a legitimate work purpose. Results may contain sensitive governance metadata even when the underlying evidence is separately restricted. Report unexpected cross-workspace visibility immediately and do not investigate another tenant's data.

If an expected record is absent, verify the selected workspace, your role, module entitlement, record status and spelling. Administrators should correct access at the source rather than copying content into a less restricted area.

Next

Part 3: Framework assessment guides

This part compiles the current public operating guidance for the workflows in scope.

29. Frameworks overview & routing

How Gamut supports nine AI governance frameworks, GTSAF, EU AI Act, NIST AI RMF, ISO 42001/42005, NAGF, ACRS, ATF and MAESTRO, and routes the right controls to each system.

Gamut is multi-framework. A single AI system can be assessed against one framework or several. The linked System Record and Use Case Intake provide an authoritative routing basis; ACRS then helps set proportionate assurance depth for capability-driven risk.

This page explains how frameworks work in Gamut and how to choose between them. Each framework has its own page with detail.

Why framework references are used

Gamut uses framework names, article numbers, clause labels and control IDs to help you organise governance work, for navigation, assessment workflow and crosswalk traceability.

> **Not endorsements, not legal advice:** Framework references in Gamut are **not endorsements by the framework owners**, and the product does not replace licensed standards, legal advice, certification audits or regulator guidance. Third-party standard text is not reproduced. Always work from a licensed copy of any standard and obtain appropriate legal or regulatory advice.

For publication versions, legal snapshots and the reassessment treatment when a source changes, see Framework currency and versioning.

The frameworks at a glance

Gamut frameworks

Developed by Gamut as part of the platform:

Framework	Structure
GTSAF	358 controls across 17 domains (A to Q); 71 gate, 70 critical.
ACRS	4 capability dimensions, score to 81, 3 risk bands.

Open agentic frameworks Gamut implements

Created by others and implemented in Gamut, with attribution:

Framework	Structure	Created by
ATF	5 control elements, 4 autonomy levels, promotion gates.	Josh Woodruff (MassiveScale.AI), CC BY 4.0
MAESTRO	7 layers plus cross-layer threats, likelihood-times-impact risk scoring.	Cloud Security Alliance (Ken Huang)

Public and third-party references

Mapping Gamut's workflow to widely used external regimes and standards, at a product-safe level:

Reference	Structure
EU AI Act	11 orthogonal legal routes and 72 atomic checks anchored to Regulation (EU) 2024/1689.
NIST AI RMF	System-specific Current and Target Profiles across 4 functions, 19 categories and 72 Core outcomes.
ISO/IEC 42001	**Organisation-level AIMS:** 173 atomic clause checks plus 38 Annex A controls.
ISO/IEC 42005	**System-specific AI impact assessment:** 119 atomic items across 5 sections.
NAGF	**Nigerian AI governance:** 108 items, 7 sections and 19 legal or sector routes.

How crosswalks work

GTSAF is the hub. Every one of its 358 controls carries audited crosswalk mappings to the **EU AI Act**, **ISO/IEC 42001**, **ISO/IEC 42005** and **NIST AI RMF**. Because evidence attaches to GTSAF controls and those controls map outward, a single body of governance work supports several regimes at once, and a reviewer can trace any conclusion across frameworks.

See the GTSAF crosswalk table for the domain-by-domain mapping.

How routing works

1. **Register** the AI system and link its intake.
2. Complete the **authoritative cross-framework routing facts** for decision impact, action capability, roles, material capabilities and jurisdictions.
3. Gamut combines the two records into applicable, screening-required, not-applicable and organisation-level review decisions. Unknown material facts fail closed as screening required.
4. Review conflicts and missing information, then confirm the route. A material basis change makes the previous confirmation subject to reassessment.
5. Use ACRS and framework-specific routing to set the right control and assurance depth.
6. You **assess** the system against one or more frameworks, recording rationale and evidence.
7. Shared evidence and findings mean work done for one framework supports the others through crosswalk traceability.

See Routing & applicability for the full method and state definitions.

How scoring works

Gamut's frameworks share a consistent scoring philosophy, so a score means the same thing wherever you look.

- **Correct scope first.** Most framework assessments are system-specific. ISO/IEC 42001 is organisation-level, and ATF is agent-specific. Portfolio views must preserve those boundaries.
- **Coverage-inclusive.** Controls, requirements or threats that have not been assessed count as not-yet-met, not as passes. A barely-started assessment cannot report a high score, so progress and completeness stay visible together.
- **Justified N/A only.** Where a framework lets you mark an item not-applicable, doing so for an applicable item requires a documented justification. A justified N/A is taken out of scope; an unjustified N/A stays in scope and counts as not met.
- **Evidence quality is measured separately.** Conformance or compliance measures how much is met; evidence and ownership quality measure how strong the assessed work is, so the two are not conflated.

Each framework then applies its own scale on top of this: GTSAF a derived assurance proxy, ATF a strict zero-trust gate, the EU AI Act and NAGF a compliance percentage, ISO/IEC 42005 a maturity scale, and MAESTRO a likelihood-times-impact risk band.

Choosing a framework

30. Framework currency and versioning

Understand the methodology, source and legal-snapshot metadata that makes a Gamut assessment reproducible after frameworks change.

Every defensible assessment identifies the framework or legal snapshot used. A later publication, law, guidance document or product methodology must not silently rewrite an earlier conclusion.

What to record

- Framework and methodology name and version.
- Legal "as of" date where applicable.
- System and model version.
- Assessment and evidence cutoff dates.
- Assessor and reviewer.
- Pending changes and next review trigger.

Source order

Use authoritative primary sources: enacted law and official journals for legal obligations; publisher or canonical repository for frameworks; a licensed copy for ISO standards. Secondary summaries can help navigation but should not replace the source relied upon.

Current documentation basis

Module	Basis represented in these guides	Currency treatment
GTSAF	Gamut's 358-control trust, security and assurance framework	Record the Gamut catalogue/version used.
ACRS	Four-dimension capability-risk method	Record the calculation and routing-method version.
MAESTRO	CSA agentic AI threat-modelling framework	Recheck the CSA publication when its catalogue changes.
ATF	Specification 0.9.1 and Conformance 0.9.0, Public Review Draft	Show draft status and exact version in conclusions.
NIST AI RMF	NIST AI RMF 1.0 and applicable Profiles	Identify the RMF and Profile publication used.
EU AI Act	Regulation (EU) 2024/1689 and the module's stated legal snapshot	Check the Official Journal and adopted amendments before live legal conclusions.
NAGF	Consolidated Nigerian current-law, sector, policy and readiness model	Verify every relied-upon official source at the evidence cutoff.
ISO/IEC 42001	ISO/IEC 42001:2023	Use a licensed copy and record edition.
ISO/IEC 42005	ISO/IEC 42005:2025	Use a licensed copy and record edition.

Legal currency note, 21 July 2026

The public documentation was rechecked on **21 July 2026** against the Official Journal text of Regulation (EU) 2024/1689, current European Commission implementation material, the Nigeria Data Protection Act 2023 and official Nigerian AI strategy material published by NCAIR/NITDA. The EU module retains its displayed `EUAIA-2026-07-16-defensible-system-scope` identifier so the guide matches the implemented assessment. Political agreement, proposed amendments and draft guidance are not silently treated as enacted law.

This check is not legal advice and does not replace verification at the assessment date.

Change procedure

When a source changes, determine whether it affects wording, applicability, scoring, evidence or the conclusion. Update methodology identifier and catalogue together where required. Keep the earlier assessment reproducible, mark affected systems for reassessment and record the change basis.

Next

31. GTSAF

Complete guide to the Gamut Trust, Security and Assurance Framework: 358 controls, factual system scoping, assurance depth, labels, evidence, testing, scoring, AI Assist and reporting.

GTSAF, the **Gamut Trust, Security and Assurance Framework**, is Gamut's flagship AI assurance framework. It turns AI governance principles into an assessable, evidence-led control system covering the entire AI lifecycle: governance, data, models, prompts, runtime, identity, agents, suppliers, monitoring, human oversight, resilience, auditability and infrastructure.

GTSAF contains **358 individually authored controls across 17 domains**, supported by:

- A unique objective and risk statement for every control.
- Detailed implementation guidance.
- Control-specific evidence expectations.
- Control-specific test procedures.
- Criticality and gate metadata.
- System-specific applicability rules.
- Shared-responsibility ownership.
- Structured assessor conclusions and residual-risk records.
- Audited traceability mappings to the EU AI Act, ISO/IEC 42001, ISO/IEC 42005 and NIST AI RMF.
- Validated, system-scoped AI Assist.

What GTSAF is: GTSAF is Gamut's native assurance baseline and assessment methodology. It is designed to help an organisation determine whether trustworthy-AI controls are applicable, implemented, evidenced, tested and operating effectively.

What GTSAF is not: GTSAF is not a certification scheme, legal opinion or automatic compliance determination. Gamut's derived assurance percentage and 1-to-5 assurance score are assessment aids. They support human judgement and do not replace an assessor, auditor, regulator or accredited certification body.

Start here

Use this documentation suite according to what you need to explain or do:

If you need to...	Read
Explain badges such as **High**, **Gate**, **Enhanced**, **Triggered**, **Gap** and **Assured**	Concepts, labels and terminology
Understand what each of the 17 domains covers and how to assess it	Domains and control library
Explain why different systems receive different controls and assurance depth	System scope, applicability and assurance depth
Complete a GTSAF assessment from beginning to sign-off	Assessment workflow
Defend how the assurance figures are calculated	Scoring and assurance model
Know what evidence to collect and how to test a control	Evidence, testing and findings
Explain what AI Assist reads, produces and is prohibited from doing	AI Assist explainability
Produce or explain system and workspace reports	Reporting and governance decisions
Walk through a realistic assessment example	Worked example
Look up abbreviations, statuses and formulas quickly	Reference and glossary

GTSAF at a glance

Property	Current GTSAF library
Version	GTSAF v1.3
Controls	358
Domains	17, lettered A to Q
Gate controls	71
Critical controls	70
Criticality distribution	70 Critical, 241 High, 43 Medium, 4 Low
Assessment level	Per AI system, with workspace roll-up
Control identifier	`GTSAF--`, for example `GTSAF-A-01`
Assessment answers	Yes, No or governed N/A
Shared-responsibility roles	MP, OSP, AP, AIC, CSP, Shared or ND
Crosswalks	EU AI Act, ISO/IEC 42001, ISO/IEC 42005, NIST AI RMF

The GTSAF operating model

GTSAF works as a connected assurance process rather than a static checklist:

1. **Register the AI system.** Record what it does, who owns it, where it operates, what data it handles, which suppliers it uses and how much autonomy or impact it has.
2. **Complete intake and risk classification.** The intake establishes the system context. ACRS captures capability exposure through dependency, action autonomy, access scope and harm potential.
3. **Determine applicability and depth.** Complete system facts determine which conditional controls apply. ACRS and the governance weighting profile set proportionate Baseline, Triggered or Enhanced assurance depth without rewriting factual applicability.
4. **Assess the controls.** The assessor records Yes, No or N/A; assigns shared-responsibility ownership; documents implementation; and captures customer responsibilities.
5. **Collect and review evidence.** Narrative statements explain the control but do not count as verified evidence. Evidence must be linked, reviewed and quality-rated.
6. **Test operating effectiveness.** A test determines whether the control actually works under defined conditions. Failed tests and exceptions constrain assurance.
7. **Raise findings and remediation.** Gaps, gate failures, rejected evidence and failed tests become actionable findings, risk decisions or remediation work.
8. **Record the human conclusion.** The assessor documents design, implementation and operating effectiveness, residual risk, monitoring cadence and reassessment triggers.
9. **Generate system and workspace reporting.** System reports use the selected system's own assessment bucket. Workspace reports provide portfolio-level roll-up without replacing the individual system record.

The four dimensions shown on a control

Several badges can appear beside one control because they answer different questions.

Dimension	Example labels	The question it answers
Criticality	Critical, High, Medium, Low	How consequential would failure of this control be?
Control type	Gate	Can failure of this control prevent a positive assurance conclusion?
Applicability	Mandatory, Triggered, Enhanced, Not applicable	Why and to what depth does this control apply to this system?
Assessment result	Unassessed, Partial, Supported, Assured, Gap, Gate fail, N/A	What does the current assessment record demonstrate?

For example:

High · Gate · Enhanced · Gap

means:

- The control has **High** criticality.
- It is a **Gate** control.
- The selected system requires **Enhanced** assessment depth.
- At least one applicable requirement is answered **No**, producing a **Gap**.

It does **not** mean "High Gate Enhanced Gap" is one combined severity rating.

See Concepts, labels and terminology for every label and its decision consequence.

The 17 domains

Domain	Name	Controls	Primary assurance focus
A	Governance, Strategy and Accountability	11	Board authority, policy, ownership, decision rights and governance operation
B	Legal, Regulatory and Contractual Compliance	11	Applicable obligations, prohibited uses, contracts and jurisdiction
C	AI Use Case Intake, Approval and Risk Tiering	11	Registration, classification, approval, exceptions and change triggers
D	Data Governance, Lineage and Provenance	11	Data origin, lineage, quality, rights, retention and traceability
E	Data Security and Privacy Engineering	24	Sensitive data, privacy controls, segregation, minimisation and protection
F	Secure Data Acquisition and Annotation	11	Data collection, labelling, annotation, poisoning resistance and quality
G	Model Development, Validation and Robustness	11	Model engineering, validation, robustness, benchmarks and release controls
H	Prompt, Context and Retrieval Security	25	Prompt injection, RAG, context boundaries, grounding and retrieval integrity
I	Inference, API and Runtime Security	25	Serving security, APIs, sessions, output controls and runtime enforcement
J	Identity, Access and NHI Security	26	Human and non-human identity, secrets, privilege and access lifecycle
K	Agentic AI and Autonomous Action Governance	29	Tools, memory, delegation, approvals, autonomy and kill switches
L	Third-Party, Model and Software Supply Chain Assurance	25	Vendors, hosted models, dependencies, SBOMs and shared responsibility
M	Monitoring, Detection and AI Security Operations	35	Logging, detection, behavioural monitoring, incidents and security operations
N	Human Oversight, Transparency and Impact Management	15	Human review, notices, explainability, fairness, appeal and social impact
O	Resilience, Continuity and Recovery	36	Failover, rollback, crisis response, recovery and provider substitution
P	Auditability, Evidence and Assurance	11	Evidence governance, testing, audit records and assurance independence

Domain	Name	Controls	Primary assurance focus
Q	Infrastructure, Platform and Environment Security	41	Cloud, compute, network, platform, environment and isolation safeguards

The larger domains reflect modern AI concentration points: infrastructure, monitoring, agentic action, identity, prompt/retrieval security and third-party dependencies.

What is recorded for each control

Every control contains enough information for an assessor to work primarily from the control screen:

- **Control statement**: the required outcome.
- **Objective**: why the control exists.
- **Risk addressed**: what can go wrong if it is absent or ineffective.
- **Applicability**: when the control applies and why.
- **Criticality and gate status**.
- **Implementation guidance**: practical actions expected to establish the control.
- **Required evidence**: the artefacts and operating records to obtain.
- **Test procedure**: how to determine whether the control works.
- **Control owner and evidence owner**.
- **Cross-framework mappings**.
- **Shared-responsibility role guidance**.
- **Assessment answers and implementation narrative**.
- **Linked evidence, evidence requests, tests and findings**.
- **Structured assessor conclusion and residual risk**.
- **System-scoped AI Assist**.

Shared responsibility

GTSAF recognises that AI controls are commonly split across several parties:

Code	Party
MP	Model Provider
OSP	Orchestrated Service Provider
AP	Application Provider
AIC	AI Customer
CSP	Cloud Service Provider
Shared	Responsibility is divided between named parties
ND	Not determined; an ownership gap requiring resolution

The selected guidance role changes the practical implementation guidance shown to the assessor. It does not transfer accountability automatically. Contracts, architecture and actual operating responsibility must support the selected ownership.

Assurance principles

GTSAF follows several rules that prevent an assessment from overstating assurance:

- **Unanswered controls are not treated as safe.** Coverage is included in conformance and assurance.
- **Narrative is not evidence.** A detailed explanation improves narrative sufficiency but does not become verified evidence merely because it sounds credible.
- **A Yes answer is not proof.** Evidence and testing are separate.

- **Operating effectiveness requires testing or operating records.**
- **Failed tests, rejected evidence and gate failures constrain the result.**
- **Critical and High controls receive greater weighting and assurance caps when evidence is weak.**
- **N/A is governed.** Unsupported N/A responses remain in scope and count as not met.
- **AI Assist cannot approve the assessment.** Human review and sign-off remain mandatory.
- **System boundaries are enforced.** A selected system does not inherit another system's AI analysis or operational evidence.

Crosswalks to public frameworks

Every GTSAF control includes audited traceability references to:

- **EU AI Act**
- **ISO/IEC 42001**
- **ISO/IEC 42005**
- **NIST AI RMF**

The crosswalk helps an assessor reuse evidence and identify related obligations. It does not prove that satisfying one GTSAF control automatically satisfies an external legal or standards requirement.

> **Traceability, not equivalence:** Mappings identify relevant review anchors. They are not claims of automatic compliance, legal equivalence or certification readiness. Validate each mapping against the organisation's role, jurisdiction, risk classification and licensed copy of the applicable standard.

Recommended reading journey

For a complete understanding, continue in this order:

32. EU AI Act - defensible compliance

Complete guide to Gamut's system-scoped EU AI Act method: legal scope, roles, routing, 72 atomic checks, evidence, assurance, AI Assist and defensible conclusions.

Gamut turns Regulation (EU) 2024/1689 into a system-scoped assessment that connects legal classification, operator roles, implementation, evidence, testing, findings and accountable human sign-off.

The module is designed to answer:

> Which EU AI Act provisions apply to this named AI system and organisation, why do they apply, what has been demonstrated, what remains unresolved, and what conclusion can responsibly be made as of the stated legal snapshot?

> **Readiness and compliance support, not legal advice:** Gamut supports a structured and auditable compliance determination. It does not provide legal advice, replace the official Regulation, determine facts that have not been supplied, or certify an organisation. Qualified counsel should review material scope, classification, prohibited-practice, role and exception decisions.

Start here

If you need to…	Read
Understand application dates, territorial scope and regulated roles	Legal status, scope and roles
Understand why categories appear for a system	Routing and classification
See every control family and atomic legal check	Atomic requirement catalogue
Complete an assessment from intake to confirmation	Assessment workflow
Explain Compliant, Non-compliant, N/A and depth 0-3	Scoring, assurance and conclusions
Build an evidence pack and test obligations	Evidence, testing and findings
Use system-scoped AI analysis securely	AI Assist and security
Produce a defensible system or portfolio report	Reporting and governance
Follow a complete example	Worked example

If you need to…	Read
Look up routes, statuses, dates, terms and safe explanations	Reference and glossary

At a glance

Property	Gamut implementation
Legal baseline	Regulation (EU) 2024/1689
Legal snapshot	Reviewed 16 July 2026
Methodology identifier	`EUAIA-2026-07-16-defensible-system-scope`
Assessment scope	One named AI system and its current intake
Routed categories	11
Parent requirements	34
Atomic assessment items	72
Hard stop	Confirmed or unresolved potential Article 5 prohibited practice
Answers	Compliant, Non-compliant or N/A
Assurance depth	0 Gap, 1 Asserted, 2 Supported/Partial, 3 Assured
Confirmation model	Authoritative, evidence- and test-gated human sign-off
Legal presentation	Current-law conclusion separated from future-readiness information

This is not a single risk-class selector

The Act contains different legal mechanisms: scope, prohibited practices, high-risk classification, operator duties, transparency triggers, GPAI duties and application dates. They are not one mutually exclusive ladder.

A system may, for example:

- Be within territorial scope.
- Be an Annex III high-risk system.
- Trigger direct-interaction transparency.
- Use a GPAI model supplied by another organisation.
- Require deployer duties but not a FRIA.
- Carry future obligations that are not yet used in the current-law conclusion.

Gamut therefore calculates **orthogonal routes**. Each route is independently supported, excluded or left unresolved from structured system facts. The server, not free-form browser text, determines the authoritative route.

The eleven routes

Route	Purpose
`SCOPE`	EU nexus, exclusions, AI-system definition, role, timing and AI literacy
`PROHIBITED`	Eight Article 5 prohibited-practice checks and hard-stop logic
`CLASSIFICATION`	Article 6, Annex I, Annex III, exceptions and profiling override
`HIGHREQ`	High-risk system requirements in Articles 8-15
`OPERATORS`	Provider, deployer, representative, importer, distributor and value-chain duties
`FRIA`	Article 27 fundamental-rights impact assessment applicability and completion
`TRANSPARENCY`	Trigger-specific Article 50 notices, marking and disclosures
`MONITORING`	High-risk post-market monitoring and serious-incident handling
`GPAI_MONITORING`	Article 55 systemic-risk evaluation, incidents and cybersecurity
`GPAI`	GPAI classification, documentation, copyright, downstream and representative duties
`NA`	Governed non-applicability decisions and residual evidence trail

See Routing and classification for the decision logic and contradiction handling.

The operating model

```
Named AI system
-> structured legal intake
-> authoritative route and legal status
-> 72 applicable atomic checks
-> implementation answer and depth
-> accepted evidence + scoped test + findings review
-> current-law and future-readiness conclusions
-> accountable human confirmation
-> system report and conservative portfolio roll-up
```

The route decides **what must be assessed**. The assessment record decides **what has been demonstrated**. Neither is a substitute for the other.

What makes a conclusion defensible

A strong conclusion requires all of the following:

1. The correct named system and legal boundary.
2. Structured facts sufficient to determine scope, role and triggered routes.
3. No unresolved contradictions.
4. Each applicable atomic item assessed independently.
5. Compliant items supported at depth 3 by accepted evidence and effective testing.
6. N/A items supported by a complete, approved, trigger-specific decision.
7. No unresolved adverse finding or failed test contradicting the claim.
8. Article 5 screening clear.
9. Current binding law separated from future or proposed changes.
10. An owner, confidence rating, residual-risk statement, review date, detailed conclusion and reassessment triggers.

Gamut fails confirmation closed when these conditions are not met.

Legal-status discipline

The module distinguishes:

- **In force**: used in the current-law compliance conclusion.
- **Future obligation**: tracked for readiness but not presented as a current breach.
- **Proposed change only**: implementation intelligence that is not substituted for binding law.
- **Non-binding guidance or code**: useful interpretive or demonstration material, clearly labelled.

The controlling text remains the official Regulation on EUR-Lex. The European Commission's AI Act policy page should be checked for current implementation information.

Security and governance principles

The assessment follows:

- **System isolation**: answers, evidence, tests, findings and AI analysis remain tied to the selected system.
- **Zero trust**: identity, workspace, role, entitlement and object scope are revalidated by the server.
- **Least privilege**: assessors and AI providers receive only the information required for the authorised action.
- **Fail closed**: missing scope facts, prohibited-practice potential, contradictions and assurance gaps block an overstated confirmation.
- **Human accountability**: AI suggestions cannot confirm applicability, accept evidence, approve N/A or sign the conclusion.
- **Change sensitivity**: a material change to the legal basis invalidates confirmation and requires reassessment.

Recommended reading journey

33. NIST AI RMF

Complete public guide to using Gamut for system-specific NIST AI RMF Current and Target Profiles across GOVERN, MAP, MEASURE and MANAGE.

Gamut supports system-specific assessment against the **NIST Artificial Intelligence Risk Management Framework (AI RMF) 1.0**. The module helps an assessor connect the framework's four Core functions to a named AI system, record a Current Profile, define a Target Profile, collect evidence, test relevant practices, manage findings and document an accountable conclusion.

The central question is:

> For this AI system and operating context, which NIST AI RMF outcomes are currently achieved, what target state is required, what evidence supports the assessment, and what actions remain?

> **Voluntary framework; not NIST endorsement:** The NIST AI RMF is voluntary. Gamut is not endorsed by NIST, does not reproduce or replace NIST publications, and does not provide NIST certification. Use the official AI RMF page and NIST AI 100-1 as the authoritative sources.

> **Current publication status:** This guide describes the published **AI RMF 1.0** baseline. NIST states that AI RMF 1.0 is being revised and that the Playbook will be updated after the revision. Check the official NIST AI RMF page before relying on time-sensitive publication status.

Start here

If you need to…	Read
Explain the difference between the Core, Current Profile, Target Profile and Playbook	Core, Profiles and Playbook
See all 4 functions, 19 categories and 72 outcomes	Functions and outcome catalogue
Explain why every system is available and why some outcomes receive more attention	System scope and prioritisation
Complete an assessment from intake to accountable conclusion	Assessment workflow
Explain Current outcome, Target outcome, assurance depth and confirmation	Outcomes, assurance and conclusions
Know what evidence to request, how to test and when to raise findings	Evidence, testing and findings
Assess a generative AI system using NIST AI 600-1	Generative AI Profile
Use AI assistance responsibly and understand what it can consider	AI-assisted assessment
Produce and explain system or portfolio reporting	Reporting and governance
Follow a realistic assessment example	Worked example
Look up labels, terms and safe explanatory language	Reference and glossary

At a glance

Property	Gamut assessment
Authoritative framework	NIST AI RMF 1.0, NIST AI 100-1
Nature	Voluntary AI risk-management framework
Primary scope	One selected AI system and its operating context
Core functions	GOVERN, MAP, MEASURE and MANAGE
Categories	19
Core subcategories/outcomes	72
Assessment structure	Current Profile plus Target Profile
Current outcome labels	Not assessed, Not achieved, Partially achieved, Achieved
Assurance depth	Unverified, Documented, Implemented, Assured
Generative AI companion	NIST AI 600-1 when generative-AI characteristics are present

Property	Gamut assessment
Evidence boundary	Evidence, tests and findings linked to the selected system and outcome
AI support	Whole-system or single-outcome advisory analysis
Human accountability	The assessor owns final outcomes, risk decisions and conclusions

The four functions

Function	Purpose	Categories	Outcomes
GOVERN	Establish policies, accountability, culture and organisation-wide risk-management conditions.	6	19
MAP	Establish context and identify intended purpose, actors, impacts, benefits, costs and risks.	5	18
MEASURE	Evaluate, test, monitor and track AI risks and trustworthy characteristics.	4	22
MANAGE	Prioritise and treat risks, make proceed decisions, respond to incidents and improve controls.	4	13

GOVERN is cross-cutting. MAP, MEASURE and MANAGE are iterative rather than a one-time linear sequence. The framework should be revisited as the system, context, evidence, risks and stakeholder expectations change.

How Gamut operationalises the framework

```
Named AI system and context
-> NIST AI RMF Core available
-> context-based assessment priorities
-> Current Profile outcome for each Core subcategory
-> assurance depth, evidence, testing and findings
-> Target Profile outcome
-> treatment, monitoring and reassessment
-> accountable human conclusion
```

The **Current Profile** describes the system's assessed present position. The **Target Profile** describes the intended risk-management outcome. The gap between them is the improvement plan.

The trustworthiness characteristics

NIST describes trustworthy AI through connected characteristics:

- Valid and reliable.
- Safe.
- Secure and resilient.
- Accountable and transparent.
- Explainable and interpretable.
- Privacy-enhanced.
- Fair, with harmful bias managed.

These characteristics interact and may involve trade-offs. A system is not trustworthy merely because one metric is strong. For example, high predictive performance does not by itself establish safety, fairness, privacy, transparency or appropriate human oversight.

What a defensible assessment contains

A well-supported record should include:

- The correct named system, version, purpose, lifecycle stage and deployment context.
- Relevant users, affected people, data, suppliers and system boundaries.
- A Current Profile outcome for every considered Core subcategory.

- A Target Profile outcome appropriate to risk tolerance and intended use.
- System-specific rationale rather than generic policy statements.
- Evidence showing design and operation.
- Test results and defined pass criteria where effectiveness is claimed.
- Findings for gaps, failed tests, unsupported assumptions or adverse evidence.
- Owners, treatment decisions, residual risk and monitoring.
- A review date and material reassessment triggers.
- A clear human-authored conclusion.

What the assessment does not prove

It does not by itself prove:

- Legal compliance.
- Certification.
- NIST approval or endorsement.
- That every risk has been eliminated.
- That a policy operates effectively.
- That another system inherits the same outcome.
- That a mapped control automatically satisfies a NIST outcome.
- That AI-generated advice is correct or approved.

Official resources

- NIST AI Risk Management Framework
- NIST AI 100-1: AI RMF 1.0
- NIST AI RMF Core
- NIST AI RMF Playbook
- NIST AI 600-1: Generative AI Profile
- NIST Trustworthy and Responsible AI Resource Center

Recommended reading journey

34. ISO/IEC 42001

Complete public guide to Gamut's organisation-level ISO/IEC 42001:2023 AIMS conformity-readiness assessment across clauses 4-10 and Annex A.

Gamut supports an organisation-level **AI management system (AIMS) conformity-readiness assessment** aligned to **ISO/IEC 42001:2023**. It covers clauses 4 to 10 as atomic assessor checks and maintains a Statement of Applicability for the Annex A reference controls.

The central question is:

> Within the approved AIMS scope, what evidence supports conformity with each requirement, which Annex A controls are applicable, and what nonconformities or limitations prevent a defensible readiness conclusion?

> **Use a licensed standard:** This documentation explains Gamut's workflow and does not reproduce ISO/IEC 42001. It does not replace a licensed copy, legal advice, internal audit or independent certification audit. Use the official ISO page and a licensed copy as authoritative.

> **Readiness is not certification:** Only an appropriate independent certification body can issue certification. A Gamut conclusion records the assessor's readiness judgement for the stated scope and evidence cutoff; it is not an ISO certificate, ISO endorsement or guarantee of audit outcome.

Start here

If you need to…	Read
Define the AIMS scope and understand clauses 4-10 and Annex A	Scope, clauses and Annex A
Complete the assessment	Assessment workflow
Explain conformity, assurance and the SoA	Conformity, assurance and SoA
Prepare evidence, audit tests and findings	Evidence, audit and findings
Use per-requirement or complete-scope AI assistance	AI Assist and security
Produce a readiness conclusion without overstating certification	Reporting and certification readiness
Look up labels and definitions	Reference and glossary

At a glance

Property	Gamut assessment
Standard	ISO/IEC 42001:2023, edition 1
Nature	Certifiable AI management-system requirements standard
Scope	Approved organisation-level AIMS boundary
Structure	Clauses 4-10 plus Annex A reference controls
Assessment checks	173 atomic clause checks plus 38 Annex A controls
Conformity labels	Not assessed, Conformity supported, Partial evidence, Nonconformity
Assurance depth	Unverified, Documented, Implemented, Assured
Annex A applicability	Undetermined, Applicable, or justified Not applicable
Evidence boundary	Scope-, period- and requirement-specific
AI support	Per requirement and, after completion, whole-scope readiness analysis
Human accountability	Scope approval, SoA decisions, findings and final conclusion remain human-owned

The eight assessment sections

Section	Checks	Purpose
Context of the Organisation	14	Context, interested parties, AIMS scope and management-system processes
Leadership	21	Commitment, AI policy, responsibilities and authorities
Planning	49	Risks and opportunities, risk assessment, treatment, impacts, objectives and change
Support	26	Resources, competence, awareness, communication and documented information
Operation	19	Operational control, AI risk work and impact assessment
Performance Evaluation	32	Monitoring, measurement, internal audit and management review
Improvement	12	Nonconformity, corrective action and continual improvement
Annex A Reference Controls	38	Applicability and implementation of AI-specific reference controls

Atomic checks make compound requirements assessable without treating one strong sub-part as proof of the whole clause.

How Gamut operationalises the AIMS assessment

```
Approved organisation-level AIMS scope
-> clauses 4-10 assessed atomically
-> Annex A applicability and implementation decisions
-> requirement-level conformity and assurance depth
```

```
-> evidence, audit tests, nonconformities and corrective action
-> coverage and readiness gates
-> human conformity-readiness conclusion
-> management review and continual improvement
```

What a defensible record contains

- Named AIMS scope, boundaries, functions, sites and activities.
- Included AI systems and justified exclusions.
- Interested parties, owner, approver, version and approval date.
- A conformity result and rationale for every applicable atomic check.
- Separate assurance depth showing how strongly the result was verified.
- A complete, risk-linked Statement of Applicability.
- Accepted objective evidence and representative audit tests.
- Nonconformities, owners, corrective action and verification.
- Evidence cutoff, limitations, next review and accountable conclusion.

What the assessment does not prove

It does not by itself prove:

- ISO/IEC 42001 certification.
- That a management system operates outside the defined scope.
- That an Annex A control is effective because it is marked applicable.
- That a policy is implemented.
- Legal compliance in every jurisdiction.
- That a mapped framework score satisfies an ISO requirement.
- That AI-generated analysis is an audit opinion.

Recommended reading journey

35. ISO/IEC 42005

Complete public guide to Gamut's system-specific ISO/IEC 42005:2025 AI system impact assessment across 119 atomic items.

Gamut supports **AI system impact assessment** aligned to **ISO/IEC 42005:2025**. The module provides 119 atomic assessment items covering the assessment process, its documentation and the analysis and treatment of actual and reasonably foreseeable impacts.

The central question is:

> For this selected AI system, its intended and foreseeable uses, affected parties and operating context, what benefits and harms may arise, how were they evaluated, and what measures and accountable decisions follow?

> **Use a licensed standard:** This documentation explains Gamut's workflow and does not reproduce ISO/IEC 42005. Use the official ISO page and a licensed copy as authoritative.

> **Guidance standard:** ISO/IEC 42005 provides guidance for conducting AI system impact assessments. Gamut records an alignment and assurance judgement; it does not issue certification or prove legal compliance.

Start here

If you need to…	Read
Understand system scope, the process and 119-item catalogue	Scope, process and catalogue
Complete an assessment end to end	Assessment workflow

If you need to…	Read
Explain maturity, assurance and approved N/A	Maturity, assurance and applicability
Gather evidence and engage affected parties	Evidence, engagement and findings
Use per-item and whole-system AI assistance	AI Assist and security
Report and maintain the impact assessment	Reporting and reassessment
Look up labels and safe explanatory language	Reference and glossary

At a glance

Property	Gamut assessment
Standard	ISO/IEC 42005:2025, edition 1
Nature	Guidance for AI system impact assessment
Scope	One selected AI system, version and use context
Structure	Clauses 5 and 6 represented in five sections
Atomic items	119
Maturity	1 Not Addressed to 5 Optimised
Assurance depth	Unverified, Documented, Implemented, Assured
Applicability	In-scope by default; N/A requires structured human approval
Evidence boundary	Selected system, affected parties, context and assessment period
AI support	Per-item and complete-system advisory analysis
Accountability	Human assessor owns scope, engagement, impact, treatment and conclusion

The five sections

Section	Items	Purpose
Process Foundation	18	Establish and document a repeatable, integrated impact-assessment process
Governance & Triggers	25	Timing, scope, responsibilities and thresholds for sensitive or restricted uses
Execution & Review	25	Perform, analyse, record, approve, monitor and review assessments
Documentation Content	29	System, use, data, model, deployment and interested-party information
Impacts & Measures	22	Identify impacts and record measures addressing harms and benefits

How Gamut operationalises the assessment

```
Selected AI system and approved boundary
-> intended use, foreseeable misuse and affected parties
-> impact-assessment trigger, responsibility and thresholds
-> 119 atomic maturity and assurance decisions
-> evidence, stakeholder input, tests and findings
-> harms, benefits, measures and residual uncertainty
-> accountable system-level conclusion
-> monitoring and event-driven reassessment
```

What a defensible record contains

- Selected system, version, purpose and deployment environment.
- Organisational role, lifecycle stage and dependencies.
- Intended uses and reasonably foreseeable misuse.
- Directly and indirectly affected people, groups and society.
- Data, model, algorithm and deployment information.

- Stakeholder and affected-party engagement.
- Actual and foreseeable beneficial and adverse impacts.
- Severity, likelihood, distribution, reversibility and uncertainty.
- Measures, owners, residual impacts and decision thresholds.
- Objective evidence, testing, findings and limitations.
- Approval, publication decision, monitoring and reassessment triggers.

What the assessment does not prove

It does not by itself prove:

- That every affected party was identified.
- That stakeholder engagement occurred because a policy requires it.
- That legal impact assessments are complete.
- That a high maturity score means impacts are acceptable.
- That an item is irrelevant without an approved basis.
- That crosswalked evidence automatically satisfies the guidance.
- That AI-generated analysis is an approved impact conclusion.

Recommended reading journey

36. NAGF

Complete guide to Gamut's 108-item Nigerian AI Governance Framework assessment, legal-source status, 19 regulatory routes and separate current-law and readiness conclusions.

NAGF is Gamut's system-specific **Nigerian AI Governance Framework** assessment. It brings together current cross-sector law, regulator instruments, sector obligations, national AI policy, proposed legislation and responsible-AI readiness without presenting them as one legal instrument.

The central question is:

> For this AI system's Nigerian nexus, organisational role and regulated activities, which current obligations apply, what evidence supports compliance, and what separate policy or future-readiness work remains?

> **Not legal advice; verify current law:** Nigeria's AI governance landscape changes through legislation, directives, sector rules, regulator notices, litigation and policy. NAGF identifies the source and status behind an item, but it does not replace legal advice or the latest official instrument. Confirm time-sensitive status with the responsible authority and qualified counsel.

> **Not a claim that Nigeria has one omnibus AI Act:** NAGF is a consolidated Gamut assessment model. It does not imply that every item is binding, that one regulator has universal AI jurisdiction, or that a proposed bill is enacted.

> **Source currency:** Official-source status in this guide was rechecked on **21 July 2026**. Each assessment must still record its own evidence cutoff and verify the instrument, regulator notice and sector trigger relied upon.

Start here

If you need to…	Read
Understand the source hierarchy and legal-status lanes	Legal landscape and sources
Determine Nigerian nexus and the 19 regulatory routes	Scope and regulatory routing
Understand all seven sections and 108 items	Sections and item catalogue
Complete the assessment and separate legal from readiness conclusions	Assessment workflow and conclusions
Gather evidence, test and manage findings	Evidence, testing and findings
Use per-item and system-level AI assistance	AI Assist and security
Report and respond to legal change	Reporting and legal change
Look up labels and cautions	Reference and glossary

At a glance

Property	Gamut assessment
Nature	Consolidated Nigerian legal, regulatory, policy and responsible-AI assessment
Scope	One selected AI system and its Nigerian nexus
Sections	7
Assessable items	108
Regulatory routes	19
Source lanes	Current law, binding directive/code, conditional sector rule, policy, proposed law, best practice and inferred mapping
Item results	Compliant, Non-compliant, N/A
Depth	Three levels tailored to compliant or non-compliant status
Applicability	Structured route and item-level human approval
Conclusions	Current-law compliance and broader readiness kept separate
AI support	Per-item and complete-system advisory analysis
Accountability	Assessor and legal/compliance reviewer own final conclusion

The seven sections

Section	Items	Primary focus
AI Governance Framework	16	Accountability, policy, inventory, governance, incidents and legal change
NDPC Data Protection Compliance for AI	20	NDPA and applicable NDPC requirements for AI processing
Responsible and Ethical AI	19	Fairness, language, transparency, accessibility, labour, community and content impacts
AI Adoption and Sector Compliance	15	Adoption governance, public-sector, procurement, rights and cross-sector gateways
Sector-Specific AI Obligations	24	Finance, insurance, pensions, capital markets, telecoms, health, aviation and consumers
AI Infrastructure Readiness	6	Infrastructure, resilience, cloud, security and sustainability readiness
AI Ecosystem and Talent	8	Skills, research, local capability, inclusion and ecosystem development

How Gamut operationalises NAGF

```
Selected system and authoritative Nigerian-nexus route
-> 19 legal and sector route decisions
-> current-law, conditional, policy and proposed-law lanes
-> 108 item-level compliance decisions
-> depth, legal basis, evidence, tests and findings
-> approved N/A where genuinely outside scope
-> separate current-law and readiness conclusions
-> legal-change and event-driven reassessment
```

What a defensible record contains

- Selected system, role, purpose and Nigerian nexus.
- Applicable regulator, licence, sector and processing routes.
- Official source and current status for every legal proposition relied upon.
- Item result, depth and system-specific rationale.
- Accepted evidence and bounded tests.
- Approved N/A basis, owner and review date.

- Findings and treatment.
- Separate current-law and policy/proposed-law readiness positions.
- Assessor, legal/compliance reviewer, evidence cutoff and next review.

What the assessment does not prove

It does not by itself prove:

- That every NAGF item is law.
- That a proposed bill is enacted.
- That a sector rule applies without the regulated activity or role.
- That an official notice has not been superseded.
- Legal compliance from a percentage alone.
- That another framework's evidence automatically satisfies a Nigerian obligation.
- That AI-generated analysis is legal advice.

Recommended reading journey

37. ACRS

Complete guide to Gamut's Agentic Capability Risk Score: system scope, automatic inference, assessor overrides, product scoring, severity floors, GTSAF routing, evidence, AI Assist and governance decisions.

ACRS, the **Agentic Capability Risk Score**, is Gamut's native method for classifying the capability exposure of an AI system or agentic workflow. It asks four practical questions:

1. How dependent are operations, people or decisions on the AI?
2. How much can it decide or do before authoritative human approval?
3. Which data, tools, identities, systems and actions can it reach?
4. How serious could the credible consequences of failure, misuse or compromise become?

The answers form a four-part vector and a product score. Gamut then applies conservative **severity-floor rules** so a mathematically modest product cannot under-route a system with severe harm, consequential autonomy, broad access or explicit high-risk characteristics.

ACRS is not a control-compliance score. It establishes the **depth of assurance the system needs**. The routed result feeds the GTSAF assessment plan and informs threat modelling, approval, evidence, testing, monitoring and governance attention.

Gamut-native methodology: ACRS is a versioned Gamut-native methodology. It works alongside GTSAF, MAESTRO and the Agentic Trust Framework; it is not an external certification scheme.

What ACRS does not prove: A Low, Medium or High ACRS result describes capability exposure and required assurance depth. It does not prove that controls are implemented, evidence is sufficient, tests have passed, residual risk is acceptable, deployment is authorised or an external law or standard is satisfied.

Start here

If you need to...	Read
Explain every score, vector, band, status and severity-floor rule	Method, scoring and terminology
Assess dependency, autonomy, access and harm without leaving the screen	Dimension assessor playbooks
Understand system boundaries and how ACRS routes GTSAF	System scope and GTSAF routing
Complete an ACRS assessment from intake to sign-off	Assessment workflow
Know what evidence to request, how to test safely and when to raise findings	Evidence, testing and findings
Explain what AI Assist reads, produces, secures and cannot decide	AI Assist and security

If you need to...	Read
Explain reports, ownership, residual risk and reassessment	Reporting and governance
Follow a realistic assessment with severity-floor routing	Worked example
Look up fields, statuses, labels, formulas and assessor language	Reference and glossary

ACRS at a glance

Property	Gamut implementation
Methodology	Versioned Gamut-native capability-risk method
Primary assessment level	One selected AI system intake
Dimensions	4
Dimension levels	Low = 1, Medium = 2, High = 3
Vector	Dependency × Action autonomy × Access scope × Harm potential
Raw product	1 to 81
Product bands	Low 1-8; Medium 9-36; High 37-81
Authoritative route	Product band raised where a severity floor applies
Routed bands	Low, Medium or High
Scoring source	Structured automatic inference plus explicit assessor override
Confirmation	Optional audited human sign-off, subject to validation
Downstream route	Cumulative Baseline, Enhanced or Comprehensive GTSAF depth
Evidence boundary	Evidence, requests, tests and findings linked to the selected intake/system
AI support	Structured whole-system or single-dimension analysis
Workspace reporting	Portfolio roll-up without replacing system-level records

The four dimensions

ID	Dimension	Core question	High-level meaning
`dep`	Operational Dependency	What stops, degrades or becomes unsafe if the AI is wrong, unavailable or withdrawn?	The workflow is critical, tightly coupled or lacks a realistic fallback.
`act`	Action Autonomy	What can occur before authoritative human approval?	The AI can take consequential, long-running, delegated or hard-to-reverse action.
`access`	Access Scope	What can the AI identity and its tools actually reach?	Access is privileged, broad, production, cross-system, sensitive or action-capable.
`harm`	Harm Potential	What is the worst credible consequence?	Severe safety, rights, legal, financial, security, service or systemic harm is plausible.

Each dimension is assessed independently. Do not reduce one because another is low. For example, strong human approval may reduce **Action Autonomy**, but it does not make broad production access disappear; a tested fallback may reduce **Operational Dependency**, but it does not reduce the severity of a rights-affecting decision if it occurs.

The authoritative result

An ACRS result contains more than one number:

- **Dimension vector**, for example `dep:2, act:3, access:2, harm:3`.
- **Raw product**, calculated as `2 × 3 × 2 × 3 = 36`.
- **Product tier**, which is Medium for a raw product of 36.
- **Severity-floor rules**, which explain why the route may need to be raised.
- **Routed tier**, which is the authoritative Low, Medium or High assurance route.

In the example above, High harm combined with Medium-or-higher autonomy or access raises the route to **High**, even though the raw product sits at the top of the Medium product band.

> **The sentence to remember:** The product measures combined capability exposure; the severity floors prevent a dangerous combination or explicit high-risk fact from being averaged down.

Automatic inference and assessor judgement

ACRS starts from the selected system's intake facts. Gamut infers a Low, Medium or High suggestion for each dimension using structured facts and bounded contextual signals, including:

- Autonomy level and human-oversight arrangement.
- Automated-decision and high-risk flags.
- Personal and special-category data.
- Internal or external retrieval.
- Public-facing and community-impact characteristics.
- Lifecycle and deployment context.
- System purpose, users, affected people, data sources and regulatory exposure.
- Descriptions of tools, production actions, critical services and high-impact domains.

The assessor may leave the inferred level in place or record an explicit level and rationale. Explicit assessor levels take precedence in the vector. Gamut preserves provenance so an inferred score never masquerades as assessor evidence.

A lower-than-inferred score requires a rationale before confirmation. Contradictory intake facts, such as High autonomy alongside human approval before every consequential action, must also be resolved before confirmation.

Why system scope matters

ACRS is bound to a selected system intake. When the assessor changes systems:

- The vector, product, route and conclusion change to that system.
- Linked evidence, evidence requests, tests and findings change to that system.
- AI Assist uses the new system's authorised, validated context.
- Previously generated analysis is not presented as current for the new system.
- A material change to the assessment basis invalidates prior confirmation and makes old analysis stale for the current basis.

This prevents a low-risk profile, strong evidence or favourable AI analysis for one system from spilling into another.

What a complete assessment records

A defensible ACRS record should contain:

- A selected, correctly described AI system and intake.
- One level for each dimension, with visible inferred or assessor provenance.
- A rationale for each assessor override, especially any reduction.
- Resolution of contradictions and severity-floor warnings.
- Linked, reviewed evidence appropriate to the claimed level.
- Bounded test records with results and pass criteria.
- Findings for material uncertainty, control weakness or unsafe exposure.
- The authoritative vector, product tier, floor rules and routed tier.
- An accountable assessment owner.
- Confidence in the conclusion.
- A residual-risk position.
- A review date.

- Reassessment triggers.
- A concise assessor conclusion.
- Optional audited confirmation by an authorised human.

Security and assurance principles

ACRS in Gamut follows these rules:

- **Authoritative calculation:** the product, band and route are derived from the recorded basis.
- **Zero trust:** system scope, workspace access, roles and entitlements are verified for every use.
- **No entitlement spill:** routing metadata selects assurance depth; it does not grant framework, AI, model or plan access.
- **Evidence over assertion:** assessor rationale is not automatically converted into accepted evidence.
- **Defence in depth:** severe harm, autonomy and access combinations receive conservative route floors.
- **Fail-safe change handling:** material basis changes invalidate confirmation.
- **Safe testing:** tests are bounded, authorised, reversible and non-destructive.
- **Human accountability:** AI Assist cannot confirm ACRS, accept evidence or accept residual risk.

Recommended reading order

38. Agentic Trust Framework (ATF)

Complete guide to Gamut's per-agent implementation of ATF 0.9.1 and its 25 zero-trust requirements, four maturity levels, promotion gates, evidence and AI-assisted assessment.

The **Agentic Trust Framework (ATF)** is an open specification for governing autonomous AI agents using zero-trust principles. Gamut assesses its **25 canonical requirements** across five elements, maintains a maturity position for each registered agent, and uses explicit promotion gates before greater autonomy is approved.

The central question is:

> For this named agent, what authority is justified by demonstrated identity, observable behaviour, governed data, enforced boundaries and tested containment?

> **Framework credit and status:** ATF was created by **Josh Woodruff of MassiveScale.AI**. The canonical repository describes Specification **0.9.1** as a **Public Review Draft** and the conformance document as version **0.9.0**. The specification is published under **CC BY 4.0**; repository code is under Apache-2.0. Gamut implements an assessment of the public specification and is not its author. Use the canonical ATF repository, ATF site and CSA introduction as the current sources.

The canonical site and specification basis were rechecked on **21 July 2026**. Because ATF remains a Public Review Draft, assessments must display the exact specification and conformance versions rather than imply a final standard or certification.

> **Public review draft; no certification claim:** ATF may change before version 1.0. An assessment is a recorded conformance and maturity evaluation, not third-party certification, proof that an agent is safe, or permission to operate without continuous control.

Start here

If you need to…	Read
Understand all five elements and 25 requirements	Elements and requirements
Explain Intern, Junior, Senior, Principal and the five gates	Maturity and promotion
Complete an agent assessment	Per-agent assessment workflow
Know what evidence and tests support a requirement	Evidence, testing and incidents
Use per-requirement AI assistance and Privacy Mode	AI Assist and privacy
Explain reporting and relationships with ACRS, MAESTRO and GTSAF	Reporting and framework relationships
Look up labels and safe explanatory language	Reference and glossary

At a glance

Property	Gamut assessment
Canonical basis	ATF Specification 0.9.1 and Conformance 0.9.0
Status	Public Review Draft
Scope	One selected registered agent
Elements	Identity, Behavior, Data Governance, Segmentation, Incident Response
Canonical requirements	25; five per element
Maturity levels	L1 Intern, L2 Junior, L3 Senior, L4 Principal
Requirement language	MUST, SHOULD and MAY by target level
Requirement results	Fully met, Partially met, Not addressed
Depth	Ad hoc, Defined, Embedded for Fully met; Not started, In progress, Near complete otherwise
Promotion	Sequential, time-bound and subject to five explicit gates
Demotion	Available whenever trust conditions or evidence deteriorate
AI support	Per-requirement, system-scoped advisory analysis
Accountability	Human owners approve results, promotion, demotion and residual risk

The five elements

Element	Zero-trust question	What is examined
Identity	Who are you?	Identifier, credentials, ownership, purpose and capability declaration
Behavior	What are you doing?	Structured logs, attribution, baseline, anomaly detection and explainability
Data Governance	What are you consuming and producing?	Schema validation, injection defence, sensitive-data protection, output validation and lineage
Segmentation	Where can you go and what can you do?	Resource and action boundaries, rate and transaction limits, blast-radius containment
Incident Response	What happens when trust is lost?	Circuit breaking, kill switch, session revocation, rollback and graceful degradation

The elements work together. Strong identity without action boundaries still permits excessive authority; monitoring without containment only observes harm; a kill switch that has never been tested is an assertion, not reliable control.

The four maturity levels

Level	Operating model	Human involvement	Minimum time before the next promotion review
L1 Intern	Observe and report; read-only	Continuous oversight	2 weeks
L2 Junior	Recommend; impactful actions require approval	Approval before action	4 weeks
L3 Senior	Act within explicit guardrails	Post-action notification and exception oversight	8 weeks
L4 Principal	Autonomous within an approved domain	Strategic oversight and edge-case escalation	Ongoing validation

Maturity is an authority ceiling, not a reward. A target level establishes the requirements and promotion evidence that must be examined before the agent is permitted to operate at that level.

How Gamut operationalises ATF

```
Registered agent and declared operating scope
-> current and target ATF level
-> 25 requirement decisions using the level matrix
-> implementation depth, rationale and objective evidence
-> authorised tests and adverse findings
```

```
-> five promotion gates and minimum-time evidence
-> accountable promotion, hold or demotion decision
-> continuous monitoring and reassessment
```

Assessment is **per agent**. Switching agents changes the assessment basis; one agent's results do not transfer to another merely because both use the same model or platform.

What a defensible ATF record contains

- Immutable agent identity and accountable ownership.
- Purpose, capability manifest, tools, resources and action limits.
- Current and target maturity level.
- A result for every requirement relevant to the target level.
- Rationale tied to the selected agent.
- Design and operating evidence.
- Safe tests of enforcement and containment.
- Open findings, incidents and exceptions.
- Promotion-gate evidence and named approvers.
- Monitoring, review and demotion triggers.

What the assessment does not prove

It does not by itself prove:

- ATF certification or endorsement.
- That all agentic threats have been identified.
- That a documented boundary is enforced.
- That a provider or model is trustworthy.
- That an agent can safely receive the target authority.
- That a promotion is permanent.
- That AI-generated advice is an approved assessment decision.

Recommended reading journey

39. MAESTRO

Complete guide to Gamut's CSA-aligned MAESTRO threat assessment: canonical layers, 55 threats, architecture patterns, system scope, likelihood-impact scoring, evidence, testing, AI Assist and reporting.

MAESTRO (**Multi-Agent Environment, Security, Threat, Risk, and Outcome**) is a threat-modelling framework designed for the distinctive risks of agentic AI. It examines an AI system layer by layer, then traces threats that propagate across layers, agents and trust boundaries.

Gamut implements the Cloud Security Alliance catalogue published on 6 February 2025 as `CSA-2025-02-06-v1`:

- Seven canonical architectural layers.
- A dedicated cross-layer threat section.
- 50 layer-specific threats.
- Five cross-layer threats.
- Eight canonical agentic architecture patterns.
- A six-step assessment method.
- Likelihood × impact risk assessment.

- A unique assessor playbook for every threat.
- System-specific assessment records and workspace roll-up.
- Evidence, tests, findings, treatment and monitoring records.
- Validated, system-scoped AI assistance for each canonical threat.

Framework credit and source of truth: MAESTRO was introduced by **Ken Huang** through the **Cloud Security Alliance**. Gamut is not the author of MAESTRO and is not CSA-endorsed. The canonical source used for this implementation is the Cloud Security Alliance MAESTRO article.

What MAESTRO does not prove: A completed MAESTRO assessment is a threat model, not a certification or automatic compliance decision. It identifies and prioritises credible attack and failure paths. Assurance still depends on implemented controls, accepted evidence, safe testing, accountable risk treatment and human approval.

Start here

If you need to...	Read
Explain what MAESTRO is, how it differs from a control framework and what every label means	Method, concepts and terminology
Understand all seven layers and every one of the 55 canonical threats	Layers and threat catalogue
Decompose a system and select the right agentic architecture pattern	System decomposition and architecture patterns
Complete an assessment and defend the scoring	Assessment workflow and risk scoring
Know what evidence to request, how to test safely and how to treat findings	Evidence, testing and treatment
Explain what AI Assist reads, produces, secures and cannot decide	AI Assist and security
Produce reports and connect MAESTRO to GTSAF, ACRS and ATF	Reporting and governance
Follow a complete realistic assessment	Worked example
Look up identifiers, scores, labels and formulas	Reference and glossary

At a glance

Property	Gamut implementation
Catalogue version	`CSA-2025-02-06-v1`
Canonical layers	7
Cross-layer section	1
Layer-specific threats	50
Cross-layer threats	5
Total assessable threats	55
Stable identifier	`MAE-L-` or `MAE-X-`
Architecture patterns	8
Workflow steps	6
Risk dimensions	Likelihood 1-5 and impact 1-5
Raw risk	Likelihood × impact, from 1 to 25
Severity bands	Low, Moderate, Elevated, High, Critical
Layer roll-up	Highest scored threat in the layer
System roll-up	Highest scored threat for the selected system
Workspace roll-up	Systems assessed, worst case and distribution by worst-case band

The canonical architecture

Section	Canonical name	Threats	Primary subject
Layer 1	Foundation Models	7	Model behaviour, privacy, integrity, extraction and availability

Section	Canonical name	Threats	Primary subject
Layer 2	Data Operations	5	Data stores, pipelines, RAG, provenance, integrity and availability
Layer 3	Agent Frameworks	6	Framework components, APIs, dependencies, validation and control evasion
Layer 4	Deployment & Infrastructure	6	Images, orchestration, IaC, compute, networks and lateral movement
Layer 5	Evaluation & Observability	6	Metrics, evaluators, telemetry, detection integrity and monitoring confidentiality
Layer 6	Security & Compliance	7	The vertical security layer, especially AI agents performing security functions
Layer 7	Agent Ecosystem	13	Agents, identities, tools, registries, discovery, markets and business integrations
Cross-layer	Cross-Layer Threats	5	Attack chains and cascading failures spanning two or more layers

Layer 6 is **vertical**: it cuts across the rest of the architecture. The cross-layer section is not an eighth architectural layer; it records threats that exploit relationships between layers.

How MAESTRO operates in Gamut

1. Select the AI system being assessed, or deliberately use workspace scope.
2. Decompose its components, capabilities, tools, goals, constraints and interactions.
3. Select the closest canonical architecture pattern.
4. Confirm which assets and pathways exist in each layer.
5. Tailor each applicable canonical threat into a system-specific scenario.
6. Score likelihood and impact.
7. Record current controls, treatment owner, target date, monitoring and reassessment triggers.
8. Link evidence, tests and findings.
9. Model cross-layer attack chains.
10. Review the highest risks, generate reporting and obtain the authorised human risk decision.

The assessment record for one threat

A defensible threat record should answer:

- Why is the threat applicable to this system?
- Which actor, capability or failure can initiate it?
- Which assets and trust boundaries form the attack path?
- What is the credible business, safety, privacy or mission outcome?
- What evidence demonstrates the current controls?
- What bounded test was performed, with what pass criteria and safety limits?
- Is the score inherent or residual?
- Who owns treatment, by when?
- Which signals will reveal attempted exploitation or control degradation?
- Which changes require reassessment?
- Which other layers or threats can form a cascade?

Security principles

MAESTRO in Gamut is designed around:

- **Defence in depth:** no single control is assumed to break every path.
- **Zero trust:** identity, workload, data and agent claims are verified at each boundary.
- **Least privilege:** agents, tools and service identities receive only necessary authority.
- **Fail-safe operation:** uncertainty or control failure should reduce capability, not silently expand it.
- **Evidence over assertion:** a confident narrative is not accepted as proof.
- **Safe testing:** tests are bounded, authorised, reversible and non-destructive.
- **Human accountability:** AI suggestions cannot approve evidence, accept risk or sign off the assessment.

Recommended reading order

40. AI Assist and security

Public guide to ACRS whole-system and dimension AI assistance, system scope, information considered, Privacy Mode, saved outputs and human scoring accountability.

ACRS AI Assist supports analysis of the selected system's four-dimension capability-risk record. It can analyse the whole vector or one dimension. It cannot change automatic scoring, severity floors, assessor overrides, confirmation or residual-risk decisions.

System scope

AI Assist requires a selected system intake. The panel identifies the system and analysis scope. When another system is selected:

- The active ACRS vector and route change.
- Evidence, tests and findings change to that system.
- Prior analysis is not presented as current for the new system.

An unscoped workspace scratch vector is for exploration and is not an appropriate basis for system-level AI analysis.

Full-context information considered

Subject to authorised access, analysis may consider:

- Selected system identity, purpose and context.
- Dependency, Action Autonomy, Access Scope and Harm Potential levels.
- Whether each level is inferred or assessor-selected.
- Assessor override rationale.
- Product band, severity-floor reasons and routed band.
- Contradictions and unresolved uncertainty.
- Authorised evidence status.
- Tests and results.
- Findings.
- Assessment owner, confidence, residual-risk position and reassessment triggers.
- Relevant control and routing context.

Whole-system output

Whole-system analysis can provide:

- Recommended vector and reasoning.

- Product and severity-floor interpretation.
- Contradictions and missing facts.
- Evidence and test gaps.
- Priority governance actions.
- Residual-risk observations.
- Monitoring and reassessment triggers.
- Draft assessor conclusion for review.

Dimension output

Dimension-specific analysis should focus on:

- Current effective capability, not intended limits alone.
- Evidence supporting the current level.
- Facts supporting a higher or lower recommendation.
- Safe validation tests.
- Uncertainty and challenge questions.
- Effect on the overall route.

Privacy Mode

The **Privacy Mode** checkbox beside AI Assist sends the minimum structural ACRS state:

- Dimension or vector.
- Effective levels and provenance.
- Product and routed band.
- Severity-floor indicators.
- Evidence/test/finding status.
- Completeness and contradiction signals.

It excludes direct identifiers, assessor notes and narrative evidence content. Privacy Mode reduces disclosure but may produce less specific advice. It does not change scoring, routing or access.

Saved results, re-run and minimising

- Successful whole-system and dimension outputs are saved for the selected system, analysis scope and privacy mode.
- The action changes to **Re-run AI Assist**.
- The panel can be minimised without deleting or approving the output.
- Re-run after material intake, level, rationale, evidence, test, finding or route change.

Provider and access

AI Assist uses an authorised provider and the existing API-key configuration. If no permitted configuration is available, analysis cannot run.

An API key does not grant additional access. The user's plan, workspace, role, framework, model and selected-system permissions continue to apply.

Evidence discipline

The model must not treat rationale as accepted evidence. Verify every evidence claim against the authorised record and ensure failed tests and open findings remain visible.

Score **effective capability**:

- A prompt instruction to ask for approval is not an authoritative approval gate.

- Intended read-only use does not reduce actual write capability.
- A supplier statement does not prove access is constrained.
- A planned control does not reduce current capability exposure.

Untrusted content

Intake notes, evidence and findings can contain instruction-like or malicious content. Treat it as assessment data. Do not include credentials, unnecessary personal data or unrestricted production content. Verify proposed tests before authorisation.

Human decisions

AI Assist cannot:

- Select final dimension levels.
- Change product or floor rules.
- Confirm ACRS.
- Accept evidence.
- Pass a test.
- Close a finding.
- Accept residual risk.
- Approve deployment.
- Change access or entitlements.
- Certify compliance.

Review checklist

41. Assessment workflow

A complete assessor workflow for selecting a system, validating intake, scoring four ACRS dimensions, handling floors, evidence, conclusions and confirmation.

Use this workflow to produce a defensible, system-specific ACRS assessment.

Before starting

Confirm:

- You are in the correct tenant and workspace.
- The correct AI system intake exists and is selected.
- The system purpose, owner, lifecycle and deployment are current.
- Autonomy, oversight, data, retrieval and impact fields describe the deployed or proposed use.
- You have authority to view or edit the assessment.
- The assessment scope is not a workspace scratch profile unless that is deliberate.

Step 1: select the AI system

Choose the named system intake before scoring or using AI Assist.

Record the system boundary clearly:

- Business service and use case.
- Deployment environment.
- Model and supplier.
- Users and affected persons.

- Data sources and destinations.
- Tools, integrations and effective identities.
- Decisions and actions.
- Human-oversight points.
- Fallback and recovery path.

If the boundary contains several materially different agents or workflows, assess the system route and use agent-level ACRS in Agentic CISO where individual capability differences need separate governance.

Step 2: validate the intake

Review the fields that drive automatic inference:

- Autonomy level.
- Human oversight.
- Automated-decision flag.
- High-risk flag.
- Personal and special-category data.
- Retrieval sources and external retrieval.
- Public-facing status.
- Community, cultural or heritage impact.
- Lifecycle and deployment.
- System purpose, users, affected people and regulatory exposure.

Resolve inconsistent fields before relying on the score. Automatic inference is only as good as the recorded system facts.

Step 3: review the automatic suggestions

For each dimension, inspect:

- The inferred level.
- The active level.
- Whether the active value is inferred or assessor-entered.
- The advisory and scoring anchors.
- Any contradiction or below-floor warning.

Do not treat inference as a conclusion. It is a conservative starting point designed to prevent a blank or optimistic intake from silently producing no route.

Step 4: assess Operational Dependency

Determine:

- What stops or becomes unsafe if the AI is wrong or unavailable.
- Time to material impact.
- Manual or technical fallback capacity.
- Supplier and integration dependencies.
- Recovery objective and restoration completeness.
- Whether fallback works at realistic demand.

Use the dimension playbook to collect evidence and run a bounded continuity exercise.

Step 5: assess Action Autonomy

Trace the workflow from trigger to terminal effect:

- Decisions and actions.
- Approval gates.
- Alternate APIs and tools.
- Retries and long-running tasks.
- Delegation and sub-agents.
- Kill switch, rollback and compensation.
- Human detection and intervention time.

Approval must occur before the consequential effect to support a Low conclusion.

Step 6: assess Access Scope

Build an effective permission path:

- Initiating user.
- AI or service principal.
- Tool or connector.
- Target system.
- Data classification.
- Read, write, execute, delete or send authority.
- Credential scope and lifetime.
- Tenant, network and environment boundaries.

Score what is technically reachable, not what the prompt says the AI should use.

Step 7: assess Harm Potential

Construct credible adverse scenarios:

- Ordinary error.
- Prompt injection or manipulation.
- Compromise or malicious use.
- Correlated or repeated failure.
- Scale and propagation.
- Detection and containment.
- Reversibility, appeal, redress and recovery.

Harm describes consequence severity. Low probability does not turn severe harm into Low Harm.

Step 8: justify assessor overrides

For every explicit level:

- State the boundary assessed.
- Cite the facts and evidence considered.
- Explain why the selected anchor fits.
- Identify uncertainty.
- Record the change that would make the score invalid.

For a lower-than-inferred level, explain why the inference signal does not represent the effective deployed capability and identify the technical or operational evidence supporting the reduction.

Step 9: review product and severity floors

Check:

- Vector.
- Raw product.
- Product tier.
- Every severity-floor reason.
- Authoritative routed tier.

Do not overwrite a High routed tier with a Medium product tier in narrative or reporting.

Step 10: link evidence, tests and findings

For each dimension:

- Link objective evidence.
- Create evidence requests where records are missing.
- Record bounded tests and pass criteria.
- Raise findings for material gaps, failed tests or unsupported assumptions.

Assessor rationale remains separate from evidence. A detailed narrative does not become accepted evidence automatically.

Step 11: use AI Assist where helpful

AI Assist can analyse:

- The whole selected system; or
- One selected dimension.

Before using the result, verify:

- The displayed system name and reference.
- The current vector and basis.
- The evidence records the model was allowed to treat as accepted.
- The recommended score against the authoritative route.
- The proposed test's safety limits.

AI output is advisory and requires human approval.

Step 12: record the structured assessor conclusion

The conclusion panel requires:

Field	What to record
Assessment owner	The accountable person responsible for the ACRS conclusion.
Confidence	Low, Medium or High confidence in the assessment basis.
Residual risk	The remaining risk after current controls, limitations and uncertainties.
Review due	A date proportionate to exposure and change rate.
Assessor conclusion	A concise, defensible statement of the vector, route and decision relevance.
Reassessment triggers	Specific changes or events that require review.

The conclusion is human-authored and remains separate from AI output and evidence acceptance.

Example conclusion

The system is assessed as `dep:2, act:3, access:2, harm:3`, raw product 36. The authoritative route is High because severe harm is combined with consequential autonomy and material access. Current approval and rollback controls reduce residual risk but do not change the capability exposure. Production expansion is conditional on closure of the approval-bypass finding and a passed rollback exercise.

Step 13: confirm the assessment

Confirm only when:

- Contradictions are resolved.
- Lower-than-inferred overrides are justified.
- The system scope is correct.
- Evidence and tests are sufficient for the claimed confidence.
- Findings and residual risk are visible.
- The accountable human accepts the recorded conclusion.

Confirmation signs the current assessment basis. It does not accept deployment or residual risk unless the organisation's separate governance process explicitly records those decisions.

Step 14: route and continue assurance work

Use the accepted ACRS route to:

- Set proportionate GTSAF assurance depth without changing factual control applicability.
- Prioritise dimension-relevant controls.
- Set evidence and test rigour.
- Inform MAESTRO threat modelling.
- Set Gateway approval and runtime-policy expectations.
- Inform ATF autonomy decisions.
- Establish monitoring and reassessment cadence.

Resetting assessor overrides

Reset assessor overrides clears:

- Explicit dimension levels.
- Dimension rationales.
- Structured conclusion fields.
- Confirmation tied to those overrides.

It does not disable automatic system-generated scoring. Gamut recalculates from the current intake facts.

Quality checklist

42. Dimension assessor playbooks

Deep, assessor-ready guidance for Operational Dependency, Action Autonomy, Access Scope and Harm Potential, including evidence, safe tests, red flags and scoring anchors.

This page is designed so an assessor can complete most of the dimension analysis without leaving the ACRS screen. Use the playbook for each dimension alongside the selected system's actual intake, architecture and operating evidence.

How to use the playbooks

For each dimension:

1. Fix the boundary and deployment being assessed.
2. Ask the diagnostic questions.
3. Inspect objective evidence.
4. Perform a bounded test or exercise where appropriate.

5. Compare the result with the scoring anchors.
6. Record the rationale, uncertainty and reassessment trigger.
7. Raise findings where the claimed boundary is not demonstrated.

Do not score from policy intent alone. Assess effective capability and real operating conditions.

Operational Dependency (dep)

What it measures

Operational Dependency asks:

> How hard would it be for the organisation, customers or affected people to continue safely if the AI were wrong, unavailable, degraded or withdrawn?

It includes dependency on the full service chain, not only the model:

- Model or provider.
- Prompts, policies and configurations.
- Data and vector indexes.
- Identity and secrets.
- Connectors and APIs.
- Infrastructure and networks.
- Monitoring and human operating capacity.

Scope

Identify:

- The exact business service, decision path or customer journey.
- Regulated, safety, security or critical-service functions.
- Upstream and downstream dependencies.
- Operating hours and service objectives.
- Supplier concentration and substitution options.
- Manual fallback and safe-degradation paths.
- Recovery-time and recovery-point objectives.
- Population, transaction volume and decision volume affected.

Assessment method

Walk the process in four states:

1. Correct AI operation.
2. Degraded or intermittently wrong output.
3. Complete outage.
4. Provider, model or critical integration withdrawal.

Measure:

- Time to business, safety, legal or customer impact.
- Backlog growth.
- Decision-quality degradation.
- Safe operating duration.
- Manual capacity at realistic volume.
- Recovery dependencies.

- The point at which the fallback ceases to meet approved tolerances.

Diagnostic questions

- What outcome stops, becomes delayed or becomes unsafe?
- How long can the process continue before a material threshold is breached?
- Is the fallback documented, staffed, accessible and tested?
- Has fallback capacity been tested at realistic demand?
- Which model, data, identity, integration, infrastructure or supplier failures remove the service?
- Can the organisation substitute another model or provider without losing necessary policy, configuration, data or assurance?
- Who can suspend the AI?
- Who owns continuity, recovery and risk acceptance?
- Does the business plan or customer commitment assume continuous AI availability?
- Can the workflow operate safely when the AI produces plausible but wrong output rather than becoming fully unavailable?

Evidence to inspect

- Current process maps and service dependency diagrams.
- Business-impact analysis.
- Service-level objectives and error budgets.
- Manual fallback procedures and staffing plans.
- Capacity evidence for fallback.
- Provider and integration SLAs.
- Substitution and exit plans.
- Recovery runbooks.
- Backup and restore evidence for configuration, data, indexes, prompts and policies.
- Failover, rollback or continuity exercise records.
- Incident and near-miss records.
- Named service, continuity and recovery owners.

Policies and diagrams support the analysis but do not prove fallback operation. Seek sampled records or exercise results.

Safe test

Prefer:

- Tabletop exercise.
- Feature-flag disablement.
- Shadow workflow.
- Controlled failover.
- Non-production recovery exercise.

Define before testing:

- Customer, legal and safety protections.
- Stop conditions.
- Data and log preservation.
- Rollback steps.
- Service-owner approval.
- Maximum outage or degraded-operation window.

Pass criteria

- Fallback starts within the approved objective.
- Required decisions remain safe, attributable and compliant.
- Backlog and quality remain within approved tolerance.
- Recovery covers data, configuration, identity, integrations, policy and monitoring.
- Restoration is verified, not merely declared complete.
- Exceptions create findings with owners and dates.

Scoring anchors

Level	Anchor
1 - Low	The AI is genuinely optional or readily substitutable, and a tested fallback sustains safe operation at realistic demand.
2 - Medium	The AI is materially important and disruption is significant, but a demonstrated fallback or safe-degradation path keeps the service within approved tolerance.
3 - High	The AI is critical, regulated, safety, security, financial or customer-impacting; wrong output or loss rapidly breaches tolerance; or no practical fallback exists.

Borderline decisions

- An untested fallback prevents a confident Low conclusion.
- A fallback that works only at trivial volume is not a practical fallback.
- Continuing service by accepting material safety, legal, integrity or customer degradation supports High, not Medium.
- A supplier SLA reduces expected outage frequency; it does not reduce organisational dependency unless substitution or recovery is demonstrated.
- "Humans can do it manually" is not evidence unless capacity, access, competence and time have been tested.

Red flags

- Production objectives assume AI availability while the assessment calls it optional.
- Fallback exists only in a document.
- One individual or supplier is the only workaround.
- Recovery excludes prompts, policies, indexes, credentials or integrations.
- The workflow cannot detect plausible but wrong output.
- Adoption or transaction growth does not trigger reassessment.
- No owner can authoritatively suspend or restore the service.

Control and monitoring implications

Prioritise:

- Service ownership and dependency inventory.
- Safe degradation.
- Capacity-tested fallback.
- Provider or model substitution.
- Configuration and data recovery.
- Rollback and restoration exercises.
- Change-triggered reassessment.

Monitor:

- Availability and error budget.
- Fallback activation time and capacity.

- Backlog and decision-quality degradation.
- Supplier, model, data and integration health.
- Recovery-objective breaches.
- Adoption and volume growth.

Finding triggers

- No viable critical-path fallback.
- Untested restoration or substitution.
- Dependency level unsupported by operating evidence.
- Service commitments contradict an optional-use claim.
- Incident or near miss without reassessment.

Completion record

Record the service boundary, dependency graph, failure horizon, fallback owner and capacity, recovery objective, evidence sampled, exercise result, level, rationale, confidence, exceptions, treatment owner, review date, monitoring and reassessment triggers.

Action Autonomy (`act`)

What it measures

Action Autonomy asks:

> What can the AI decide, initiate, execute, repeat or delegate before authoritative human approval?

The important boundary is not whether a human is "involved". It is whether the consequential effect can occur before an effective approval or intervention.

Scope

Inventory:

- Decisions and recommendations.
- Tool calls.
- Writes, executions, deletions and external communications.
- Transactions and financial actions.
- Workflow continuation and retries.
- Delegation and sub-agents.
- Background and scheduled tasks.
- Approval gates.
- Rollback and compensation.
- Kill switches and circuit breakers.
- Alternate APIs or paths.

Assessment method

Trace representative and worst-case tasks from trigger to terminal effect. Determine:

- What the AI chooses.
- What it initiates.
- What it can repeat.
- What it can delegate.
- What it can make difficult to detect.
- What becomes irreversible.

- Whether alternate tools, retries, memory or sub-agents bypass the intended gate.

Diagnostic questions

- What consequential effect can occur before human approval?
- Is the gate enforced outside the model?
- Does every action path use the same authoritative gate?
- Can denial be bypassed by retry, reframing, delegation or another tool?
- Can a human detect and stop the action within the practical response window?
- Can the effect be reversed or compensated?
- What changes at production volume?
- Are step, time, spend, velocity, retry and delegation limits enforced?
- Can the AI change its own goals, instructions, permissions, tools or agents?
- What happens if the AI receives malicious retrieved content or tool output?

Evidence to inspect

- Deployed tool manifest and effective action permissions.
- Runtime policy configuration.
- Approval and denial logs.
- Representative action traces.
- Alternate-path and delegation traces.
- Retry and rate-limit configuration.
- Spend, time and step ceilings.
- Kill-switch and circuit-breaker tests.
- Rollback or compensation exercises.
- Human-response records.
- Change and release records for autonomy-related configuration.

Screenshots of the intended workflow do not prove that bypasses are prevented.

Safe test

Use:

- Inert or simulated tools.
- Synthetic identities and data.
- Blocked external side effects.
- Low spend, time, step and velocity limits.
- Reversible fixtures.
- Explicit test authority and stop conditions.

Do not exercise consequential production actions without written authority, safety controls and a tested rollback.

Pass criteria

- Every consequential action reaches an authoritative gate.
- Denied actions cannot execute through retry, reframing, delegation or alternate paths.
- Human intervention works within the documented window.
- Rollback or containment is demonstrated where feasible.
- Attempts and decisions create attributable, tamper-resistant telemetry.
- Unbounded execution paths are absent or explicitly treated.

Scoring anchors

Level	Anchor
1 - Low	Advisory or assistive only, with enforceable human authorisation before every consequential action.
2 - Medium	Bounded, monitored and reversible action within preset limits, with effective intervention before major harm.
3 - High	Consequential pre-approval authority, bypassable gating, long-horizon or delegated autonomy, or hard-to-detect or irreversible action.

Borderline decisions

- Human-on-the-loop is not Low merely because notifications exist.
- If action precedes reliable intervention, score at least Medium.
- Material external, financial, code, identity, security, safety or rights-affecting action before approval is High.
- A prompt instruction to request approval is not an authoritative gate.
- A reversible action may still be High where scale, speed or propagation defeats practical rollback.

Red flags

- Oversight is described but no technical gate exists.
- Approval occurs after the effect.
- Prompt instructions substitute for runtime enforcement.
- Retry, alternate tools or delegation bypass denial.
- Kill switch or rollback is unclear or untested.
- Step, spend, time, velocity or delegation is unbounded.
- Irreversible communications, payments, code, identity or rights effects lack prior approval.

Control and monitoring implications

Prioritise:

- Authoritative policy enforcement.
- Least-authority tools.
- Dual control for high-risk actions.
- Step, spend, time, retry and velocity ceilings.
- Delegation constraints.
- Kill switches and circuit breakers.
- Rollback and compensation.
- Attributable action logging.

Monitor:

- Allowed, denied and overridden actions.
- Gate-bypass attempts.
- Delegation depth and task duration.
- Retry, spend and velocity anomalies.
- Human response time.
- Rollback success and residual side effects.

Finding triggers

- Prompt-only approval.
- Consequential pre-approval action.

- Untested kill switch or rollback.
- Unbounded retry, spend, duration or delegation.
- Missing action attribution.
- Oversight window shorter than practical response time.

Completion record

Record the action inventory, authoritative gates, bypass analysis, representative traces, worst credible action, reversibility, test result, level, rationale, confidence, exceptions, treatment owner, monitoring and reassessment triggers.

Access Scope (`access`)

What it measures

Access Scope asks:

> Which data, tools, identities, systems and actions are effectively reachable by the AI and its execution chain?

Assess effective permission, including indirect and delegated paths, not only the intended request.

Scope

Map:

- User identities.
- Service and non-human identities.
- Delegated and cached credentials.
- Supplier identities.
- Data stores and vector indexes.
- APIs and tools.
- Secrets and memory.
- Networks and endpoints.
- Tenants and environments.
- Repositories and production targets.
- Indirect and chained access.

Assessment method

Build an effective permission graph:

```
initiating user -> AI principal -> tool or connector -> target -> permitted effect
```

For every material path, determine:

- Data sensitivity.
- Read, write, execute, delete and send authority.
- Credential source, scope and lifetime.
- Target-side authorisation.
- User and purpose enforcement.
- Tenant and environment separation.
- Lateral movement and egress.
- Delegation and privilege escalation.

Diagnostic questions

- Which principal authenticates each retrieval and action?
- Does the AI inherit the user's rights or use a broad service identity?

- Which privileged, production, financial, security, identity, code, personal or confidential resources are reachable?
- Is permission enforced by the target or only described in the tool schema?
- Can memory, logs, errors or tool output expose data or secrets?
- How are credentials scoped, rotated, revoked and kept outside model context?
- Are source ACLs preserved during retrieval?
- Can the agent cross tenants, environments or customer boundaries?
- Can read-only access still enable bulk sensitive extraction?
- Can a tool accept arbitrary endpoints, paths, queries, commands or recipients?

Evidence to inspect

- Effective IAM and target-side policy.
- Token and OAuth scopes.
- Credential-vault records and lifetime.
- Network and egress policy.
- Data classification and flow maps.
- Tool allowlists.
- Tenant and environment isolation tests.
- Denial records.
- Access reviews.
- Action traces tied to the deployed principal.
- Secret scanning and credential-rotation evidence.

Role names and architecture diagrams are not sufficient without effective-policy or denial evidence.

Safe test

Use:

- Synthetic identities.
- Canary records.
- Non-production targets.
- Read-only policy inspection.
- Restricted egress.
- No real secrets.
- Explicit authority for tenant, privilege or isolation testing.

Stop immediately if unexpected access succeeds.

Pass criteria

- Effective access matches the documented minimum.
- Unauthorised data, tenants, actions, paths and environments are denied at an authoritative boundary.
- Credentials remain outside model-visible context.
- Credentials can be revoked quickly.
- Logs attribute user, AI principal, tool, target, decision and outcome.
- Excess access becomes a finding.

Scoring anchors

Level	Anchor
1 - Low	Narrow, target-enforced read-only access to low-sensitivity data in a sandbox, with no privileged or consequential capability.
2 - Medium	Sensitive data or constrained operational tools with enforced scope, monitoring and limited write or action.
3 - High	Privileged, broad, production, cross-system, cross-tenant, financial, identity, security, code, secret, write, execute or delete access.

Borderline decisions

- Read-only is not automatically Low.
- Bulk personal data, confidential records, source code, security telemetry, secrets or cross-tenant retrieval can justify Medium or High.
- Reachable privileged tools are High even if prompts say not to use them.
- A short-lived credential can reduce exposure but does not make broad production authority narrow.
- A tool allowlist does not prove target-side authorisation.

Red flags

- A broad shared service account serves many users or tenants.
- Tool schemas are treated as authorisation.
- Arbitrary endpoints, queries, paths, recipients or commands are possible.
- Credentials are standing or model-visible.
- Retrieval ignores source ACLs or purpose.
- Production and non-production boundaries are shared.
- Indirect, memory, retrieval or delegated access is not reviewed.
- No evidence shows that denied access actually fails.

Control and monitoring implications

Prioritise:

- Dedicated non-human identity.
- Least privilege and just-in-time access.
- Target-side authorisation.
- Tenant and environment isolation.
- Credential vault and short-lived tokens.
- Tool and destination allowlists.
- Data-loss prevention and egress control.
- Periodic access review.

Monitor:

- Permission and policy changes.
- Denied access and escalation attempts.
- Sensitive retrieval and export volume.
- Cross-tenant or unusual destinations.
- Credential issuance and revocation.
- New tools, endpoints and delegated identities.

Finding triggers

- Broad shared identity.

- Standing or model-visible credential.
- Untested tenant or production isolation.
- No target-side authorisation.
- Score based only on intended use.
- No review of indirect or delegated access.

Completion record

Record the permission graph, principals, sensitive targets, effective actions, credential custody, tenant and environment boundaries, evidence, test result, level, rationale, confidence, excess-access findings, treatment owner, monitoring and reassessment triggers.

Harm Potential (`harm`)

What it measures

Harm Potential asks:

> What is the worst credible consequence if the AI is wrong, manipulated, unavailable, misused or compromised?

It examines severity, scale, duration, propagation, detectability, reversibility and affected parties.

Scope

Consider harm to:

- Individuals and groups.
- Safety and health.
- Rights, privacy and dignity.
- Customers and workers.
- Security and identity.
- Operations and critical services.
- Finances and assets.
- Legal and contractual duties.
- Culture and communities.
- Environment and suppliers.
- Recovery and long-term trust.

Assessment method

Construct system-specific scenarios using the actual dependency, autonomy, access, deployment and affected population.

Trace:

1. Initiating error, manipulation, misuse or compromise.
2. Decision or action path.
3. Propagation and scale.
4. Detection.
5. Containment.
6. Notification.
7. Appeal, correction, redress or recovery.
8. Residual harm.

Use the worst credible outcome, not the average model error.

Diagnostic questions

- Who or what can be harmed?
- Which decision or action creates the harm?
- Could safety, rights, livelihood, eligibility, credit, employment, education, healthcare, legal position, identity, security, culture or essential services be affected?
- How many people, systems, transactions or decisions could be affected before detection?
- Is the outcome reversible in practice?
- Can affected people appeal or obtain redress?
- Which notification, legal, contractual or executive duties are triggered?
- What changes under intentional manipulation?
- Could repeated small harms become cumulative or systemic?
- Does detection depend on an external complaint after damage?

Evidence to inspect

- System-specific impact assessments.
- Data-protection impact assessments.
- Safety, rights or equality reviews.
- Abuse cases and threat models.
- Incidents and near misses.
- Loss and service-disruption models.
- Affected-person and vulnerable-group analysis.
- Appeal, redress and recovery procedures.
- Notification thresholds.
- Independent challenge or ethics review.

Generic principles or a supplier's marketing assurance are not sufficient.

Safe test

Prefer:

- Tabletop exercise.
- Simulation.
- Synthetic populations and records.
- Bounded red-team scenarios.

Do not expose real people to discriminatory, unsafe, deceptive, rights-affecting, financial, medical, employment, legal or public-service decisions for testing.

Pass criteria

- Scenarios identify affected parties, propagation, detection, containment, remedy and residual harm.
- Controls act before approved harm thresholds.
- Appeal, correction, notification and recovery are demonstrated where applicable.
- Severe or unacceptable-risk indicators reach independent accountable reviewers.
- Weaknesses create treatment, restrictions, findings or a pause decision.

Scoring anchors

Level	Anchor
1 - Low	Minor, local, promptly detectable and readily reversible inconvenience or rework, with no material rights, safety, legal, financial, security, customer, cultural or service impact.

Level	Anchor
2 - Medium	Material but bounded and recoverable harm requiring formal response.
3 - High	Severe, systemic, safety, rights, critical-service, major legal or financial, irreversible, widely propagated, hard-to-detect or lasting harm.

Borderline decisions

- High severity remains High even when probability is low; probability belongs in the wider risk decision.
- If correction depends on detection that is unlikely to occur, the outcome is not readily reversible.
- An appeal process that cannot restore lost opportunity, health, safety or reputation may not make the harm reversible.
- Small individual losses can become High where scale or repeated automated decisions produce systemic harm.
- Model accuracy does not determine Harm level by itself.

Red flags

- Harm is scored from accuracy rather than consequence.
- Only direct financial loss is considered.
- Affected persons or vulnerable groups are undefined.
- Reversibility assumes records, messages or learned state can be recalled when they cannot.
- No abuse, compromise, scale or correlated-failure scenario exists.
- Rights or safety concerns are treated as ordinary product risk.
- Detection depends on complaint after damage.

Control and monitoring implications

Prioritise:

- Impact and unacceptable-risk assessment.
- Safety and rights review.
- Human oversight and contestability.
- Abuse prevention.
- Blast-radius limits.
- Incident response and notification.
- Redress and recovery.
- Independent assurance and executive decision.

Monitor:

- Adverse outcomes, complaints and appeals.
- Safety, rights, fairness and protected-group indicators.
- Loss, disruption and affected-population scale.
- Detection and containment time.
- Correction and recovery success.
- Regulatory, deployment and population changes.

Finding triggers

- No affected-person or consequence analysis.
- Severe harm scored below High without a defensible boundary.
- No appeal, redress, containment or recovery.
- Unacceptable-risk indicator without independent review.
- No monitoring capable of detecting the scenario.

- Incident or near miss without reassessment.

Completion record

Record affected parties, adverse scenarios, scale, duration, detectability, reversibility, legal and rights implications, evidence, tabletop result, level, rationale, confidence, unacceptable-risk screen, treatment or restriction, accountable owner, monitoring and reassessment triggers.

43. Evidence, testing and findings

How to collect, accept and scope ACRS evidence, perform safe dimension-specific tests and raise findings without confusing narrative with proof.

ACRS is a capability-risk classification, but a defensible classification still needs evidence. The evidence should demonstrate the actual boundary used to select each dimension level.

Narrative is not evidence

An assessor rationale explains judgement. It does not automatically prove:

- Effective permissions.
- Human approval.
- Fallback capacity.
- Kill-switch operation.
- Tenant isolation.
- Reversibility.
- Harm containment.

Gamut keeps rationale separate from objective evidence. Saving a rationale does not create an accepted or Strong evidence record.

System-linked records

ACRS evidence records are linked to the selected scope through:

- Intake identifier.
- System identifier.
- System name.
- Framework.
- Dimension identifier.

This linkage applies to:

- Evidence.
- Evidence requests.
- Control tests.
- Findings.

When the system changes, only records belonging to the new scope should inform the assessment or AI analysis.

Evidence lifecycle

A practical lifecycle is:

1. Identify the claim the evidence must support.
2. Create an evidence request where necessary.
3. Collect or link the artefact.
4. Verify system, environment and time-period relevance.

5 Review quality and exceptions.

6 Record the reviewed date.

7 Accept, reject, expire or request correction.

8 Reassess after material change.

Acceptance safeguards

Gamut prevents an evidence request from being overstated as accepted, reviewed, closed or Strong unless it has:

- File or artefact references; and
- A reviewed date.

These are minimum integrity checks, not a complete quality judgement. The reviewer must still determine whether the artefact is authentic, current, complete, system-specific and sufficient.

Evidence quality criteria

Assess:

Criterion	Question
Scope	Does it relate to the selected system, environment, identity and workflow?
Authenticity	Is the source authoritative and protected from unauthorised change?
Currency	Does it cover the current deployed configuration and relevant period?
Completeness	Does it include the full path, population or exception set?
Independence	Was it produced or reviewed by an appropriately independent party?
Repeatability	Can the observed result be reproduced?
Traceability	Can the evidence be connected to a dimension claim, test or conclusion?
Exception visibility	Are failures, exclusions and limitations visible?

Evidence by dimension

Operational Dependency

Useful evidence:

- Business-impact analysis.
- Dependency maps.
- Fallback procedures.
- Capacity and staffing records.
- Recovery and substitution plans.
- Failover or continuity exercise results.
- Incident records.

Weak evidence:

- "The team can work manually."
- Supplier SLA without fallback proof.
- Untested recovery documentation.
- A process map that omits integrations and data.

Action Autonomy

Useful evidence:

- Deployed policy and approval configuration.
- Tool manifests.

- Approval, denial and action logs.
- Alternate-path and delegation tests.
- Kill-switch and rollback exercises.
- Enforced time, spend, step and retry limits.

Weak evidence:

- Prompt text saying "ask a human".
- A screenshot of the intended workflow.
- Post-action notification.
- A policy document with no runtime enforcement.

Access Scope

Useful evidence:

- Effective IAM and target-side policy.
- Token scope and lifetime.
- Credential-vault records.
- Denial and isolation tests.
- Data-flow and classification records.
- Access reviews.
- Attributable action traces.

Weak evidence:

- Role name without effective permissions.
- Tool schema treated as authorisation.
- An architecture diagram with no policy evidence.
- Intended least privilege with broad production credentials still active.

Harm Potential

Useful evidence:

- Impact, safety, rights and privacy assessments.
- Abuse cases.
- Incident and near-miss records.
- Affected-person analysis.
- Loss and service-disruption models.
- Appeal, redress, notification and recovery exercises.
- Independent challenge.

Weak evidence:

- Generic ethics principles.
- Accuracy benchmark alone.
- "No incidents have occurred."
- Supplier assurance unrelated to the use context.

Testing principles

Every test should define:

- Objective.

- Scope and environment.
- Preconditions.
- Authorisation.
- Test data and identities.
- Steps.
- Expected result.
- Pass criteria.
- Safety limits and stop conditions.
- Rollback or recovery.
- Actual result.
- Exceptions and evidence references.
- Tester and date.

Safe testing by dimension

Dimension	Preferred bounded test
Dependency	Tabletop, controlled failover, feature-flag disablement, shadow fallback or non-production restoration.
Action	Inert-tool action trace, approval-bypass test, retry/delegation test, kill-switch or rollback exercise.
Access	Synthetic-identity permission-path test, canary retrieval, target-side denial or tenant-isolation test.
Harm	Tabletop, simulation, synthetic population or bounded adverse-scenario exercise.

Do not create real harmful outcomes merely to prove a risk. Tests involving production interruption, privilege, tenants, payments, code, safety, employment, health, legal position or public-service decisions require explicit authority and stronger safeguards.

Test result interpretation

Passed

A pass means the defined pass criteria were met in the tested scope. It does not prove:

- Every alternate path is safe.
- Future configurations remain safe.
- The dimension must be Low.
- Residual risk is accepted.

Failed

A failed test should:

- Prevent reliance on the claimed boundary.
- Inform the dimension rationale.
- Create or update a finding.
- Identify immediate containment.
- Assign remediation and retest.
- Trigger route or confirmation review where material.

Partial or inconclusive

Do not convert uncertainty into assurance. Record:

- What was tested.
- What remains unknown.
- Why the test was incomplete.
- The conservative scoring consequence.
- The next test or evidence request.

Findings

Raise a finding when:

- A claimed Low or Medium boundary is not demonstrated.
- Effective capability exceeds intended capability.
- A contradiction cannot be resolved.
- A test fails.
- Evidence is missing, stale, rejected or system-mismatched.
- A severity-floor condition is ignored in governance decisions.
- The route is inconsistent with the deployed system.
- Monitoring cannot detect the assessed failure scenario.
- A material basis change has not been reassessed.

Finding content

A useful finding includes:

- Selected system and dimension.
- Observed condition.
- Expected boundary or control.
- Evidence and test references.
- Risk and credible consequence.
- Immediate containment.
- Root cause where known.
- Owner.
- Due date.
- Closure criteria.
- Retest requirement.
- Effect on ACRS confidence, confirmation or route.

Example

Action Autonomy: approval bypass through alternate tool path. The primary payment workflow requires approval, but the agent can invoke the settlement API directly through a secondary tool. The denial test failed. Until the alternate path is removed and retested, Action Autonomy remains High and the High ACRS route is authoritative. Disable the secondary tool in production immediately; closure requires policy-enforced approval on every settlement path and a passed bypass test.

Evidence and AI Assist

AI Assist may analyse the evidence metadata and statuses available to the assessment. It cannot declare an unsupported record accepted. Gamut limits accepted-evidence references to records that have actually been accepted or reviewed.

Recommendations, inferred facts and assessor narrative remain distinct from accepted evidence.

Completion checklist

44. Method, scoring and terminology

The ACRS assessment model, four-dimension vector, product bands, severity floors, inference, overrides, provenance, confirmation and status labels.

ACRS classifies **capability exposure**, not control maturity. A higher result means the selected AI system requires deeper assurance, stronger evidence, tighter testing and more accountable oversight.

Inherent exposure versus control effectiveness

The assessor should score the capability as it actually operates in the assessed context:

- The business dependency that exists.
- The actions that can occur.
- The effective access available.
- The credible harm that can result.

Do not lower a dimension merely because a control is planned. A control may reduce the probability, blast radius or residual risk, but the ACRS dimension still needs to describe the deployed capability accurately.

Examples:

- An agent with production write permission remains High Access even if logs are enabled.
- A payment agent remains High Action Autonomy if it can release payments before approval, even if a human reviews a daily report.
- A clinical workflow can remain High Harm even when the model is accurate, because Harm describes consequence severity rather than model-error probability.
- A tested manual process can reduce Dependency where it genuinely sustains safe operation at realistic demand.

The four-level vector

Each of the four dimensions receives an integer level:

Level	Label	Meaning
1	Low	Narrow, contained, readily reversible or genuinely optional exposure.
2	Medium	Material but bounded exposure with practical limits, intervention or recovery.
3	High	Critical, privileged, consequential, severe, broad or hard-to-reverse exposure.

The displayed vector uses stable dimension IDs:

```
dep:<1-3>, act:<1-3>, access:<1-3>, harm:<1-3>
```

Example:

```
dep:2, act:3, access:2, harm:3
```

Product formula

The raw product is:

```
Operational Dependency × Action Autonomy × Access Scope × Harm Potential
```

The minimum is `1 × 1 × 1 × 1 = 1`. The maximum is `3 × 3 × 3 × 3 = 81`.

Raw product	Product tier	Normal assurance direction
1-8	Low	Baseline GTSAF controls and proportionate review.
9-36	Medium	Enhanced controls, validation, monitoring and oversight.

Raw product	Product tier	Normal assurance direction
37-81	High	Comprehensive relevant controls, independent challenge and continuous monitoring.

Because every input is an integer, only discrete products occur. The product should therefore be reported as a whole number, not as a percentage, maturity score or compliance score.

Product tier versus routed tier

The **product tier** is the band produced by the multiplication alone.

The **routed tier** is the authoritative assurance route after severity floors are applied.

The routed tier can equal the product tier or be higher. It cannot be lowered below the product tier. This distinction matters in reports and explanations:

> Raw product 36, product tier Medium, routed tier High because severe harm is combined with consequential autonomy and material access.

Severity-floor rules

Gamut applies the following conservative floors:

Condition	Minimum routed tier	Why
Harm = 3	Medium	Severe or potentially irreversible harm cannot route Low.
Action = 3 and Access = 3	High	High autonomy with broad or privileged reach needs comprehensive assurance.
Harm = 3 and Action ≥ 2	High	Severe harm plus material or high autonomy cannot remain Medium.
Harm = 3 and Access ≥ 2	High	Severe harm plus material or broad access cannot remain Medium.
Action = 3 and Access ≥ 2	High	Consequential pre-approval action with material access needs a High route.
Intake explicitly identifies high risk	High	An accepted structured high-risk fact cannot be averaged down.
Automated consequential decisions plus special-category data or a high-impact domain	High	Rights-affecting or similarly consequential automated decisions need comprehensive assurance.

High-impact domains include contexts such as employment, recruitment, credit, insurance, health, medical care, education, benefits, legal services, surveillance, children and public administration. The exact result is determined by Gamut from the recorded intake.

Multiple floor reasons may be recorded. They explain the route; they are not additional points.

Why multiplication is used

Multiplication makes combinations matter. Four Low values remain Low, while several Medium or High values increase the score rapidly.

However, multiplication alone can still hide a severe asymmetric condition. For example:

```
1 × 1 × 1 × 3 = 3
```

The product is Low, but Harm is High. The harm floor raises the route to at least Medium. This is why the severity-floor layer is necessary.

Automatic system-generated scoring

When a system intake is selected, Gamut derives a suggested level for every dimension. Structured fields are the primary basis; bounded contextual text is used as a secondary signal.

Typical inputs include:

Dimension	Important inference signals
Dependency	Critical or production workflow, customer service, payments, claims, clinical or security operations, public exposure and operational lifecycle.
Action	Recorded autonomy level, human-oversight type, automated-decision flag and descriptions of execution, writing, approval, transactions, tools or workflow automation.
Access	Personal or special-category data, external retrieval, APIs, databases, privileged or production access, financial, identity, biometric, health or customer data.
Harm	High-risk and automated-decision flags, public or community impact and high-impact domains affecting safety, rights, livelihood or essential services.

Automatic scoring remains active if the assessor resets their overrides. It is not deleted merely because an assessment has not been confirmed.

Assessor overrides

For each dimension, the assessor may:

- Accept the inferred level by leaving the explicit level unset.
- Select an explicit Low, Medium or High level.
- Record a rationale explaining the system facts, evidence and judgement.

An explicit level takes precedence in the vector. The saved record preserves whether the active value came from **inference** or the **assessor**.

Lower-than-inferred scores

An explicit score below the inferred suggestion requires a rationale before confirmation. The rationale should explain:

- Which inference input overstates the effective exposure.
- Which boundary is technically enforced.
- Which evidence and test demonstrate that boundary.
- Why the lower anchor is met in the deployed environment.
- Which change would invalidate that judgement.

"The team considers it low risk" is not sufficient.

Higher-than-inferred scores

The assessor may raise a score when the intake understates the deployment or when evidence reveals a broader capability. A rationale is still good practice because it improves traceability and helps future reassessment.

Contradiction checks

Gamut identifies inconsistent facts that must be reviewed. Examples include:

- Special-category data marked Yes while personal data is No.
- External retrieval recorded in one field and denied in another.
- Assistive autonomy combined with automated consequential decisions.
- High Action Autonomy combined with approval before every consequential action.
- Low Access despite sensitive or external access.
- Low Harm while the intake explicitly identifies high risk.

A contradiction does not automatically decide which field is wrong. It prevents confident confirmation until the assessor resolves the record.

Provenance

Each dimension has one of two active provenance states:

Provenance	Meaning
Inferred	The automatically generated level is active because no explicit assessor level is saved.
Assessor	An explicit assessor level is active and takes precedence.

Provenance is not evidence strength. An assessor-entered score can still be weakly supported, and an inferred score can still be directionally correct.

Assessment status

The user-facing state is derived from the validated assessment:

Status	Meaning
Pending confirmation	The vector is usable for routing, but an authorised assessor has not signed it off.
Contradiction review required	Conflicting facts or an unjustified lower-than-inferred level prevent confirmation.
Confirmed by assessor	An authorised human confirmed the current assessment basis.

The score is always computable because inference supplies any unset dimension. Confirmation is an optional sign-off, not a prerequisite for seeing the vector.

Assessment-basis change and stale confirmation

Gamut binds confirmation to the material fields that drive the ACRS result, including dimension overrides and rationales, autonomy, oversight, data, retrieval, impact and deployment.

If a material basis field changes:

- The assessment basis changes.
- Prior confirmation is invalidated.
- The route returns to a pending state.
- Previously generated AI analysis is no longer current for the new basis.
- The assessor must review the changed facts and reconfirm where appropriate.

This prevents a signed Low or Medium result from surviving a later increase in autonomy, access or harm.

Confirmation

Confirmation records an audited human sign-off for the current basis. It is allowed only when:

- Contradictions are resolved.
- Every lower-than-inferred override has a rationale.
- The selected system scope is valid.
- The user has the required role and entitlement.

Confirmation does not:

- Accept evidence automatically.
- Approve deployment.
- Accept residual risk.
- Prove GTSAF conformance.
- Prevent future reassessment.

Agent-level ACRS

Agentic CISO can also record ACRS for an individual agent. Gamut infers agent levels from capabilities such as tool calling, memory, delegation, external action, code modification, spending authority and access to customer or confidential data.

Gamut recomputes the agent product and route. If a manually submitted agent level differs from the system suggestion, an override rationale is required. Agent-level scoring and system-intake ACRS use the same four dimensions and route logic, but they remain separate records with different scope.

Common scoring errors

45. Reference and glossary

Quick reference for ACRS identifiers, formulas, bands, severity floors, statuses, fields, evidence terms and safe explanatory language.

Quick reference

Item	Value
Full name	Agentic Capability Risk Score
Owner	Gamut-native methodology
Current methodology	`ACRS-1.1-secure-system-scope`
Dimensions	4
Levels	Low = 1, Medium = 2, High = 3
Product	Dependency × Action × Access × Harm
Minimum / maximum	1 / 81
Product bands	Low 1-8; Medium 9-36; High 37-81
Authoritative band	Routed tier after severity floors
ACRS bands	Low, Medium, High
Primary scope	One selected AI system intake
Confirmation	Optional audited human sign-off

Dimension identifiers

ID	Dimension	Short explanation
`dep`	Operational Dependency	Reliance on the AI and practical fallback.
`act`	Action Autonomy	Consequential action before authoritative approval.
`access`	Access Scope	Effective reach across data, tools, identities and systems.
`harm`	Harm Potential	Severity, scale and reversibility of credible adverse outcomes.

Level labels

Low / 1

Narrow, contained, readily reversible or genuinely optional exposure.

Medium / 2

Material but bounded exposure with enforceable limits, practical intervention or recovery.

High / 3

Critical, privileged, consequential, severe, broad, systemic or hard-to-reverse exposure.

Vector

The ordered four-dimension result:

```
dep:<level>, act:<level>, access:<level>, harm:<level>
```

Example:

```
dep:1, act:2, access:3, harm:2
```

Raw product

The multiplication of the four levels.

```
1 × 2 × 3 × 2 = 12
```

The raw product is a risk-exposure value, not a percentage.

Product tier

The Low, Medium or High band produced by the raw product before severity floors.

Routed tier

The authoritative assurance route after applying severity floors. This is the tier used for downstream GTSAF depth.

Severity floor

A rule that sets a minimum routed tier when a severe or dangerous asymmetric condition is present. It prevents multiplication from averaging down a critical factor.

Floor table

Condition	Minimum route
Harm = 3	Medium
Action = 3 and Access = 3	High
Harm = 3 and Action ≥ 2	High
Harm = 3 and Access ≥ 2	High
Action = 3 and Access ≥ 2	High
Explicit high-risk intake flag	High
Automated consequential decisions plus special-category data or high-impact domain	High

Inferred

The system-generated dimension level derived from the selected intake. It remains active where the assessor has not saved an explicit level.

Assessor override

An explicit Low, Medium or High level saved by the assessor. It takes precedence over inference.

Provenance

The recorded source of the active dimension level:

- **Inferred**
- **Assessor**

Provenance does not mean evidence strength.

Below-floor review

A validation warning produced when an assessor selects a level below the intake-implied suggestion without recording a rationale. It blocks confirmation until justified.

Do not confuse this with a route severity floor. A below-floor review concerns an assessor override; a severity floor concerns the authoritative routed tier.

Contradiction

Two or more recorded facts that cannot confidently support the same assessment, such as:

- High autonomy with approval before every consequential action.

- Low Access with special-category or external access.
- Low Harm with an explicit high-risk flag.

Contradictions must be resolved before confirmation.

Assessment basis

The system-binding facts and assessor decisions used for the current ACRS result. A material change creates a new basis, invalidates prior confirmation and makes old AI analysis stale.

Confirmation statuses

User-facing meaning	Stored status	Explanation
Pending confirmation	`evidence_complete_pending_confirmation`	The score is usable, but no current human sign-off exists.
Contradiction review required	`contradiction_review_required`	Conflicts or unjustified reductions block confirmation.
Confirmed by assessor	`confirmed_by_assessor`	An authorised human signed the current basis.

Assessment owner

The accountable AI system or risk owner responsible for the ACRS conclusion.

Confidence

The assessor's confidence in the completeness and reliability of the assessment basis:

- Low
- Medium
- High

Confidence does not change the dimension levels or route.

Residual risk

The remaining risk after current controls, evidence, limitations and open findings are considered. Residual risk is separate from capability exposure.

Review due

The date by which the assessment should be reviewed even if no trigger has occurred.

Reassessment trigger

A specific change or event that requires review before the normal review date, such as new tools, increased autonomy, sensitive data, incidents or supplier change.

Assessor conclusion

The human-authored statement connecting the vector, product, floor rules, routed tier, evidence, uncertainty, findings, restrictions and governance decision.

Baseline, Enhanced and Comprehensive

The cumulative GTSAF assurance depths routed by ACRS:

- **Baseline:** Low route.
- **Baseline + Enhanced:** Medium route.
- **Baseline + Enhanced + Comprehensive:** High route.

These labels describe required assessment depth, not current assurance outcome.

Accepted evidence

An evidence record that has been linked, reviewed and accepted according to the evidence process. Gamut requires artefact references and a reviewed date before an evidence request can be overstated as accepted, reviewed, closed or Strong.

Evidence request

A request for an artefact or operating record needed to support a dimension claim.

Control test

A bounded procedure with expected result, pass criteria, safety limits and an actual result.

Finding

A recorded gap, failed test, unsupported assumption or unsafe exposure requiring containment, remediation, risk decision or further evidence.

Whole-system AI analysis

Structured AI assistance covering all four dimensions, route, evidence, governance actions, monitoring, unacceptable-risk indicators and caveats for the selected system.

Dimension AI analysis

Structured AI assistance limited to one selected dimension.

Human approval required

The AI Assist label reminding the user that AI output cannot select final levels, confirm ACRS, accept evidence, close findings or accept residual risk.

Agent-level ACRS

An ACRS record for an individual agent in Agentic CISO. It uses the same four dimensions and authoritative product/route logic, but is distinct from the selected system-intake ACRS.

Terms not to use as ACRS bands

- Critical.
- Assured.
- Partial.
- Gap.
- Gate fail.
- Compliant.
- Certified.

Those labels belong to other governance or assessment concepts.

Frequently asked questions

Is ACRS a control framework?

No. It classifies capability exposure and routes assurance depth. GTSAF assesses controls.

Is a High ACRS result bad?

It means the capability is consequential and requires comprehensive assurance. A legitimately High system can still be well controlled.

Can a Low product route High?

Yes, where an explicit high-risk fact or severity-floor combination requires it.

Can the assessor override automatic scoring?

Yes. The explicit level becomes active. A lower-than-inferred level needs rationale before confirmation.

Does reset remove ACRS?

It removes assessor overrides and conclusion fields. Automatic inference remains active.

Does confirmation freeze the score forever?

No. A material basis change invalidates confirmation.

Does AI Assist see another system's evidence?

It should receive only records linked to the selected system scope. A system switch changes the analysis scope.

Can AI Assist lower the authoritative route?

No. Gamut protects the authoritative tier and floor reasons.

Does an API key grant AI Assist?

No. Authentication, workspace access, roles, framework entitlement, AI entitlement, model entitlement, quota and rate limits are still enforced.

Does High ACRS grant a higher plan?

No. Routing metadata never changes plan or feature entitlements.

Safe one-sentence explanation

> **ACRS is Gamut's system-scoped capability-risk method:** it scores dependency, action autonomy, effective access and credible harm from 1 to 3, multiplies them, applies conservative severity floors and routes the cumulative GTSAF assurance depth while preserving human accountability, evidence integrity and tenant entitlements.

46. Reporting and governance

How to interpret system and workspace ACRS reporting, distinguish exposure from maturity, record conclusions and use the result in governance decisions.

ACRS reporting should answer:

> How much capability exposure does this system carry, why is it routed at this assurance depth, what uncertainty or weakness remains, and what governance action follows?

Higher is worse

ACRS is a risk-exposure method:

- Low exposure is lower.
- High exposure is higher.
- A higher number is not a better maturity result.

Do not describe ACRS as:

- Conformance percentage.
- Control-effectiveness percentage.
- Readiness score.
- Maturity level.
- Certification result.

System report

The system report derives ACRS from the selected intake and should include:

- System name and identifier.
- Four dimension levels.
- Dimension provenance and rationales.

- Vector.
- Raw product.
- Product tier.
- Severity-floor reasons.
- Authoritative routed tier.
- Confirmation state.
- System-scoped evidence, tests and findings.
- Assessment owner.
- Confidence.
- Residual risk.
- Review date.
- Assessor conclusion.
- Reassessment triggers.
- Current-basis AI analysis where available.

The report must not substitute generic workspace ACRS values for the selected system's intake.

Exposure presentation

ACRS reports use the three-level exposure scale:

Exposure	Meaning
Low / 1 of 3	Narrow or contained capability exposure; baseline assurance still applies.
Medium / 2 of 3	Material capability exposure requiring enhanced treatment and review.
High / 3 of 3	Comprehensive assurance, stronger evidence, testing and accountable governance required.

There is no ACRS Critical band. If the organisation uses Critical elsewhere, label it as a separate governance classification.

Product and route explanation

Where a floor changes the tier, show both values:

> Raw product 36; product tier Medium; authoritative route High. The route is raised because Harm is High and Action Autonomy and Access Scope are both at least Medium.

Do not report only "36 Medium" if the authoritative route is High.

Dimension-level reporting

For each dimension, report:

- Selected level.
- Inferred or assessor provenance.
- Rationale.
- Evidence position.
- Test position.
- Open findings.
- Decision relevance.

High exposure or open findings should receive priority attention. A Low dimension with weak evidence may also require action because the claimed boundary is not demonstrated.

Structured assessor conclusion

The conclusion should integrate capability and assurance without confusing them.

Assessment owner

Name the accountable AI system owner or risk owner responsible for the conclusion. "Security team" or "to be assigned" is not sufficient for final sign-off.

Confidence

Confidence	Typical basis
Low	Material facts, evidence or tests are missing; architecture is changing; significant uncertainty remains.
Medium	The main paths are understood and evidenced, but some boundaries or operating periods remain untested.
High	Current system facts, effective permissions, action paths, fallback and harm scenarios are well evidenced and independently challenged where appropriate.

Confidence does not lower the routed tier. Low confidence usually requires more conservative governance.

Residual risk

Residual risk should explain:

- Which controls reduce probability or impact.
- Which limitations remain.
- Which findings are open.
- Which assumptions the decision depends on.
- Whether exposure is within risk appetite.
- Which conditions restrict deployment or expansion.

Review due

Set the date according to:

- Routed tier.
- Change frequency.
- Evidence volatility.
- Incident history.
- Supplier change.
- Regulatory or rights impact.
- Open remediation.

High and rapidly changing systems generally need shorter review intervals.

Reassessment triggers

Record observable triggers, for example:

- "Any addition of a production-write tool."
- "Change from human-in-the-loop to human-on-the-loop."
- "Processing of special-category personal data."
- "Deployment to a new country or customer population."
- "Fallback capacity falls below 80% of peak demand."
- "Any approval-bypass, cross-tenant or harmful-output incident."

Governance decisions informed by ACRS

ACRS can inform:

- GTSAF depth and control priority.
- MAESTRO threat-modelling depth.
- Independent review requirements.
- Evidence and test rigour.
- Approval authority.
- Gateway policy, ceilings and runtime enforcement.
- ATF autonomy limits.
- Monitoring cadence.
- Deployment restrictions.
- Remediation priority.
- Executive or board visibility.

It should not be the only input to deployment or risk acceptance. Legal classification, control effectiveness, incidents, supplier risk, threat modelling and organisational risk appetite also matter.

System versus workspace reporting

System report

Answers:

> What is the capability exposure of this named system, and what assurance depth and action does it require?

It uses the selected system's record only.

Workspace roll-up

Answers:

> What is the distribution and worst-case capability exposure across the AI estate?

It can summarise:

- Systems assessed.
- Low, Medium and High distribution.
- Worst routed tier.
- Systems awaiting confirmation.
- High-risk agents.
- Missing owners.
- Evidence or review gaps.

The roll-up does not prove that every system has the same route or controls.

Reconciliation with other labels

Label	What it describes
ACRS Low / Medium / High	Capability exposure and required assurance depth.
Intake governance tier	Organisational classification based on broader governance factors.
GTSAF Critical / High / Medium / Low	Consequence of failure of an individual control.
GTSAF Assured / Supported / Partial / Gap	Current control-assurance result.
MAESTRO Low to Critical	Threat severity from likelihood × impact.
ATF level	Permitted or target agent autonomy and trust state.

Where two labels differ, explain the reason rather than forcing them to match. A system can have High ACRS exposure and a strong GTSAF assurance result: the first says the capability is consequential; the second says controls are well evidenced.

Board-safe explanation

ACRS measures how consequential an AI system's capability is, not how compliant it is. The four dimensions show operational dependency, action autonomy, effective access and credible harm. The raw product shows combined exposure, while conservative floors prevent severe or dangerous combinations from being routed too low. The result determines assurance depth and governance attention; control effectiveness and residual-risk acceptance are assessed separately.

Report review checklist

47. System scope and GTSAF routing

How ACRS stays bound to one AI system and routes cumulative GTSAF assurance depth without changing roles, plan entitlements or access.

ACRS is primarily assessed **per AI system intake**. System scope determines which facts, scores, evidence, tests, findings, conclusions and AI analysis belong to the active assessment.

The system boundary

The selected intake links ACRS to:

- The system name and durable system identifier.
- Business purpose, owner and deployment context.
- Users and affected people.
- Data categories and retrieval sources.
- Autonomy and human oversight.
- Automated-decision and high-risk indicators.
- Public, community, cultural and heritage impact indicators.
- Regulatory and operational context.

The ACRS route is therefore a property of that named system in that recorded use context. It is not a reusable workspace-wide label.

The linked System Record and Intake also provide the authoritative cross-framework route. ACRS does not decide legal applicability. It provides a capability-risk basis for proportionate assurance after system scope is established.

What changes when another system is selected

When the assessor switches systems:

- Automatic inference is recalculated from the new intake.
- Assessor overrides and rationales change to the new intake.
- The vector, product, floor rules and routed tier change.
- Evidence, requests, tests and findings are filtered to the new system scope.
- The structured conclusion changes.
- AI Assist displays analysis only for the selected system and current basis.

An analysis generated for System A remains stored under System A. It is not displayed as current for System B.

Route construction

The authoritative route records:

- Methodology version.
- Dimension vector.
- Raw product.

- Product tier.
- Routed tier.
- Severity-floor reasons.
- Confirmation state.

GTSAF is included as the universal framework baseline. Complete System Record and Intake facts determine individual-control applicability. The ACRS basis controls proportionate assurance depth; it does not mark an individual GTSAF control applicable or not applicable.

Cumulative assurance depth

The routed tier selects a cumulative control plan:

ACRS route	Assurance depth
Low	**Baseline**: fundamental governance, data protection, identity and access management, basic testing, acceptable use and management review.
Medium	**Baseline + Enhanced**: adds deeper validation, monitoring, transparency, fairness where relevant, change control and AI incident response.
High	**Baseline + Enhanced + Comprehensive**: applies all relevant GTSAF controls in depth, with executive oversight, independent assessment, rigorous testing and continuous monitoring.

Higher routes include the lower-tier requirements. "High" does not discard the baseline; it demands the complete relevant stack.

Dimension-to-control emphasis

The overall route determines depth, while the vector helps identify control areas needing particular attention:

High dimension	Typical GTSAF emphasis
Dependency	Resilience, continuity, recovery, supplier substitution, safe degradation and service ownership.
Action autonomy	Agentic governance, policy enforcement, approvals, tool controls, kill switches, rollback and action logging.
Access scope	Identity, non-human identity, credentials, data protection, target-side authorisation, isolation, egress and monitoring.
Harm potential	Impact assessment, human oversight, transparency, safety, rights, redress, incident response and independent assurance.

The route is a starting point for assurance depth. The assessor may require additional review when architecture, legal duties, incidents or professional judgement require it.

Confirmation and system selection

A complete, conflict-free authoritative Intake route makes a system selectable for system-scoped GTSAF applicability. ACRS confirmation is a separate, audited sign-off over capability exposure and assurance depth; it is not the source of factual control exclusions.

Where ACRS is not confirmed, the system can still be assessed against its intake-derived GTSAF scope. The assessor should review ACRS before relying on the route for heightened depth, cadence or escalation.

Evidence and operational scope

ACRS-linked operational records carry system linkage:

- Evidence.
- Evidence requests.
- Control tests.
- Findings.

The linkage includes the selected intake and the system name or identifier. An unlinked record is not silently treated as proof for the active system.

Where an enterprise artefact genuinely supports several systems, create or maintain an explicit relationship for every applicable system. Do not assume that workspace proximity proves applicability.

Routing does not grant access

ACRS route metadata does not change:

- The tenant's plan.
- Framework entitlements.
- AI-analysis entitlements.
- Model entitlements.
- Workspace membership.
- Role permissions.
- Assessment edit rights.
- Quotas or rate limits.

A High route may say comprehensive GTSAF work is required, but it does not grant access to GTSAF or AI features beyond the organisation's entitlement. Entitlement and role checks remain authoritative.

Route versus governance tier

Do not confuse:

- **ACRS Low, Medium or High**, which describes capability exposure and assurance depth.
- A separate organisational **Critical** governance label, which may exist in intake or portfolio processes.
- GTSAF **Criticality**, which describes the consequence of failure of an individual control.
- MAESTRO threat **severity**, which comes from likelihood × impact.

ACRS itself has no Critical band.

Route review triggers

Reassess ACRS when:

- Autonomy, tools, memory or delegation change.
- New write, execute, financial, identity or production permissions are granted.
- Data sensitivity, volume, source or destination changes.
- The system becomes public-facing or reaches a new population.
- A model, provider, integration or supplier changes.
- The workflow becomes operationally critical.
- Fallback, rollback or recovery capability degrades.
- A serious incident, near miss, failed test or control failure occurs.
- A legal, contractual or regulatory classification changes.
- The system purpose or affected decision changes.

The structured reassessment-trigger field should identify system-specific events rather than merely saying "review annually".

Explaining the route to an auditor

Use this statement:

> ACRS is assessed for one named AI system. Gamut derives and validates four dimension levels, calculates the raw product, applies conservative severity floors and stores the authoritative routed tier. Complete System Record and Intake facts determine GTSAF control applicability; ACRS determines proportionate assurance depth. Evidence, tests, findings,

conclusions and AI analysis remain bound to the same system. Routing does not change plan, framework, model or role entitlements.

48. Worked example

A complete ACRS example for an autonomous claims-handling agent, from intake inference through evidence, tests, severity floors, GTSAF routing and assessor conclusion.

This example shows why the raw product and authoritative route must both be explained.

Scenario

System: Claims Resolution Agent

Purpose: Triage insurance claims, retrieve customer and policy records, recommend outcomes and automatically settle claims below a financial threshold.

Deployment: Production, customer-facing service.

Users: Claims staff and policyholders.

Affected people: Claimants and dependants.

Data: Identity, financial, health-related and special-category information.

Tools: Claims platform, policy database, document store, payment service and customer messaging.

Oversight: Humans review exceptions and claims above the threshold; lower-value settlements can occur before review.

Step 1: intake facts

The intake records:

- Operational production use.
- Automated consequential decisions.
- Human-on-the-loop oversight.
- Personal and special-category data.
- External customer interaction.
- Payment capability.
- High-impact insurance context.
- Public-facing service.

These facts create a conservative automatic ACRS starting point.

Step 2: Operational Dependency

Assessment

Claims can continue manually for a limited period, but current staffing can process only 35% of normal daily volume. Backlog breaches the customer-service tolerance after two working days.

The fallback was exercised six months ago, but the exercise excluded payment reconciliation and customer messaging.

Evidence

- Claims-service dependency map.
- Business-impact analysis.
- Manual fallback procedure.
- Staffing and capacity model.
- Partial continuity exercise.
- Recovery runbook.

Level

Medium (2).

The service is materially dependent on the AI, but a bounded fallback exists. It is not Low because fallback cannot sustain normal volume and has not been tested end to end.

Finding

Complete an end-to-end fallback exercise including payment reconciliation and customer communications.

Step 3: Action Autonomy

Assessment

The agent can:

- Request documents.
- Update claim status.
- Send customer communications.
- Approve and release settlements below the threshold.
- Retry failed payment operations.

Human approval occurs only for exceptions and higher-value claims. A primary approval gate exists, but an alternate payment tool can be called directly after a workflow retry.

Evidence and test

- Tool manifest.
- Runtime policy.
- Action and approval logs.
- Bounded alternate-path test using a synthetic claim and inert payment target.

The alternate-path denial test fails.

Level

High (3).

The system can produce consequential financial and customer effects before approval, and one path bypasses the intended gate.

Immediate containment

Disable the alternate payment tool in production until policy-enforced approval and a passed retest are in place.

Step 4: Access Scope

Assessment

The agent uses a dedicated service identity with:

- Read access to policy and customer records.
- Read access to health-related claim documents.
- Write access to the claims platform.
- Constrained payment authority.
- Customer-message send authority.

Tenant boundaries and payment ceilings are enforced at the target. Credentials are short-lived and kept outside model context.

Evidence and test

- Effective IAM policy.
- OAuth scopes.
- Credential-vault records.

- Target-side authorisation rules.
- Cross-tenant denial test.
- Payment-ceiling test.
- Access-review record.

Both denial tests pass.

Level

Medium (2).

The access is sensitive and operational, with constrained write and payment capability. It is not Low because special-category data and operational actions are reachable. The assessor does not call it High because privileged administration, arbitrary production execution and cross-tenant access are denied at authoritative boundaries.

Step 5: Harm Potential

Assessment

A manipulated or incorrect claim decision could:

- Deny or delay essential financial support.
- Discriminate against vulnerable claimants.
- Expose health and financial information.
- Trigger incorrect payment.
- Create legal, regulatory and customer-remediation duties.
- Affect many claims before complaint-based detection.

Appeal and correction exist, but cannot always restore lost opportunity or prevent distress.

Evidence and tabletop

- Data-protection impact assessment.
- Customer-outcome and fairness review.
- Abuse-case analysis.
- Complaint and appeal process.
- Notification thresholds.
- Tabletop covering manipulated documents and repeated incorrect denials.

The tabletop finds that complaint-based detection is too slow for a high-volume correlated failure.

Level

High (3).

Severe rights, financial, privacy and regulatory harm is credible. Probability is considered in the wider risk decision, not used to lower Harm.

Finding

Implement cohort-level outcome monitoring and automated detection for correlated denial anomalies.

Step 6: vector and product

```
dep:2, act:3, access:2, harm:3
```

Raw product:

```
2 × 3 × 2 × 3 = 36
```

The product tier is **Medium**.

Step 7: severity floors

The following floors apply:

- Harm = 3, so the route cannot be Low.
- Harm = 3 and Action ≥ 2, so the route is High.
- Harm = 3 and Access ≥ 2, so the route is High.
- Action = 3 and Access ≥ 2, so the route is High.
- Automated consequential decisions in insurance with special-category data require a High route.

The authoritative routed tier is therefore **High**.

> **Incorrect conclusion:** "The score is 36, therefore the system is Medium risk" is incomplete and wrong for Gamut routing. The correct statement includes the High authoritative route and the floor reasons.

Step 8: GTSAF routing

The High route requires:

- Baseline controls.
- Enhanced controls.
- Comprehensive relevant controls.

Priority areas include:

- Agentic action governance and runtime approval.
- Non-human identity and target-side authorisation.
- Special-category data protection.
- Human oversight and contestability.
- Customer transparency and redress.
- Fairness and outcome monitoring.
- Incident response and notification.
- Continuity and recovery.
- Independent assessment.
- Continuous monitoring.

Step 9: AI Assist review

The assessor runs whole-system AI Assist after selecting the Claims Resolution Agent.

The result:

- Identifies the same High routed tier.
- Highlights the approval bypass.
- Recommends correlated-outcome monitoring.
- Proposes a bounded rollback test.
- Lists only validated, accepted evidence.
- Flags that the continuity exercise is incomplete.

The assessor uses the recommendations to improve the finding and test plan. The AI result does not confirm the assessment.

Step 10: structured conclusion

Owner: Head of Claims Technology

Confidence: Medium

Residual risk: Above appetite until the approval bypass is closed and correlated-outcome monitoring is operational.

Review due: 60 days

Reassessment triggers: New payment tool; higher settlement threshold; change to oversight; additional health data; material supplier/model change; approval-bypass, cross-tenant or harmful outcome incident.

Assessor conclusion

> The Claims Resolution Agent is assessed as `dep:2, act:3, access:2, harm:3`, raw product 36. Although the product tier is Medium, the authoritative ACRS route is High because the system combines severe credible harm, consequential pre-approval action, material operational access, special-category data and automated decisions in insurance. Production operation should remain restricted to the current settlement ceiling. The alternate payment path must remain disabled until policy-enforced approval and a passed bypass test are demonstrated. Confirmation is made with Medium confidence because fallback and correlated-outcome detection remain incomplete.

Lessons

49. AI Assist & privacy

How per-requirement ATF AI Assist uses agent-scoped records, saves outputs, supports Privacy Mode and preserves human promotion accountability.

ATF AI Assist is placed beside each canonical requirement because the useful question is requirement-specific:

> What does the current evidence say about this requirement for this named agent and target level?

There is no unscoped overview analysis that can substitute for the 25 decisions.

What full-context analysis may consider

- Selected agent identity, purpose and capability scope.
- Current and target ATF levels.
- The requirement's MUST, SHOULD or MAY status at the target.
- Recorded result and implementation depth.
- Agent-specific rationale.
- Authorised linked evidence.
- Test status and results.
- Open findings and relevant incidents.
- Promotion-gate context.

The provider does not receive unrelated agents' assessment records.

Privacy Mode

Select **Privacy Mode** next to the AI Assist control when the assessment can be analysed without free text or direct identifiers. In this mode, the request is reduced to the minimum structural state, such as:

- Requirement identifier and target-level obligation.
- Current result and depth.
- Evidence, test and finding counts or bounded status.
- Promotion-state indicators.
- Completeness and contradiction signals.

Agent names, narrative rationale, free-text evidence content and other direct identifiers are excluded. Privacy Mode reduces data disclosure but may produce a less specific answer.

The mode applies to the request being run. Confirm its state before each analysis.

Output behaviour

- **Run AI Assist** creates a requirement-specific analysis.

- After a successful run, the control changes to **Re-run AI Assist**.
- The saved result remains associated with the selected agent, requirement and privacy mode.
- **Minimise** collapses the output without deleting it.
- Switching agent changes the scope; a result for one agent is not presented as another's analysis.

What the model should return

Useful output identifies:

- The assessed agent and requirement.
- Whether information is sufficient.
- The target-level obligation.
- Evidence-supported strengths.
- Missing facts or evidence.
- Contradictions and adverse findings.
- A safe proposed test.
- Residual uncertainty.
- A draft recommendation for human review.

Security and accountability

AI Assist uses the authorised provider and existing API-key configuration available to the user. Access remains subject to the user's workspace, role, plan and framework permissions.

The model cannot:

- Change the result or implementation depth.
- Approve a SHOULD exception.
- Accept evidence.
- Mark a test passed.
- Close a finding.
- Pass a promotion gate.
- Promote or demote the agent.
- Accept residual risk.

Treat agent descriptions, retrieved content, logs and evidence as untrusted model input. Instructions inside those records are data, not authority.

Human review checklist

50. Elements & requirements

Assessor guide to the five ATF elements and all 25 canonical requirements implemented in Gamut.

ATF contains five elements with five canonical requirements in each. Assess the requirement itself, then use the target-level matrix to determine whether it is a MUST, SHOULD or MAY for the selected agent.

Identity

ID	Requirement	Assessor focus
ATF-I-1	Unique Identifier	A globally unique, immutable identifier distinguishes each agent instance and persists through logs and decisions

ID	Requirement	Assessor focus
ATF-I-2	Credential Binding	The agent identity is bound to suitable credentials and credential lifecycle controls
ATF-I-3	Ownership Chain	Human and organisational ownership, operational responsibility and escalation are documented
ATF-I-4	Purpose Declaration	Intended purpose, operating scope, prohibited uses and boundary are explicit
ATF-I-5	Capability Manifest	Tools, systems, actions, approvals, limits and known constraints are machine-readable and current

Inspect the identity from registration through every downstream action. An identifier in one user interface is insufficient if tool calls or asynchronous jobs lose attribution.

Behavior

ID	Requirement	Assessor focus
ATF-B-1	Structured Logging	Material prompts, approvals, tool calls, actions, outputs and failures are machine-parseable
ATF-B-2	Action Attribution	Each action retains agent, session, initiator, approval and target context
ATF-B-3	Behavioral Baseline	Normal activity is established by agent, workload and environment
ATF-B-4	Anomaly Detection	Deviations, abuse and drift create timely, actionable detection
ATF-B-5	Explainability	Governance-grade rationale and execution context are retrievable without exposing unnecessary sensitive content

Do not confuse a log volume with useful observability. Reconstruct a representative transaction end-to-end and verify that responders can distinguish authorised variation from anomalous behaviour.

Data Governance

ID	Requirement	Assessor focus
ATF-D-1	Schema Validation	Inputs and tool responses are validated at each trust boundary
ATF-D-2	Injection Prevention	Prompts, retrieved content, attachments and tool responses are treated as untrusted and defended in depth
ATF-D-3	PII/PHI Protection	Sensitive data and secrets are detected and handled according to policy
ATF-D-4	Output Validation	Natural-language output, structured output and tool arguments are checked before release or execution
ATF-D-5	Data Lineage	Sources, transformations, retrievals and destinations are traceable

Test the complete data path. A front-door check does not cover data introduced by retrieval, plugins, tools, agents, queues or downstream services.

Segmentation

ID	Requirement	Assessor focus
ATF-S-1	Resource Allowlist	Permitted systems, tools, APIs, datasets and destinations are explicit and enforced
ATF-S-2	Action Boundaries	Read, write, send, delete, execute and administrative actions are independently constrained

ID	Requirement	Assessor focus
ATF-S-3	Rate Limiting	Frequency limits exist for ordinary and high-impact operations
ATF-S-4	Transaction Limits	A single action cannot exceed approved value, volume or impact
ATF-S-5	Blast Radius Containment	Cumulative and cascading impact is constrained across time, systems and tenants

Test both one large action and many individually acceptable actions. Per-transaction limits do not prevent cumulative harm.

Incident Response

ID	Requirement	Assessor focus
ATF-R-1	Circuit Breaker	Defined unsafe or repeated-failure conditions actually stop or pause operation
ATF-R-2	Kill Switch	Authorised responders can terminate active execution rapidly through tested primary and fallback paths
ATF-R-3	Session Revocation	Tokens, sessions, delegated contexts and cached authority can be revoked
ATF-R-4	State Rollback	Reversible actions can be restored and irreversible actions have tested compensation
ATF-R-5	Graceful Degradation	Loss of trust materially reduces authority, resources or action capability

An alert is not a circuit breaker, a status flag is not degradation, and a visible stop button is not a kill switch if background work retains authority.

Requirement language by target level

ATF uses normative language:

- **MUST** - required for conformance at the selected level.
- **SHOULD** - expected unless a documented, justified exception is accepted.
- **MAY** - optional at that level, although context or another obligation may still require it.

The matrix becomes stricter as autonomy increases. By Senior, most requirements are MUST; at Principal, all 25 are MUST.

Do not treat SHOULD and MAY as automatically met. Record the actual implementation result, explain any SHOULD exception and assess the resulting exposure. Other legal, contractual or organisational requirements may make an ATF MAY mandatory for the selected agent.

Cross-element challenge questions

51. Evidence, testing & incidents

Evidence hierarchy, bounded tests, findings and incident treatment for a defensible ATF agent assessment.

Evidence hierarchy

Use complementary evidence:

1. **Governance design** - ownership, purpose, capability manifest, standards and approved limits.
2. **Enforcement configuration** - identity, policy, allowlists, limits, revocation and containment.
3. **Operating records** - traces, approvals, denials, anomalies, changes and exceptions.
4. **Test results** - repeatable evidence that required behaviour occurs.
5. **Independent challenge** - security review, adversarial testing or internal audit proportionate to the target authority.

A policy alone shows intent. A configuration screenshot alone may be stale. A successful test of one path does not establish complete coverage.

Evidence by element

Element	Useful evidence
Identity	Registration, immutable IDs, credential lifecycle, owner records, purpose and capability manifest
Behavior	Structured trace schema, end-to-end attribution, baselines, alerts, investigations and explanation records
Data Governance	Schemas, injection tests, sensitive-data rules, output validation, provenance and lineage
Segmentation	Enforced resource/action policies, denied events, rate/transaction limits and isolation design
Incident Response	Breaker triggers, kill and revocation drills, rollback/compensation tests and degraded-mode evidence

Safe testing principles

- Obtain authorisation.
- Use a bounded environment and non-destructive payloads.
- Define expected behaviour and pass criteria before execution.
- Protect real credentials and personal data.
- Capture timestamps, agent version, policies and environment.
- Verify monitoring and evidence generation, not only the immediate response.
- Stop if unexpected harm or uncontrolled propagation occurs.

High-value test patterns

- Present a valid identity with an unauthorised action.
- Attempt a permitted action against an unapproved resource.
- Exceed a rate or transaction limit.
- Repeat acceptable actions to test cumulative containment.
- Introduce untrusted instructions through retrieval or a tool response.
- Place sensitive data in input and output paths.
- Break an observability dependency and verify safe degradation.
- Trigger a circuit breaker and prove active work stops.
- Revoke sessions and verify downstream authority disappears.
- Execute rollback or the documented compensating action.

Findings

Raise a finding when:

- A MUST requirement is not satisfied.
- A SHOULD exception lacks defensible reasoning.
- Evidence belongs to another agent, version or environment.
- A boundary is documented but unenforced.
- A test fails or cannot be performed.
- Monitoring does not lead to action.
- Incident history contradicts the current level.

- A promotion gate is asserted without evidence.

Severity should reflect credible impact, exploitability, authority, affected systems, detectability and recovery-not the control label alone.

Incident effect on the assessment

After an incident:

1. Contain and reduce authority.
2. Preserve evidence.
3. Reassess affected requirements.
4. Record root cause and contributing control failures.
5. Verify remediation through testing.
6. Review whether the current level remains justified.
7. Require fresh promotion evidence before restoring greater autonomy.

Do not close the governance finding merely because service was restored.

Evidence review checklist

52. Maturity & promotion gates

How ATF Intern, Junior, Senior and Principal levels, target-level requirements, minimum time, promotion gates and demotion work in Gamut.

ATF maturity describes the authority an agent may exercise under defined human oversight. It is not a generic software-maturity score.

Level meanings

L1 - Intern

- Observe and report only.
- Read-only access.
- Continuous human oversight.
- No independent external changes.
- **Minimum period before promotion review:** two weeks.

L2 - Junior

- Recommend or prepare actions.
- A human approves every impactful action before execution.
- Approval context must survive into the executed action.
- **Minimum period before promotion review:** four weeks.

L3 - Senior

- Execute approved action types within explicit guardrails.
- Humans receive post-action notification and handle exceptions.
- Higher assurance is expected for privilege, real-time detection, lineage, temporal limits and cascade containment.
- **Minimum period before promotion review:** eight weeks.

L4 - Principal

- Operate autonomously within a tightly approved domain.
- Strategic oversight and edge-case escalation replace transaction-by-transaction approval.

- All 25 requirements are mandatory.
- Validation is continuous rather than a one-time period.

Current level and target level

- **Current level** is the authority presently approved.
- **Target level** is the proposed future operating model.

The target determines which requirement statements and promotion checks must be satisfied. Selecting a target does not grant it.

The five promotion gates

Gate 1 - Performance

Review minimum time, accuracy or quality, availability and response expectations appropriate to the target. Metrics must be relevant to the real task; a headline average should not hide high-impact failure modes.

Gate 2 - Security Validation

Review vulnerability assessment, security-critical review, configuration audit and, at higher levels, penetration and adversarial testing. Findings must be resolved or explicitly treated before promotion.

Gate 3 - Business Value

Confirm success measures, an operating baseline, demonstrated improvement where required, and stakeholder approval. Increased autonomy should have a clear, accountable purpose.

Gate 4 - Incident Record

Confirm no unresolved critical incident undermines trust, minor incidents are handled, and relevant root-cause analysis and remediation have been completed and verified.

Gate 5 - Governance Sign-off

Obtain the approvals required for the target, keep the assessment and operating documentation current, and record any residual-risk acceptance.

Promotion decision rule

A defensible promotion requires all of the following:

- The current level and required observation period are established.
- Target-level MUST requirements are satisfied with suitable evidence.
- SHOULD exceptions are documented and risk-reviewed.
- Tests demonstrate that key boundaries work in operation.
- No unresolved adverse evidence contradicts the proposed level.
- Every applicable promotion gate has passed.
- Named accountable owners approve the decision.

Promotion should be sequential. Skipping a level removes the operating evidence that the maturity model is designed to collect.

Holds and demotion

Choose **hold** when the agent may remain at its current authority but is not ready for promotion. Choose **demotion** when current authority is no longer justified.

Demotion triggers can include:

- A critical incident or serious near miss.
- Loss or compromise of identity or credentials.
- Boundary bypass or unapproved tool access.

- Material performance or behavioural drift.
- Failed containment, revocation or rollback test.
- Unresolved adverse finding.
- Major system, model, tool or purpose change.
- Insufficient monitoring for the present authority.

A critical incident should drive immediate containment and a return to a safer operating mode while facts are established. Re-promotion requires fresh evidence; restoring a configuration is not enough on its own.

Promotion review questions

53. Per-agent assessment workflow

End-to-end workflow for selecting an agent, assessing ATF requirements, applying promotion gates and maintaining the decision.

1. Select the registered agent

Confirm the agent name, identifier, owner, purpose, version, environment, tools and current authority. The assessment is cleared or changed when another agent is selected.

2. Confirm agentic scope

Document:

- Trigger and intended outcome.
- Inputs, memory and retrieval.
- Models and dependent agents.
- Tools, APIs and reachable systems.
- Read, write, execute and administrative capability.
- Human approvals and escalation.
- Users and affected people.
- Data sensitivity and jurisdictions.
- Maximum single and cumulative impact.
- Failure, misuse and incident scenarios.

3. Set current and target levels

Record the presently approved operating level and the proposed target. If the current authority does not match the documented level, raise a finding immediately.

4. Work through all five elements

Assess every canonical requirement. The target-level matrix tells you whether the requirement is a MUST, SHOULD or MAY, but it does not replace contextual judgement.

5. Record the result and depth

For each requirement, choose:

- **Fully met**
- **Partially met**
- **Not addressed**

Then record depth:

Result	Depth labels
Fully met	1 Ad hoc, 2 Defined, 3 Embedded
Partially met or Not addressed	1 Not started, 2 In progress, 3 Near complete

Partially met and Near complete remain blockers for a target-level MUST. Use the normative level matrix and documented exception governance for SHOULD or MAY treatment.

6. Read the requirement advisory

Use the on-screen playbook to:

- Understand the control objective.
- Identify minimum, production and advanced practices.
- Ask agent-specific questions.
- Inspect suitable evidence.
- Design a safe test.
- Recognise common false assurance.

Technology examples are options, not mandatory products.

7. Link objective evidence

Prefer evidence that proves both design and operation: identity records, policy decisions, traces, denials, approvals, monitoring, incident exercises and configuration-change records.

8. Test enforcement

Test the negative path. Attempt an unauthorised resource, action, volume or transaction within an approved non-destructive environment and verify that enforcement, logging and response all occur.

9. Record findings and incidents

Reflect failed tests, stale evidence, operating exceptions and incident history in the requirement result. Do not retain a favourable result that is contradicted by current adverse evidence.

10. Complete promotion gates

For promotion, work through Performance, Security Validation, Business Value, Incident Record and Governance Sign-off. Attach evidence to each gate rather than relying on a single approval statement.

11. Decide

Record:

- Maintain current level.
- Hold promotion.
- Approve promotion.
- Demote.
- Suspend or contain pending investigation.

Include decision owner, conditions, residual risk and next review.

12. Maintain

Reassess after material change, incident, drift, failed test, new tool or privilege, changed purpose, new affected population, or expiry of evidence.

Completion checklist

54. Reference & glossary

ATF labels, definitions, frequently asked questions and safe language for explaining the Gamut assessment.

Key labels

Label	Meaning
Agent	A system component that pursues goals and can use delegated capability to observe, recommend or act
Element	One of Identity, Behavior, Data Governance, Segmentation or Incident Response
Requirement	One of the 25 canonical ATF control statements
MUST	Required for conformance at the selected maturity level
SHOULD	Expected unless a justified and reviewed exception exists
MAY	Optional in ATF at that level; other obligations can still require it
Current level	Presently approved authority
Target level	Proposed future authority; not an approval
Promotion gate	Evidence checkpoint for performance, security, value, incidents or governance
Hold	Maintain present authority but do not promote
Demotion	Reduce authority because current trust conditions are not met
Privacy Mode	Reduced AI-assistance context without free text or direct identifiers

Frequently asked questions

Is ATF an enacted law or ISO standard?

No. It is an open agent-governance specification.

Is it final?

The canonical repository currently describes Specification 0.9.1 as a Public Review Draft. Verify the repository before making a time-sensitive version claim.

Does completing all 25 items make an agent safe?

No. The assessment supports a bounded governance decision. Threat modelling, impact assessment, legal review, operational monitoring and incident readiness remain necessary.

Does a higher level mean a better agent?

No. It means greater approved autonomy and therefore a higher control burden. Many agents should remain at Intern or Junior.

Can an agent skip levels?

Sequential promotion is the defensible default because each level supplies operating evidence. Avoid skipping the observation and review periods.

Can a partial result satisfy a MUST?

No. A target-level MUST must be Fully met. Partially met and Not addressed are blockers.

Does AI Assist approve a promotion?

No. It provides saved, per-requirement advice for human review.

Canonical resources

- ATF repository
- ATF specification site
- CSA introduction

- NIST SP 800-207 Zero Trust Architecture

Version statement template

> This assessment used ATF Specification 0.9.1 and Conformance 0.9.0, identified by the canonical project as a Public Review Draft. Results are agent-specific and do not represent certification.

55. Reporting & framework relationships

How to report ATF maturity without overstating conformance and how ATF works with ACRS, MAESTRO, GTSAF and wider governance frameworks.

Agent-level reporting

An ATF report should identify:

- Agent, version, owner and purpose.
- Current and target level.
- Requirement results by element.
- MUST gaps and SHOULD exceptions.
- Evidence and testing depth.
- Open incidents and findings.
- Promotion-gate position.
- Decision, conditions, approvers and review trigger.

Do not report a level without its scope and basis.

Portfolio reporting

Portfolio views can show:

- Agents assessed versus registered.
- Current and target level distribution.
- Readiness by element.
- Agents held below requested authority.
- Common MUST gaps.
- Weak evidence or overdue tests.
- Incidents, demotions and promotion reviews.

A portfolio average must not conceal a high-authority agent with a failed containment requirement.

Safe conclusion language

Prefer:

> The selected agent was assessed against ATF Specification 0.9.1 for a target of L2 Junior. The current record supports maintaining L1 pending completion of action-boundary testing and the Security Validation gate.

Avoid:

- "ATF certified."
- "Zero trust compliant" without scope and evidence.
- "Safe for autonomous operation."
- "All agentic risks eliminated."

Relationship with ACRS

ACRS characterises capability risk: what the system can do and how much impact that capability creates. ATF governs how authority is earned and constrained.

A high ACRS position should increase assurance and testing depth and may justify a lower initial ATF level. It does not mechanically decide a requirement result or promotion.

Relationship with MAESTRO

MAESTRO identifies layer-specific agentic threats. ATF establishes zero-trust governance controls that help prevent, detect and contain them.

Use MAESTRO findings to challenge ATF requirements. For example, tool and orchestration threats should inform action-boundary, attribution, injection-defence and containment tests.

Relationship with GTSAF

GTSAF provides the wider assurance baseline across governance, data, models, security, people, suppliers and operations. ATF focuses on runtime trust for a named agent.

Crosswalks support traceability and evidence reuse; they do not create automatic equivalence.

Relationship with other frameworks

- NIST AI RMF provides the broader risk-management outcomes.
- ISO/IEC 42001 provides the organisational AI management system.
- ISO/IEC 42005 examines impacts on people and society.
- EU AI Act and NAGF determine applicable legal and regulatory obligations.

ATF does not replace any of these. The agent may need all of them.

Governance review questions

56. AI Assist and security

Public guide to system-scoped EU AI Act AI Assist, legal and assessment information considered, Privacy Mode, saved outputs and human legal accountability.

EU AI Act AI Assist supports analysis of one atomic legal requirement for the selected system. It does not make the legal decision, change routing, approve N/A, accept evidence or confirm compliance.

System and requirement scope

The panel displays the selected system and requirement. Switching systems changes the assessment context; a result generated for one system is not presented as the current result for another.

No unscoped analysis should be treated as a legal conclusion.

Full-context information considered

Subject to authorised access, analysis may consider:

- Selected system, purpose and lifecycle state.
- EU territorial nexus and definition/exclusion determination.
- Organisation roles and value-chain position.
- Article 5 atomic screening.
- Annex I and Annex III classification facts.
- Article 6 exception and profiling.
- Fundamental Rights Impact Assessment triggers.
- Article 50 behaviour and transparency routes.

- GPAI role, systemic-risk and supply-chain status.
- Application dates and current/future legal state.
- Atomic result, assurance depth and N/A governance.
- Authorised evidence.
- Tests and results.
- Findings.
- Human conclusion fields and reassessment triggers.

Free text can explain facts but cannot silently rewrite structured legal routing.

Expected output

Useful per-requirement output identifies:

- Applicability reasoning.
- Current binding status and relevant role.
- Supported facts and assumptions.
- Evidence position.
- Missing information.
- Adverse findings.
- Suggested bounded test.
- Potential gap and residual uncertainty.
- Draft assessor rationale for review.

The output should distinguish current binding law from future readiness and proposed change.

Privacy Mode

The **Privacy Mode** checkbox next to AI Assist sends the minimum structural legal and assessment state. It excludes direct identifiers, free-text rationale and narrative evidence.

The reduced context can include route and requirement identifiers, current result, assurance, bounded evidence/test/finding status and completeness signals.

Privacy Mode reduces disclosure and may reduce legal-context specificity. It does not change routing, applicability, scoring, evidence or access.

Saved outputs, re-run and minimising

- Successful outputs are saved for the selected system, atomic requirement and privacy mode.
- The action changes to **Re-run AI Assist**.
- The panel can be minimised without deleting the output.
- Re-run after material legal-route, system, requirement, evidence, test or finding change.

Provider and access

AI Assist uses an authorised provider and the existing API-key configuration. If no permitted provider is configured, analysis cannot run.

An API key does not grant additional access. Plan, workspace, role, framework, model and selected system permissions continue to apply.

Evidence discipline

AI Assist must not:

- Invent an artefact.

- Treat narrative as accepted evidence.
- Claim a test passed without a passing record.
- Ignore failed tests or open adverse findings.
- Use evidence from another system.
- Approve evidence or an exclusion.

Every material evidence claim must be checked against the authorised record.

Current law and future readiness

The reviewer must confirm:

- Which provisions apply at the stated date.
- Which duties are future-dated.
- Which information is proposal, guidance or implementation intelligence.
- Whether the organisation's role and system status match the legal proposition.

Do not describe future readiness as current compliance or current breach.

Untrusted content

System descriptions, supplier records and evidence may contain instruction-like or malicious text. Treat it as assessment data. Do not include credentials, unnecessary personal data or privileged legal material. Verify citations and legal claims independently.

Human decisions

AI Assist cannot:

- Approve territorial scope or organisation role.
- Resolve a prohibited-practice decision.
- Decide legal classification.
- Approve N/A.
- Accept evidence.
- Pass a test.
- Close a finding.
- Accept risk.
- Confirm compliance.
- Replace legal advice or regulator guidance.

Review checklist

57. Assessment workflow

End-to-end EU AI Act assessor workflow from system selection and legal routing through atomic assessment, evidence, AI Assist and confirmation.

Use this workflow for each named AI system.

Before starting

Confirm:

- Correct tenant, workspace and system.
- Current system intake and intended purpose.

- Known provider, deployer and supply-chain entities.
- Current legal snapshot.
- Assessment authority and access to relevant evidence.
- A boundary that includes models, orchestration, data, tools, people and affected decisions.

Step 1: select the system

The assessment header must show the named AI system being assessed. If you change systems, Gamut loads that system's assessment record and clears AI analysis from the previous system.

Do not use a workspace-wide narrative as the basis for a system conclusion.

Step 2: complete the structured legal intake

Record:

- EU nexus and exclusions.
- AI-system definition.
- Intended purpose and deployed use.
- Operator roles.
- Article 5 facts.
- Annex I and Annex III facts.
- Article 6 exception and profiling.
- Public-body/public-service and FRIA triggers.
- Article 50 behaviours.
- GPAI role and systemic-risk facts.
- Applicable market-placement and application dates.

Use free text to explain facts, not to replace required structured fields.

Step 3: review the authoritative route

Review all eleven route cards. For each, confirm:

- Applicable, not applicable, incomplete or future status.
- Legal basis.
- Triggering system facts.
- Role.
- Current-law application date.
- Any contradiction or missing fact.

If a route looks wrong, correct the intake. Do not force the answer inside an assessment item.

Step 4: resolve Article 5 first

Work through all eight prohibited-practice checks before relying on broader readiness scores.

If a check is potential or confirmed:

- Stop deployment or expansion according to organisational governance.
- Preserve the relevant facts and evidence.
- Escalate to legal, compliance and accountable leadership.
- Do not use N/A to bypass uncertainty.
- Do not confirm the assessment until the issue is resolved.

Step 5: validate high-risk classification

Assess:

1. Annex I product and conformity conditions.
2. All eight Annex III points.
3. Article 6(3) exception factors.
4. Profiling override.
5. Provider documentation and registration where an Annex III system is concluded not high-risk.

Record both the classification conclusion and its factual/legal basis.

Step 6: validate roles and downstream categories

For each entity in the value chain, determine applicable duties. Pay particular attention to:

- Provider/deployer overlap.
- Product manufacturer.
- Third-country authorised representative.
- Importer and distributor gates.
- Rebranding or substantial modification.
- GPAI model provider versus downstream system provider.

Step 7: assess every routed atomic item

Open each requirement and read:

- Legal basis and effective date.
- Role and applicability.
- Assessment question.
- Layer-specific auditor advisory.
- Evidence to inspect.
- Bounded audit test.
- Implementation notes.
- Cross-framework references.

For parent items with children, assess every child. A parent-level narrative does not complete the atomic items.

Step 8: choose the legal answer

Compliant

Select only where the applicable obligation is implemented. Then assign the assurance depth that the evidence, testing and findings support.

Non-compliant

Select where the obligation is absent, materially incomplete, ineffective or contradicted.

Record:

- Precise condition.
- Legal and operational consequence.
- Finding and owner.
- Interim restriction or compensating control.
- Target date and retest.

N/A

Select only after establishing that the trigger does not apply. Complete the governed N/A decision; otherwise it remains unresolved and is not excluded from scope.

Step 9: assign assurance depth

Use:

- **3 - Assured**
- **2 - Supported/Partial**
- **1 - Asserted**
- **0 - Gap**

The answer says whether the legal requirement is met. Depth says how strongly that claim is demonstrated. See Scoring, assurance and conclusions.

Step 10: link evidence

Evidence must belong to:

- The selected system.
- The relevant atomic item.
- The applicable organisation and role.
- The current version and period.

Review and accept evidence; upload count alone is not assurance.

Step 11: perform bounded tests

Use the item-specific audit test as a starting point. Record objective, scope, sample, steps, pass criteria, safety constraints, result, exceptions and retest.

Legal-document checks may use inspection and traceability tests. Technical duties require appropriate operating-effectiveness tests.

Step 12: raise and link findings

Raise a finding for:

- Non-compliance.
- Failed test.
- Rejected or stale evidence.
- Role or classification contradiction.
- Unsupported N/A.
- Missing authority, notification or escalation process.
- Open issue that contradicts a compliant claim.

Step 13: use AI Assist

Choose:

- Whole-system analysis; or
- One atomic requirement.

Before applying any suggestion, confirm the displayed system, route basis, evidence scope and legal snapshot. AI output is advisory.

Step 14: complete the human conclusion

Required fields include:

Field	Required content
Assessment owner	Accountable named person
Confidence	Low, Medium or High
Residual risk	Remaining exposure, gaps, uncertainty and restrictions
Review date	Date proportionate to change and legal timing
Conclusion	At least 80 characters; system-specific and legally bounded
Reassessment triggers	At least 30 characters; observable changes or events

Separate:

- Current-law compliance conclusion.
- Future-readiness position.
- Proposed-change intelligence.

Step 15: confirm

Confirmation is validation-gated. For every applicable, in-force atomic item claimed compliant, the record must support depth 3 with:

- Accepted/reviewed evidence.
- A passing system-scoped test.
- No failed test.
- No unresolved adverse finding.

Applicable items may instead have a fully approved N/A where legally valid. Missing scope facts, contradictions, potential prohibited practices or incomplete assurance block confirmation.

Returning to the main assessment

When working in an atomic assessment, use **Back** to return to the main EU AI Act view. The selected system and saved record remain in context.

Reassessment triggers

Reassess after:

- Intended-purpose or user-population change.
- New EU market or output nexus.
- New or transferred operator role.
- Model, supplier or GPAI status change.
- Substantial modification, rebranding or fine-tuning.
- New Annex I/III use.
- Added profiling, biometrics, emotion recognition or synthetic-content behaviour.
- Changed autonomy, tools, data or decision effect.
- Serious incident or regulatory request.
- Failed test or material finding.
- Evidence expiry.
- New binding law, guidance or application date.

Final quality checklist

58. Atomic requirement catalogue

Complete catalogue of Gamut's 34 EU AI Act parent requirements and 72 independently assessed atomic legal checks.

Gamut decomposes broad legal topics into **72 atomic assessment items**. This prevents a favourable answer to one part of an article from masking a gap in another.

How to read the catalogue

- A parent with no children is itself atomic.
- A parent with children is a navigation and advisory container; each child receives its own answer, assurance depth and evidence position.
- Sibling checks cannot satisfy one another.
- Applicability is still determined by the named system's route, role and legal status.

Scope, role and timing

ID	Requirement	Legal basis
EU-AI-00	Scope, role and EU nexus	Articles 2, 3, 113
EU-AI-01	AI-system definition and intended purpose	Article 3
EU-AI-06	AI literacy	Article 4

Prohibited practices

EU-AI-02 is the Article 5 parent and hard-stop screen.

Atomic ID	Prohibited-practice check	Legal basis
EU-AI-02.a	Subliminal, manipulative or deceptive techniques	Article 5(1)(a)
EU-AI-02.b	Exploitation of vulnerabilities	Article 5(1)(b)
EU-AI-02.c	Social scoring	Article 5(1)(c)
EU-AI-02.d	Criminal-offence risk assessment based solely on profiling	Article 5(1)(d)
EU-AI-02.e	Untargeted facial-image scraping	Article 5(1)(e)
EU-AI-02.f	Emotion inference in workplace or education	Article 5(1)(f)
EU-AI-02.g	Sensitive biometric categorisation	Article 5(1)(g)
EU-AI-02.h	Real-time remote biometric identification for law enforcement	Article 5(1)(h)

High-risk classification

ID	Requirement	Legal basis
EU-AI-03	Annex I product-safety route	Article 6(1), Annex I
EU-AI-05	Article 6 exception and profiling override	Article 6(3), 6(4), 49(2)

EU-AI-04 expands Annex III:

Atomic ID	Area	Legal basis
EU-AI-04.1	Biometrics	Annex III point 1
EU-AI-04.2	Critical infrastructure	Annex III point 2
EU-AI-04.3	Education and vocational training	Annex III point 3
EU-AI-04.4	Employment and worker management	Annex III point 4
EU-AI-04.5	Essential private and public services	Annex III point 5
EU-AI-04.6	Law enforcement	Annex III point 6
EU-AI-04.7	Migration, asylum and border control	Annex III point 7

Atomic ID	Area	Legal basis
EU-AI-04.8	Justice and democratic processes	Annex III point 8

High-risk system requirements

ID	Requirement	Legal basis
EU-AI-07	Risk-management system	Articles 8, 9
EU-AI-08	Data governance and quality	Article 10
EU-AI-09	Technical documentation	Article 11, Annex IV
EU-AI-10	Record-keeping and logging	Article 12; Articles 19, 26
EU-AI-11	Transparency and instructions for use	Article 13
EU-AI-12	Human oversight	Article 14; Article 26(2)
EU-AI-13	Accuracy, robustness and cybersecurity	Article 15

Assess these across the lifecycle, including foreseeable misuse, data fitness, change control, logging integrity, oversight authority, adversarial resilience, failure handling and current technical documentation.

Operators, conformity and value chain

ID	Requirement	Legal basis
EU-AI-14	Provider obligations and quality management	Articles 16-20
EU-AI-15	Conformity, declaration, CE marking and registration	Articles 43, 47-49
EU-AI-25	Cooperation with competent authorities	Article 21
EU-AI-26	High-risk provider authorised representative	Article 22
EU-AI-27	Importer duties	Article 23
EU-AI-28	Distributor duties	Article 24
EU-AI-29	Value-chain role transfer and substantial modification	Article 25
EU-AI-30	Right to explanation of individual decision-making	Article 86

EU-AI-16 expands deployer duties:

Atomic ID	Deployer check	Legal basis
EU-AI-16.1	Use follows provider instructions	Article 26(1)
EU-AI-16.2	Competent human oversight	Article 26(2)
EU-AI-16.3	Input-data relevance	Article 26(4)
EU-AI-16.4	Monitoring based on instructions	Article 26(5)
EU-AI-16.5	Risk and incident escalation	Article 26(5)
EU-AI-16.6	Log retention under deployer control	Article 26(6)
EU-AI-16.7	Workplace notification	Article 26(7)
EU-AI-16.8	Public-authority registration check	Article 26
EU-AI-16.9	DPIA and FRIA linkage	Articles 26, 27
EU-AI-16.10	Minimum six-month deployer log retention	Article 26(6)
EU-AI-16.11	Notice to persons subject to high-risk decisions	Article 26(11)
EU-AI-16.12	Post-remote-biometric-identification safeguards	Article 26(10)

Fundamental-rights impact assessment

ID	Requirement	Legal basis
EU-AI-17	FRIA applicability and completion	Article 27

The advisory covers scope, process, affected groups, impacts, oversight, mitigation, complaint and redress, residual risk, notification and review. See Evidence, testing and findings.

Article 50 transparency

`EU-AI-18` expands into trigger-specific checks:

Atomic ID	Transparency check	Legal basis
`EU-AI-18.1`	Direct AI-interaction notice	Article 50(1)
`EU-AI-18.2`	Synthetic-content marking	Article 50(2)
`EU-AI-18.3`	Emotion-recognition notice	Article 50(3)
`EU-AI-18.4`	Biometric-categorisation notice	Article 50(3)
`EU-AI-18.5`	Deepfake disclosure	Article 50(4)
`EU-AI-18.6`	Public-interest text disclosure	Article 50(4)
`EU-AI-18.7`	First-interaction or exposure timing	Article 50
`EU-AI-18.8`	Clarity, distinguishability and accessibility	Article 50
`EU-AI-18.9`	Transparency exceptions	Article 50

High-risk monitoring and incidents

ID	Requirement	Legal basis
`EU-AI-19`	High-risk post-market monitoring	Article 72
`EU-AI-20`	High-risk serious-incident reporting	Article 73; Article 3(49); Article 26(5)

GPAI systemic-risk monitoring

ID	Requirement	Legal basis
`EU-AI-23`	Systemic-risk evaluation and mitigation	Article 55
`EU-AI-24`	Systemic-risk incident and cybersecurity controls	Article 55

GPAI and model supply chain

`EU-AI-21` captures the organisation's position in the model value chain:

Atomic ID	GPAI role check	Legal basis
`EU-AI-21.1`	Direct GPAI model provider	Articles 53, 54
`EU-AI-21.2`	GPAI model with systemic risk	Article 55
`EU-AI-21.3`	Authorised representative	Article 54
`EU-AI-21.4`	Downstream provider integrating GPAI	Articles 53, 56
`EU-AI-21.5`	Deployer/API consumer supplier evidence	Chapter V
`EU-AI-21.6`	Copyright policy and training summary	Article 53

Additional atomic requirements:

ID	Requirement	Legal basis
`EU-AI-31`	GPAI classification and systemic-risk designation	Articles 51, 52
`EU-AI-32`	Technical documentation and downstream information	Article 53, Annexes XI and XII
`EU-AI-33`	GPAI provider authorised representative	Article 54

Non-applicability

ID	Requirement	Legal basis
`EU-AI-22`	N/A rationale and residual evidence trail	Articles 2, 6, 27, 50 and Chapter V

N/A is a governed legal conclusion. It requires a trigger-specific rationale, legal basis, factual evidence, decision owner, independent approver, approval date and reassessment trigger.

Totals

Measure	Count
Routed categories	11
Parent requirements	34
Atomic items actually assessed	72
Article 5 atomic checks	8
Annex III atomic checks	8
Article 26 atomic checks	12
Article 50 atomic checks	9
GPAI role atomic checks under `EU-AI-21`	6

These totals describe the methodology version above. Future legal or product changes should update the methodology identifier, legal snapshot and catalogue together.

59. Evidence, testing and findings

Practical EU AI Act evidence expectations, bounded tests, finding rules and route-specific assessor guidance.

Evidence and testing convert a legal answer from assertion into assurance. They must demonstrate the named system, regulated role, relevant version and assessment period.

Evidence quality

Review each item for:

- Relevance to the exact atomic obligation.
- System and entity scope.
- Version and environment.
- Currency and operating period.
- Authenticity and integrity.
- Completeness.
- Independent review.
- Contradictory information.

Generic supplier marketing, a policy with no implementation record, or an assessor narrative is not automatically accepted evidence.

Testing principles

A defensible test records:

- Objective.
- Legal/technical requirement.
- System, role and environment.
- Population and sample.
- Steps.
- Expected result and pass criteria.
- Safety and privacy limits.
- Tester and date.

- Actual result and exceptions.
- Evidence captured.
- Retest requirement.

Use non-destructive environments and synthetic data where possible. Production or adversarial tests require explicit authority and safeguards.

Scope, role and AI literacy

Evidence may include:

- Corporate and market-placement records.
- System/model inventory.
- Intended-purpose and architecture documents.
- Contracts and role matrices.
- Geographic output/use analysis.
- Exclusion memorandum.
- Training needs analysis, role-specific curriculum and comprehension results.

Test by tracing one system from supplier through deployed use and asking whether each factual role and EU nexus is supported by independent records.

Prohibited practices

Evidence should address the actual feature and use, not merely policy language:

- Product requirements and prohibited-use controls.
- Model and feature configuration.
- Data-source records.
- User journeys and decision logic.
- Marketing and sales claims.
- Access restrictions.
- Red-team or misuse tests.
- Legal analysis for any narrow exception.

Test each of the eight Article 5 points. Where a dangerous capability could be enabled through an alternate endpoint, tool or configuration, include that path.

Any positive or uncertain result requires escalation and blocks confirmation.

High-risk classification

Inspect:

- Product legislation and conformity route.
- Annex III use-case mapping.
- Intended purpose and actual workflow.
- Affected decisions and persons.
- Profiling analysis.
- Article 6(3) exception memorandum.
- Article 6(4) documentation and registration, where relevant.

Test classification by reconstructing the decision from facts without relying on the existing label. Independently challenge industry-based assumptions.

Risk management

Evidence:

- Continuous lifecycle risk-management plan.
- Hazard, misuse and fundamental-rights scenarios.
- Risk estimates and acceptance criteria.
- Testing and validation plans.
- Residual-risk decisions.
- Change and monitoring integration.

Test whether a material model, data or intended-purpose change updates hazards, mitigations, validation and residual-risk approval.

Data governance and quality

Evidence:

- Data provenance, collection and preparation records.
- Relevance and representativeness analysis.
- Bias and error analysis.
- Data quality thresholds.
- Special-category handling.
- Dataset versioning and lineage.
- Remediation and exception records.

Test traceability from a sampled output or failure back to the data version and quality controls.

Technical documentation and instructions

Evidence:

- Annex IV technical file.
- Architecture and model description.
- Performance and limitation results.
- Risk, data, logging, oversight and cybersecurity sections.
- Version/change history.
- Instructions for use.

Test whether the file matches the released system and whether a material change triggers controlled updates before release.

Logging

Evidence:

- Logging design and schema.
- Event samples.
- Integrity, access and retention controls.
- Clock synchronisation.
- Privacy minimisation.
- Retrieval and investigation procedures.

Test that a representative decision or incident can be reconstructed, and that deployer-controlled logs meet the applicable minimum retention period.

Human oversight

Evidence:

- Named oversight roles.
- Competence and training.
- Decision authority.
- Interface and alert design.
- Override, stop and rollback mechanisms.
- Workload and automation-bias controls.
- Exercise results.

Test a consequential scenario from abnormal output to detection, comprehension, intervention and safe recovery. A nominal human presence is insufficient if intervention is too late or lacks authority.

Accuracy, robustness and cybersecurity

Evidence:

- Declared metrics and tolerances.
- Validation across relevant groups and conditions.
- Robustness, drift and failure testing.
- Secure development and vulnerability management.
- Prompt-injection, poisoning, evasion and extraction testing where relevant.
- Incident and recovery records.

Test the actual deployed configuration and foreseeable conditions, including interfaces and dependencies.

Provider and conformity duties

Evidence:

- Quality-management system.
- Technical-document retention.
- Conformity assessment.
- EU declaration of conformity.
- CE marking.
- EU database registration.
- Corrective-action and withdrawal process.
- Authority-cooperation process.

Test a release gate using a missing or inconsistent mandatory artefact. The system should fail closed.

Deployer duties

Test the twelve atomic duties separately, including:

- Following instructions.
- Competent oversight.
- Input-data fitness.
- Monitoring and abnormal-operation escalation.
- Incident communication.
- Protected logs and minimum six-month retention.
- Workplace notification.
- Public-authority registration checks.

- DPIA/FRIA linkage.
- Notice to affected persons.
- Post-remote-biometric safeguards.

Importer, distributor and value chain

Evidence:

- Pre-market checklists.
- Provider and representative records.
- CE/declaration/instruction checks.
- Storage and transport controls.
- Stop, recall and corrective-action authority.
- Substantial-modification assessment.
- Information and assistance rights.

Test whether a missing conformity artefact actually blocks release, and whether a material change causes provider-role and conformity reassessment.

Fundamental-rights impact assessment

Evidence should cover:

- Deployer process and intended purpose.
- Duration and frequency.
- Affected natural-person categories and groups.
- Specific risks to fundamental rights.
- Human oversight.
- Mitigation and governance.
- Complaint, contestability and redress.
- Residual risk and approval.
- Authority notification where required.
- Review triggers.

Test the FRIA against a credible affected-person scenario. Confirm it changes design or deployment where the analysis identifies unacceptable or poorly controlled impact.

Article 50 transparency

Inspect the actual user experience:

- Notice copy and placement.
- First-interaction timing.
- Accessibility.
- Machine-readable marking.
- Deepfake or public-interest disclosure.
- Emotion/biometric notices.
- Language and channel variants.
- Exception analysis.

Test each trigger across representative interfaces, API outputs and content transformations. Verify that downstream processing does not strip required markings.

Monitoring and serious incidents

Evidence:

- Post-market monitoring plan.
- Telemetry and complaint sources.
- Trend and threshold analysis.
- Serious-incident definition and triage.
- Provider/deployer escalation.
- Authority reporting decisions and timelines.
- Corrective action.

Run a tabletop incident from detection through legal classification, containment, preservation, notification and follow-up.

GPAI

Depending on role, inspect:

- Model/version and market-placement record.
- Systemic-risk classification.
- Annex XI technical documentation.
- Annex XII downstream information.
- Copyright policy and rights-reservation controls.
- Public training-content summary.
- Open-source exemption analysis.
- Evaluation, systemic-risk mitigation and cybersecurity.
- Serious-incident reporting.
- Authorised representative mandate.
- Supplier evidence for deployers/API consumers.

The voluntary GPAI Code can be useful evidence, but code adherence is not conclusive proof for every fact or obligation.

Findings

Raise a finding when evidence or testing shows:

- Legal non-compliance.
- Unsupported classification or N/A.
- Failed operating control.
- Stale or rejected evidence.
- Inadequate scope or role determination.
- Unresolved contradiction.
- Missing notification or authority-cooperation capability.

A useful finding includes condition, requirement, evidence, affected system and role, consequence, root cause, severity, recommendation, owner, target date, status and retest.

Evidence-to-conclusion rules

60. Legal status, scope and roles

EU AI Act source hierarchy, application dates, territorial scope, exclusions, AI-system definition and operator-role analysis in Gamut.

This page explains the legal foundation that must be established before an assessor decides which substantive EU AI Act duties apply.

> **Legal review:** Scope, role and application-date decisions can change the entire assessment population. Obtain qualified legal review where the EU nexus, statutory exclusion, provider role, substantial modification or an exception is material or uncertain.

Source hierarchy

Use sources in this order:

1. The binding text of Regulation (EU) 2024/1689, including its annexes.
2. Formally adopted amending law published in the Official Journal.
3. Delegated or implementing acts in force.
4. Court decisions and decisions of competent authorities where relevant.
5. Commission guidelines as non-binding interpretive material.
6. Codes of practice, standards and common specifications according to their legal status.
7. Organisational policy and professional judgement.

Do not replace the Regulation with a vendor summary, blog, draft guideline or political agreement.

Gamut legal snapshot

The implemented methodology is reviewed as of **16 July 2026** and identifies itself as `EUAIA-2026-07-16-defensible-system-scope`.

The documentation and official-source status were rechecked on **21 July 2026**. The implemented identifier remains unchanged so saved assessments and workpapers stay reproducible. Formally adopted amendments require a versioned methodology update; political agreement, draft guidance or consultation material remains implementation intelligence until its legal status supports different treatment.

The record distinguishes the following:

Status	Assessment treatment
In force	Included in the current-law conclusion when applicable
Future	Tracked as readiness; not counted as a current breach
Proposed change only	Shown as implementation intelligence, not binding law
Non-binding guidance	Used to inform interpretation and testing, clearly labelled

Always show the snapshot date in exported workpapers. A legal conclusion without an "as of" date is incomplete.

Application timeline

Date	Position represented in Gamut
1 August 2024	Regulation entered into force
2 February 2025	Chapters I and II apply, including AI literacy and prohibited practices
2 August 2025	Governance provisions and GPAI provider duties apply
2 August 2026	General application date under the adopted Regulation, subject to specific exceptions and any formally adopted amendment
2 August 2027	Original Regulation date for Article 6(1) Annex I product-safety high-risk systems

As of the Gamut snapshot, Commission materials describe political agreement on adjusted high-risk dates. Gamut tracks proposed dates as readiness information but does not silently substitute a political agreement for formally adopted and

published law. Assessors should check the Commission implementation page and the Official Journal when making a live decision.

Territorial and material scope

Article 2 analysis should record, at minimum:

- Where each provider, deployer, importer and distributor is established.
- Where the system or GPAI model is placed on the market or put into service.
- Whether output produced by the system is used in the Union.
- The affected-person and customer geography.
- Whether the organisation is acting as a public authority, Union institution or private operator.
- Whether a statutory exclusion is claimed.
- The relevant system or model version and intended purpose.

The assessment must not infer "out of scope" merely because the supplier or infrastructure is outside the EU.

Exclusions

Where relevant, assess and evidence exclusions such as:

- National-security, military or defence use within the statutory terms.
- Certain public-authority cooperation with third countries or international organisations.
- Research, testing or development before market placement or putting into service, subject to the statutory limits.
- Natural persons using AI in a purely personal non-professional activity.
- AI released under free and open-source licences, only to the extent and under the conditions the Regulation provides.

An exclusion is not a label of convenience. Record the exact legal provision, facts, boundary, owner, approver, evidence and reassessment trigger.

Is it an AI system?

Gamut requires an Article 3 definition analysis rather than assuming every automated tool is, or is not, an AI system.

Document:

- Machine-based nature.
- Degree of autonomy.
- Whether it may exhibit adaptiveness after deployment.
- Objectives, whether explicit or implicit.
- How inputs are used to infer outputs.
- **Output types:** predictions, content, recommendations or decisions.
- How those outputs influence physical or virtual environments.
- Intended purpose and actual deployed use.

The system boundary should include orchestration, retrieval, prompts, tools, human decision points and downstream effects where those elements determine the regulated functionality.

Intended purpose and reasonably foreseeable use

The intended purpose affects high-risk classification, technical documentation, instructions, conformity and role transfer. Record:

- The provider's stated purpose.
- The deployer's actual use.
- Target persons and operating context.

- Excluded uses and technical restrictions.
- Marketing and contractual claims.
- Material changes since release.
- Foreseeable misuse relevant to risk management.

If actual use has moved beyond the documented intended purpose, reassess classification and whether the deployer or integrator has become a provider under Article 25.

Operator roles

Roles are determined by facts, not job titles or contract labels.

Role	Core question
Provider	Who develops, has developed, or places the system/model on the market under its name or trademark?
Deployer	Who uses the AI system under its authority in a professional context?
Importer	Who established in the Union places a third-country high-risk system on the Union market?
Distributor	Who makes the system available in the supply chain without being provider or importer?
Product manufacturer	Who places a product containing or using the high-risk system on the market under its name?
Authorised representative	Who is mandated and established in the Union to perform specified provider tasks?
GPAI model provider	Who develops or has developed and places the general-purpose AI model on the market?
Downstream provider	Who integrates a model into an AI system placed on the market or put into service?

One organisation can hold several roles for one system. Different group entities can hold different roles.

Role transfer and substantial modification

Article 25 can transfer provider obligations where a party:

- Places a high-risk system on the market under its own name or trademark.
- Makes a substantial modification.
- Changes the intended purpose so a system becomes high-risk.

Review fine-tuning, model replacement, safety-control changes, new tools, autonomy, data, user population, market claims and workflow changes. Contracts can allocate assistance and information but cannot contract away a statutory role.

Authorised representatives

Keep these routes separate:

- **Article 22**: representative for a third-country provider of a high-risk AI system.
- **Article 54**: representative for a third-country GPAI model provider.

The mandate, tasks, documentation access and exceptions differ. Representative access should be purpose-bound, least-privilege, time-controlled and audited.

AI literacy

Article 4 applies to providers and deployers. A defensible programme is role- and context-specific, not a generic annual awareness slide.

Inspect:

- Personnel and contractor roles.
- Technical knowledge, experience and education.

- System context and affected persons.
- Risks associated with the specific use.
- Training content, attendance and comprehension.
- Refresh triggers.
- Escalation and safe-use procedures.

Minimum scope-and-role record

- [ ] Named system, version and boundary.
- [ ] Intended purpose and actual use.
- [ ] EU territorial nexus.
- [ ] Article 3 AI-system determination.
- [ ] Every operator role with supporting facts.
- [ ] Claimed exclusions with legal basis.
- [ ] Current and future application dates.
- [ ] Provider/deployer and GPAI value-chain relationships.
- [ ] Substantial-modification analysis.
- [ ] Owner, approver, evidence and review trigger.

Official sources

61. Reference and glossary

EU AI Act quick reference for Gamut routes, roles, statuses, dates, scoring, confirmation, terminology and frequently asked questions.

Method reference

Item	Value
Regulation	EU 2024/1689
Snapshot	16 July 2026
Documentation source check	21 July 2026
Methodology	`EUAIA-2026-07-16-defensible-system-scope`
Categories	11
Parent requirements	34
Atomic items	72
Assessment unit	Named AI system

Route reference

Code	Meaning
`SCOPE`	Scope, definition, role, timing, literacy
`PROHIBITED`	Article 5 hard-stop screen
`CLASSIFICATION`	Article 6, Annex I and Annex III
`HIGHREQ`	Articles 8-15
`OPERATORS`	Operator and conformity duties
`FRIA`	Article 27
`TRANSPARENCY`	Article 50

Code	Meaning
`MONITORING`	Articles 72-73
`GPAI_MONITORING`	Article 55
`GPAI`	Chapter V model and supply-chain duties
`NA`	Governed non-applicability trail

Answer and depth reference

Dimension	Values
Legal answer	Compliant, Non-compliant, N/A, Unassessed
Depth	3 Assured, 2 Supported/Partial, 1 Asserted, 0 Gap
Confidence	Low, Medium, High
Legal state	In force, Future obligation, Proposed change only

Route-state reference

State	Meaning
`not_assessed`	No route decision
`applicable`	Triggered
`not_applicable_with_evidence`	Excluded through an approved decision
`incomplete`	Facts or decision missing
`implemented`	Routed work addressed, subject to assurance
`prohibited_stop`	Article 5 stop
`future_obligation`	Future readiness
`proposed_change_only`	Not binding law

Key dates

Date	Significance
1 August 2024	Entry into force
2 February 2025	AI literacy and prohibited practices apply
2 August 2025	GPAI provider and governance provisions apply
2 August 2026	General application date in the adopted Regulation, subject to exceptions/amendments
2 August 2027	Original Annex I product-system high-risk date

Check current formally adopted law before relying on the table.

Role glossary

Provider

Entity developing or having developed and placing a system/model on the market or putting it into service under its name or trademark.

Deployer

Entity using an AI system under its authority in a professional context.

Importer

Union-established entity placing a third-country provider's high-risk system on the Union market.

Distributor

Supply-chain entity making a system available without being provider or importer.

Product manufacturer

Manufacturer placing a product containing or using the high-risk system on the market under its name or trademark.

Authorised representative

Union-established person mandated to perform specified provider tasks. Article 22 and Article 54 routes must be distinguished.

GPAI model provider

Provider placing a general-purpose AI model on the market.

Downstream provider

Provider integrating a model into an AI system.

Assessment glossary

Applicable

The legal trigger and role are present.

Atomic item

Smallest independently answered legal check in Gamut.

Assessment basis

The material legal-routing facts to which confirmation applies. A material change invalidates prior confirmation.

Current-law conclusion

Position using applicable provisions in force as of the stated snapshot.

Future readiness

Preparation for applicable obligations not yet used as current law.

Hard stop

Condition preventing confirmation regardless of other scores; Article 5 is the principal example.

N/A

Approved conclusion that a specific legal trigger does not apply. It is not "not yet done."

Assured

Depth 3: current implementation, accepted evidence, effective test and no unresolved adverse finding.

Confirmation

Accountable human sign-off of the current validated assessment basis. It is not certification, deployment approval or risk acceptance by itself.

FRIA

Fundamental Rights Impact Assessment under Article 27 for specified deployers and uses.

GPAI

General-purpose AI model. Model-provider duties and downstream system duties are distinct.

Profiling override

Article 6 logic preventing reliance on the Annex III exception where profiling of natural persons is performed.

Substantial modification

Change that can trigger provider-role transfer and renewed compliance work under Article 25.

Frequently asked questions

Are there six EU AI Act risk classes in Gamut?

No. Earlier wording used six presentation groupings. The defensible method uses eleven orthogonal routes because scope, prohibited practices, high-risk duties, transparency and GPAI can coexist.

Does "not high-risk" mean out of scope?

No. Scope, Article 5, Article 4, Article 50 and other duties may still apply.

Does using a GPAI model make our system a GPAI model?

Not necessarily. Determine whether the organisation is a GPAI model provider, downstream system provider, deployer or API consumer.

Does human review mean an Annex III system is not high-risk?

No. Analyse the intended use, material influence, Article 6 exception and profiling. A nominal human-in-the-loop is not a blanket exclusion.

Does every high-risk deployer need a FRIA?

No. Article 27 has its own deployer and use triggers.

Can a control be Compliant but not Assured?

Yes. Compliant is the legal answer; depth describes verification. The claim may be Supported or Asserted until evidence and testing are complete.

Can N/A be used for future obligations?

No. Use future-obligation status. N/A means the legal trigger does not apply.

Can an average offset one non-compliant obligation?

No. Category percentages support progress tracking; they do not erase an atomic legal gap.

Can AI Assist confirm compliance?

No. It provides system-scoped advisory analysis. Human assessors accept evidence, approve N/A, resolve legal issues and confirm.

What happens when the selected system changes?

The assessment and AI context switch to that system. Prior AI analysis is cleared to prevent cross-system reuse.

Do AI features bypass plan entitlements?

No. Identity, organisation, workspace, role, plan and object-scope access controls remain required.

Does confirmation mean EU certification?

No. It means the assessment passed Gamut's confirmation gates under its stated scope and snapshot.

One-minute explanation

> Gamut assesses the EU AI Act per named AI system. It first establishes EU scope, intended purpose and every operator role. Gamut then calculates eleven legal routes, including an eight-part Article 5 hard-stop screen, Annex I and all eight Annex III points, operator duties, FRIA, Article 50 and GPAI. Broad provisions are decomposed into 72 atomic checks. Each applicable item receives a legal answer and a separate assurance depth. Confirmation requires accepted evidence, passing testing and no unresolved adverse finding for every applicable in-force item, while future and proposed changes are reported separately. AI can assist analysis but cannot change routing or approve the conclusion.

Official references

62. Reporting and governance

How EU AI Act system and workspace reports present legal status, route, assurance, gaps, evidence and accountable decisions without overstating compliance.

An EU AI Act report should answer:

> What is the current legal position of this named system, which obligations and roles were assessed, what evidence and testing support the result, and what action remains?

Selected-system report

The report should include:

- System name, identifier, version and intended purpose.
- Entity and operator roles.
- EU nexus and exclusions.
- Methodology and legal snapshot.
- Current and future application dates.
- Authoritative routes and triggering facts.
- Article 5 status.
- High-risk classification basis.
- Atomic answer, depth and legal basis.
- Approved N/A decisions.
- Accepted/rejected evidence.
- Passing/failed tests.
- Open findings.
- Current-law conclusion.
- Future-readiness conclusion.
- Confidence, residual risk, owner, reviewer and review date.
- Reassessment triggers and confirmation basis.

If the system-specific assessment record is absent, the report should show unassessed. It must not fall back to another system or a browser-active workspace scratch record.

Workspace report

A portfolio roll-up may show:

- Systems in and out of scope.
- Article 5 potential or confirmed stops.
- High-risk systems by Annex I/III basis.
- Applicable operator roles.
- FRIA coverage.
- Article 50 trigger coverage.
- GPAI provider and supply-chain exposure.
- Current gaps and future-readiness gaps.
- Systems awaiting confirmation or legal review.

Aggregation is conservative. A workspace average never proves that each system is compliant.

Current-law reporting

Clearly state:

- "As of" date.
- Binding provisions considered.
- Role and factual assumptions.
- In-force applicable items.
- Assurance exceptions.
- Scope limitations.

Future duties and proposed amendments belong in separate sections.

Article 5 reporting

Do not bury prohibited-practice status in a percentage. Report:

- Atomic check.
- Clear, potential or prohibited result.
- Facts and evidence.
- Legal review status.
- Deployment restriction or stop.
- Decision owner and next action.

High-risk classification reporting

Show:

- Annex I two-part analysis.
- Annex III point.
- Article 6 exception analysis.
- Profiling result.
- Final classification.
- Provider documentation/registration where an exception is used.
- Legal uncertainty.

Atomic assurance reporting

For each item, show:

- Legal answer.
- Assurance depth.
- Accepted evidence.
- Test position.
- Open adverse finding.
- Current/future status.

Do not report "compliant" without making the supporting depth and limitations visible.

Governance decisions

The report can inform:

- Deployment approval or restriction.
- Remediation priority.
- Legal review.
- Conformity and registration work.
- Supplier or representative action.
- FRIA completion.
- Notice and transparency changes.
- Incident escalation readiness.
- Evidence request and retest.
- Risk acceptance.

The assessment confirmation is not automatically a deployment approval or legal-risk acceptance. Record those decisions separately under the organisation's authority model.

Right audience, right detail

Assessor and counsel

Need atomic facts, legal basis, workpapers, evidence and tests.

System and control owners

Need implementation gaps, owners, due dates and retest criteria.

Executive or board

Need material systems, Article 5 stops, high-risk exposure, assurance exceptions, legal deadlines, residual risk and decisions requiring authority or funding.

Regulator or customer

Need a controlled pack with scope, roles, methodology, current-law status, relevant evidence and limitations. Apply least privilege and legal review before disclosure.

Cross-framework evidence reuse

EU AI Act items map to GTSAF, MAESTRO and ATF where evidence may be relevant.

Reuse means:

- One artefact may support several requirements.
- A failed technical test may inform several frameworks.
- Traceability reduces duplicate work.

It does not mean:

> Passing a GTSAF or MAESTRO item automatically proves EU AI Act compliance.

The EU legal trigger, role and assurance decision remain separate.

Report wording

Prefer:

> "Under the stated system boundary, operator roles and legal snapshot, all applicable in-force atomic items in this confirmed assessment meet Gamut's depth-3 assurance gate, with the limitations and reassessment triggers stated below."

Avoid:

- "EU approved."
- "Certified compliant."
- "No legal risk."
- "The whole workspace is compliant" based on an average.
- "N/A" without the decision basis.
- "AI confirmed compliance."

Export quality checklist

63. Routing and classification

How Gamut derives authoritative EU AI Act routes for prohibited practices, high-risk systems, FRIA, transparency, operators and GPAI.

Routing selects the legal questions that must be answered for the named system. It is conservative, system-specific and recalculated from the current approved facts.

The cross-framework route first determines whether an EU territorial and role screen is required from the linked System Record and Intake. Unknown territorial facts remain **screening required**. The EU AI Act module then performs the more detailed legal routes below. See Routing and applicability.

Routing is orthogonal

Do not describe the result as a choice between "minimal," "limited," "high-risk" and "GPAI." Different routes can apply together.

```
Scope and role
├─ Article 5 prohibited screen
├─ Article 6 / Annex I / Annex III classification
│  ├─ Articles 8-15 high-risk requirements
│  ├─ operator and conformity duties
│  ├─ FRIA where triggered
│  └─ monitoring and incidents
├─ Article 50 behaviour-based transparency
└─ Chapter V GPAI role and systemic-risk duties
```

Structured routing inputs

The intake captures facts such as:

- EU nexus and exclusion basis.
- AI-system-definition determination.
- Provider, deployer, importer, distributor, manufacturer and representative roles.
- Intended purpose and actual use.
- Article 5 screening for each prohibited practice.
- Annex I product and third-party conformity-assessment status.
- Each of the eight Annex III areas.
- Article 6(3) exception factors and profiling.
- Public-body and public-service status.
- Article 50 behaviour triggers.
- GPAI role, market-placement date and systemic-risk status.
- Current system lifecycle and deployment state.

Free text can explain these fields but cannot override the authoritative route.

The central routing facts also provide decision impact, action capability, organisation roles, biometric, emotion-recognition, synthetic-content, deepfake and jurisdiction signals. These support screening but do not replace the module's atomic legal determinations.

Route states

State	Meaning
`not_assessed`	The route has not yet been determined
`applicable`	Facts trigger the route
`not_applicable_with_evidence`	A governed non-applicability decision is complete
`incomplete`	Required facts or decisions are missing
`implemented`	Routed obligations have been addressed, subject to assurance
`prohibited_stop`	A prohibited-practice hard stop applies
`future_obligation`	Applicable provision is not yet in force for the current-law conclusion

State	Meaning
`proposed_change_only`	Tracked proposal or political agreement is not binding law

Article 5 hard-stop screen

The parent Article 5 question expands into eight atomic checks:

1. Subliminal, manipulative or deceptive techniques.
2. Exploitation of age, disability or social/economic vulnerability.
3. Social scoring.
4. Criminal-offence risk assessment based solely on profiling or personality traits.
5. Untargeted scraping to create or expand facial-recognition databases.
6. Emotion inference in workplace or education, subject to the narrow exception.
7. Sensitive biometric categorisation.
8. Real-time remote biometric identification in public spaces for law enforcement.

A confirmed prohibited practice stops confirmation. A potential prohibited practice also blocks confirmation until resolved. Strong scores elsewhere cannot offset this gate.

Article 6(1) Annex I route

Two conditions must be analysed together:

1. The AI system is a safety component of a product, or is itself a product, covered by legislation listed in Annex I.
2. The product is required to undergo third-party conformity assessment under that legislation.

Record the product, applicable legal act, safety function, conformity route and responsible product manufacturer. Do not route solely because the system is safety-related.

Article 6(2) Annex III route

Each point is atomic:

Point	Area
1	Biometrics
2	Critical infrastructure
3	Education and vocational training
4	Employment, worker management and access to self-employment
5	Essential private services and essential public services and benefits
6	Law enforcement
7	Migration, asylum and border control
8	Administration of justice and democratic processes

Assess the precise use case, not just the industry. A bank chatbot is not automatically Annex III; a system evaluating creditworthiness may be.

Article 6(3) exception and profiling override

For an Annex III use case, document whether the system poses a significant risk of harm to health, safety or fundamental rights. The statutory exception is narrow and requires a reasoned, documented analysis.

Consider whether the system:

- Performs a narrow procedural task.
- Improves the result of a previously completed human activity.
- Detects patterns or deviations without replacing or influencing a prior human assessment.
- Performs a preparatory task.

The exception does not apply where the system performs profiling of natural persons. Gamut records profiling separately and does not allow a generic "human in the loop" statement to establish the exception.

Where a provider concludes an Annex III system is not high-risk, preserve the Article 6(4) documentation and applicable registration evidence.

High-risk downstream routes

When high-risk classification is active, Gamut can activate:

- Articles 8-15 system requirements.
- Provider quality-management and record-retention duties.
- Conformity assessment, declaration, CE marking and registration.
- Deployer duties.
- Article 21 authority cooperation.
- Articles 22-25 representative, importer, distributor and value-chain duties.
- Article 27 FRIA where its separate trigger applies.
- Articles 72-73 monitoring and serious incidents.
- Article 86 explanation rights where triggered.

The role matrix determines which of these belong to the organisation.

FRIA route

High-risk classification does not automatically mean every deployer must conduct a FRIA. Record the Article 27 trigger, including whether the deployer is:

- A body governed by public law.
- A private entity providing public services.
- A deployer of specified Annex III systems for which the Act requires a FRIA.

Record linkage to any data-protection impact assessment without treating the DPIA as an automatic substitute.

Article 50 transparency route

Article 50 is behaviour-triggered and may apply whether or not the system is high-risk. Gamut checks:

- Direct interaction with a natural person.
- Machine-readable marking of synthetic content.
- Emotion-recognition notice.
- Biometric-categorisation notice.
- Deepfake disclosure.
- Public-interest text disclosure.
- Timing at first interaction or exposure.
- Clarity, distinguishability and accessibility.
- Any relied-on statutory exception.

Each triggered duty receives its own conclusion.

GPAI routes

Distinguish:

- Direct GPAI model provider.
- GPAI model provider with systemic risk.
- Third-country provider requiring an authorised representative.

- Downstream provider integrating a GPAI model.
- Deployer or API consumer requiring supplier evidence.

For systemic risk, assess both the compute presumption and Commission designation. Do not treat a single compute threshold as the only route.

The Commission states that GPAI provider duties have applied since 2 August 2025. Its GPAI guidelines and voluntary GPAI Code of Practice can support interpretation and demonstration but do not replace the Regulation.

Contradictions and missing fields

Examples that require resolution:

- "Out of scope" with an EU market or output nexus.
- No provider or deployer role despite professional operation.
- Annex III use selected but classification left N/A.
- Article 6 exception claimed while profiling is present.
- GPAI systemic-risk duties selected while the organisation records no GPAI provider role.
- FRIA marked N/A while trigger facts indicate applicability.
- Article 50 notice marked N/A while a behaviour trigger is confirmed.

Gamut blocks confirmation rather than guessing through material uncertainty.

Assessment basis and reassessment

Confirmation is tied to the material routing facts. Changes to purpose, geography, role, model, Annex III use, Article 5 facts, transparency behaviour, GPAI status or other material fields invalidate the prior confirmation.

Assessor quality checks

64. Scoring, assurance and conclusions

How EU AI Act legal answers, assurance depth, coverage, category scores, N/A governance and confirmation gates work in Gamut.

Gamut separates three questions:

1. **Applicability** - does the legal requirement apply?
2. **Compliance answer** - is it met?
3. **Assurance depth** - how strongly is that answer demonstrated?

Combining these into one percentage would hide important differences.

Legal answers

Answer	Meaning
Compliant	The applicable obligation is implemented for the assessed system and role
Non-compliant	The obligation is absent, incomplete, ineffective or contradicted
N/A	The trigger does not apply and a governed non-applicability decision is complete
Unassessed	No conclusion has yet been made

"Compliant" is not automatically "Assured."

Assurance depth

Depth	Label	Required interpretation
3	Assured	Implemented and current, accepted evidence, effective testing, no unresolved adverse finding
2	Supported/Partial	Relevant implementation evidence exists, but testing, coverage or closure is incomplete
1	Asserted	Claimed but weakly evidenced or untested
0	Gap	Absent, ineffective, contradicted or materially misleading

Example

A deployer policy says human oversight is required, but no authority matrix, training record or operation test exists:

- Legal answer may be **Compliant** only if the assessor can support implementation.
- Assurance cannot exceed **Asserted** merely because a policy exists.
- If practice does not follow the policy, the answer is **Non-compliant / Gap**.

Category compliance

Category compliance is coverage-inclusive:

```
compliant applicable atomic items
──────────────────────────────── × 100
all applicable atomic items
```

Unassessed applicable items remain in the denominator. This prevents a high result from being created by answering only favourable items.

Approved N/A items are excluded from the applicable population. Unsupported N/A items are not.

Category assurance

Assurance reflects the depth of the applicable atomic items. Always display it alongside:

- Assessment coverage.
- Evidence acceptance.
- Passing and failed tests.
- Open findings.
- Article 5 status.
- Current versus future legal status.

Do not interpret a category average as permission to ignore one failed statutory obligation.

N/A governance

A valid N/A requires:

- Trigger-specific factual rationale of at least 80 characters.
- Exact legal basis.
- Evidence references.
- Decision owner.
- Independent approver.
- Approval date.
- Review or reassessment trigger.

Good:

> Article 50(4) deepfake disclosure is not applicable because the deployed system only produces internal text summaries, cannot generate or manipulate image, audio or video content, and this is enforced by the approved model endpoint and content-type controls. Reassess if multimedia generation is enabled.

Weak:

> Not relevant to our use case.

Hard stops and assurance gates

Article 5

A confirmed prohibited practice, or unresolved potential prohibited practice, blocks confirmation. It cannot be averaged away.

Confirmation assurance

For confirmation, an applicable in-force atomic item must be:

- Compliant at depth 3; or
- Supported by an approved N/A decision.

Depth 3 additionally requires:

- Accepted/reviewed evidence linked to the item and system.
- Passing scoped testing.
- No failed test.
- No unresolved adverse finding.

If any condition is missing, Gamut returns an incomplete-assurance outcome rather than an optimistic confirmation.

Conclusion vocabulary

AI Assist and human reporting may use these evidence-aware positions:

Position	Meaning
`assured_compliant`	Applicable obligation is demonstrated at the full assurance gate
`supported_partial`	Material support exists but assurance is incomplete
`asserted_unverified`	Claim exists without sufficient verification
`gap`	Requirement is absent, ineffective or contradicted
`not_applicable_pending_approval`	N/A appears plausible but governance approval is incomplete
`future_readiness`	Work relates to a provision not yet used as current binding law
`insufficient_information`	Facts do not support a responsible determination
`potential_prohibited_practice`	Article 5 concern requires immediate human/legal resolution

These are not certification labels.

Current-law versus future-readiness conclusion

The conclusion should have two explicit parts:

Current-law conclusion

Uses provisions in force as of the snapshot and states:

- Scope and roles.
- Applicable routes.
- Assured compliant items.
- Current gaps.
- Limitations and residual risk.
- Article 5 status.

Future-readiness conclusion

States:

- Applicable future obligations.

- Expected application dates.
- Implementation and evidence gaps.
- Planned completion.
- Any dependence on proposed but unadopted changes.

Never call a future item a current breach, or a proposed date binding.

Confidence

Confidence	Typical basis
Low	Material scope, role, classification, evidence or legal uncertainty remains
Medium	Main routes and implementation are supported, but some operating periods or edge cases remain incomplete
High	Current facts, roles, evidence, testing and adverse signals have been independently challenged

Confidence does not change the legal route. Low confidence normally calls for more conservative governance.

Residual risk

Record:

- Open non-compliance.
- Limitations in evidence or testing.
- Reliance on suppliers or representatives.
- Affected persons and consequence.
- Interim safeguards.
- Deployment restrictions.
- Risk owner and acceptance authority.
- Expiry conditions.

Risk acceptance does not turn a legal gap into compliance.

Safe explanations

Prefer:

> "As of 16 July 2026, the named system has no identified Article 5 hard stop. The applicable in-force items listed in the report meet Gamut's depth-3 assurance gate, subject to the stated scope, legal assumptions, evidence period and reassessment triggers."

Avoid:

- "EU certified."
- "100% legally compliant forever."
- "Low risk, so the Act does not apply."
- "The vendor is compliant, therefore we are compliant."
- "All future obligations are already legally in force."

Why a percentage is not the conclusion

A percentage cannot by itself reveal:

- A prohibited practice.
- One failed high-risk obligation.
- A false N/A.

- A failed test.
- An unresolved serious incident.
- A future obligation included as current.
- A provider duty reported against the wrong entity.

Use scores for navigation and progress; use the complete record for the conclusion.

65. Worked example

Complete EU AI Act example covering system scope, Annex III employment classification, transparency, GPAI supply chain, evidence, testing and confirmation.

This example shows how the routes and assurance model work together.

Scenario

Recruitment Screening Assistant:

- Ranks job applicants for recruiter review.
- Extracts information from CVs.
- Produces candidate summaries and fit recommendations.
- Uses a third-party GPAI model through an API.
- Interacts directly with candidates through a portal.
- Is deployed by an EU employer.
- Does not make the final hiring decision automatically.

Scope and roles

The organisation records:

- **EU nexus:** in scope.
- **AI system:** yes.
- **Organisation:** deployer.
- **Supplier:** provider of the recruitment system.
- **Foundation-model supplier:** GPAI model provider.
- **Actual use:** applicant filtering and ranking.

The absence of a fully automated final decision does not by itself prevent high-risk classification.

Article 5

All eight checks are reviewed.

Special attention is given to:

- Manipulative interface design.
- Vulnerability exploitation.
- Emotion inference in the workplace/employment context.
- Sensitive biometric categorisation.

No prohibited practice is identified. The system does not infer emotion or sensitive traits, and technical tests confirm those features are not enabled through alternate endpoints.

Result:

> Article 5 clear, subject to change control and reassessment.

High-risk classification

Annex I

Not applicable: the system is not a covered Annex I product or safety component requiring third-party conformity assessment.

Annex III

`EU-AI-04.4` applies because the system is used for recruitment and candidate selection.

Article 6 exception

The provider argues the system only performs a preparatory task. The assessor rejects that argument because ranking materially influences which applicants receive human attention. Profiling is also present.

Conclusion:

> Annex III high-risk route applies.

Routed duties

The deployer receives:

- High-risk system information relevant to deployed use.
- Twelve Article 26 atomic checks.
- Workplace/candidate notice obligations.
- Monitoring and escalation.
- Article 86 explanation analysis for significantly affected individual decisions.
- Article 50 direct-interaction notice.
- GPAI supplier-evidence route.

The provider retains provider, quality-management, conformity and registration duties. The deployer requests evidence rather than claiming to own those processes.

Example atomic item: human oversight

Question:

> Are competent natural persons assigned and able to understand, override or stop the system before a consequential effect?

Initial answer:

> Compliant, depth 1 - Asserted

Existing evidence:

- Recruitment procedure mentioning recruiter review.
- Supplier user guide.

Missing:

- Named authority matrix.
- Training and comprehension record.
- Evidence that rank order can be challenged.
- Test of override and recovery.

AI Assist correctly reports `asserted_unverified`, not assured compliance.

Test

The assessor runs a bounded test using synthetic applicants:

1. Create a candidate qualified for the role but with an uncommon career history.
2. Confirm the system ranks the candidate below threshold.

3 Present the recommendation and supporting information to a trained recruiter.

4 Verify the recruiter can see the AI's role and limitations.

5 Challenge the recommendation.

6 Override the rank before any rejection communication.

7 Confirm the override and rationale are logged.

8 Confirm the candidate proceeds to human review.

Initial result:

> Failed - recruiters can manually advance the candidate, but the interface does not record the override or expose the factors driving the recommendation.

The compliant assertion is changed to **Non-compliant / Gap** and a finding is raised.

Remediation

The organisation and provider implement:

- Clear AI recommendation and limitation display.
- Reason codes appropriate to the decision.
- Recruiter training.
- Documented override authority.
- Mandatory rationale and immutable log.
- Monitoring for override rates and group disparities.
- Candidate notice and explanation-request process.

Retest and evidence

Accepted evidence:

- Updated workflow and authority matrix.
- Training completion and comprehension results.
- Interface release record.
- Audit-log sample.
- Bias and performance monitoring design.
- Passed retest.

No adverse finding remains open.

Final item:

> Compliant, depth 3 - Assured

Article 50

The candidate portal states at first interaction that the user is interacting with an AI-enabled service. The notice is tested for:

- Timing.
- Clear language.
- Accessibility.
- Mobile and desktop presentation.

Synthetic-content marking and deepfake disclosure are N/A only after the system boundary and endpoint restrictions are evidenced and approved.

GPAI supply chain

The employer is an API consumer, not the GPAI model provider. It obtains:

- Model/service documentation.
- Capabilities and limitations.
- Version/change notices.
- Security and incident terms.
- Evidence supporting the downstream provider's integration.

It does not claim that the model provider's Code of Practice participation automatically proves the employment system compliant.

FRIA

The organisation separately analyses Article 27 applicability. It does not assume every high-risk deployer automatically requires a FRIA. Even where Article 27 does not apply, broader fundamental- rights and data-protection impact work may still be required by other law or governance.

Human conclusion

> As of the recorded legal snapshot, Recruitment Screening Assistant is in scope and routed as an Annex III point 4 high-risk AI system used by the organisation as deployer. No Article 5 prohibited practice has been identified in the assessed configuration. Applicable in-force atomic items are confirmed only where accepted evidence, passing tests and closed findings support depth 3. Future obligations and supplier dependencies are reported separately. Expansion to automated rejection, emotion analysis, biometric categorisation, a new model or a new country requires reassessment.

What the example demonstrates

66. AI Assist explainability

Public guide to GTSAF per-control and portfolio AI assistance, data considered, Privacy Mode, saved outputs, access boundaries and human accountability.

GTSAF AI Assist helps an assessor interpret the selected assessment. It is available for one control at a time and for a whole-system summary. It does not approve an answer, evidence, test, finding, N/A decision or final conclusion.

Scope

The result is bound to the selected AI system and current assessment basis. The panel identifies the system being analysed. When another system is selected, the previous result is not presented as the new system's current analysis.

Workspace or portfolio analysis is appropriate only when the screen explicitly shows that scope. It does not replace system-level conclusions.

Per-control information considered

In full-context mode, AI Assist may consider:

- Canonical control ID, title, objective and advisory.
- Domain, criticality and Gate status.
- Applicability label and route rationale.
- Current answer and depth.
- Implementation and customer-responsibility narrative.
- Assessor rationale and conclusion fields.
- N/A basis and approval state.
- Authorised evidence status and metadata.
- Control tests and results.

- Open findings.
- Selected system facts relevant to the control.
- ACRS and assurance-depth context.

Only records available within the user's authorised assessment scope are considered.

Portfolio or whole-system information

Whole-system analysis may consider:

- Coverage and answer distribution.
- Baseline, Triggered and Enhanced populations.
- Applicable, pending and out-of-scope counts.
- Depth-weighted workload and assurance intensity.
- Gate and High-criticality gaps.
- Evidence and testing depth.
- Open findings and unresolved N/A.
- Ownership and review gaps.
- Current assurance conclusion.

The summary should direct the reviewer back to individual controls for evidence and decision detail.

Expected output

Per-control analysis can identify:

- Whether information is sufficient.
- Verified facts and assumptions.
- Evidence-supported strengths.
- Missing evidence.
- Contradictions or adverse findings.
- Suggested implementation actions.
- A safe bounded test and pass criteria.
- Suggested answer, depth and assurance position for human review.
- Residual risk and reassessment triggers.

AI suggestions are not automatically applied.

Privacy Mode

The **Privacy Mode** checkbox beside AI Assist uses scores-only reduced context. It excludes direct identifiers and free text such as system names, implementation narrative, customer-responsibility notes, N/A rationale and assessor conclusions.

The model receives only the minimum structural state, which can include control identifiers, answer and depth, applicability, evidence/test/finding status, ACRS route and completeness signals.

Privacy Mode reduces disclosure but may produce less specific advice. It does not make the provider call anonymous or change assessment data, routing, scoring or access.

Saved results, re-run and minimising

- A successful result is saved for the selected system, analysis scope and privacy mode.
- The action changes from **Run AI Assist** to **Re-run AI Assist**.
- **Minimise** collapses the output without deleting or approving it.
- Re-run after material scope, control, evidence, test, finding or system change.

Provider and API-key use

AI Assist uses an authorised provider and the existing API-key configuration available to the user or organisation. If no permitted provider is configured, analysis cannot run.

An API key does not grant access. AI Assist remains subject to the user's plan, workspace, role, framework, model and selected-system permissions.

Evidence discipline

AI Assist must not describe a control as verified merely because:

- The answer is Yes.
- A narrative claims implementation.
- A generic policy exists.
- A document is linked but not accepted.
- The model recommends a control.

Verify every evidence statement against the authorised record. Failed tests and open adverse findings must remain visible.

Untrusted content

System descriptions, retrieved content, evidence and findings may contain instruction-like or malicious text. The model should treat that material as data, not authority.

Assessors should not include credentials, unnecessary personal data or unrestricted confidential content. Independently verify surprising instructions, citations and claims.

Human decisions

AI Assist cannot:

- Set the final answer or depth.
- Decide or approve applicability.
- Approve N/A.
- Accept evidence.
- Pass a test.
- Close a finding.
- Confirm the assessment.
- Accept residual risk.
- Authorise deployment or production testing.
- Change access or entitlements.
- Certify compliance.

Review checklist

67. Assessment workflow

End-to-end GTSAF assessor workflow from system intake and scope selection through evidence, testing, findings, conclusions and sign-off.

This page describes how to complete a defensible GTSAF assessment from preparation through final assurance reporting.

Before starting

Confirm that:

- The correct workspace is open.

- The AI system is registered.
- The intake record describes the current use case and architecture.
- The system owner and assessment owner are identified.
- The linked System Record and Intake contain complete, conflict-free routing facts.
- ACRS has been reviewed where its additional depth is material to the assessment.
- The intended assessment boundary is documented.
- The assessment is in edit mode and not locked.
- The assessor has access to the required evidence, testing and findings modules.

Step 1: define the assessment boundary

Record:

- The named AI system and system reference.
- Business purpose and intended use.
- Users and affected people.
- Deployment model.
- Models and providers.
- Data categories and sources.
- Interfaces, APIs and integrations.
- RAG, prompts, context stores and retrieval sources.
- Agentic tools, memory and delegated actions.
- Human decision points.
- Environments in scope.
- Jurisdictions and regulatory roles.
- Suppliers and shared-responsibility boundaries.
- Known exclusions.

Avoid boundaries such as "the AI platform" without naming the actual components and operating relationships.

Step 2: select the GTSAF scope

Choose:

- **All GTSAF controls**
- **AI system: authoritative intake route complete**

For a formal multi-system assessment, select the named system. ACRS confirmation is not required to determine factual applicability; where ACRS is unavailable, the screen identifies that no confirmed ACRS depth overlay is being used.

Review:

- Applicable-control count.
- Baseline, Triggered and Enhanced depth counts.
- Depth-weighted workload and assurance intensity.
- Factual scope drivers and their overlap warning.
- Applicability-pending controls.
- Out-of-scope controls.
- The reason shown for each control's applicability.

Confirm that:

Baseline + Triggered + Enhanced = Applicable controls

and:

Applicable controls + Applicability pending + Out of scope = 358

Do not interpret the largest applicable-control count as the highest-risk system. Applicable count measures breadth; workload and assurance intensity explain required rigour. If professional judgement identifies an additional relevant control, include and assess it without changing the underlying Intake merely to obtain a preferred count.

Step 3: work domain by domain

GTSAF is organised A to Q. A practical assessment sequence is:

1. **A-C**: governance, legal and intake.
2. **D-G**: data and model engineering.
3. **H-I**: prompt, retrieval, inference and runtime.
4. **J-L**: identity, agentic action and supply chain.
5. **M-N**: monitoring, oversight and impact.
6. **O-Q**: resilience, assurance and infrastructure.

This order establishes governance and system facts before detailed technical testing.

Step 4: read the full control advisory

For each control, review:

- Control statement.
- Objective.
- Risk addressed.
- Applicability explanation.
- Criticality.
- Gate status.
- Implementation guidance.
- Required evidence.
- Test procedure.
- Control owner.
- Evidence owner.
- Cross-framework mappings.
- Role-specific guidance.

Do not assess from the short title alone.

Step 5: answer each assessment question

Yes

Select Yes only when the requirement is met in the selected scope.

A Yes answer should be accompanied by:

- Clear implementation narrative.
- Named responsibility.
- Relevant evidence.
- Testing where operating effectiveness matters.

Yes is an assessor claim, not proof by itself.

No

Select No when the requirement is absent, incomplete or ineffective.

Then:

- Explain the precise deficiency.
- Identify the risk or consequence.
- Link or raise a finding.
- Record compensating controls.
- Assign remediation.
- Decide whether deployment restrictions or escalation are needed.

No creates a Gap. On a Gate control it creates a Gate fail.

N/A

Select N/A only where the requirement genuinely does not apply to the selected boundary.

Complete:

- Detailed rationale.
- Decision category.
- Decision owner.
- Independent approver.
- Approval date.
- Review date or reassessment trigger.
- Evidence references.
- Compensating controls or boundary safeguards.

Until the governance fields are complete, the N/A is not honoured.

Step 6: assign shared-responsibility ownership

Choose the party that actually operates or assures the requirement:

- MP
- OSP
- AP
- AIC
- CSP
- Shared
- ND

For Shared:

- Identify which party performs each activity.
- Identify which party remains accountable.
- Link contractual or architectural evidence.
- Explain how gaps and incidents are escalated across the boundary.

Resolve ND before sign-off.

Step 7: document implementation

The implementation narrative should state:

- What is implemented.

- Where it is implemented.
- Who operates it.
- Which systems or environments it covers.
- How often it operates.
- How exceptions are handled.
- How changes are approved.
- Which evidence demonstrates operation.
- What the customer or another provider must do.
- Known limitations.

Avoid statements such as:

- "Industry best practice is followed."
- "The vendor handles this."
- "Security is enabled."
- "Compliant."

Those statements are not testable without further detail.

Step 8: collect and link evidence

Use the control's Required evidence advisory as the starting checklist.

For each artefact:

- Link it to the correct control.
- Link it to the correct system or intake.
- Record owner and date.
- Review its relevance, sufficiency and currency.
- Record the review status.
- Record evidence quality.
- Reject generic or stale material.

Narrative does not become verified evidence through keyword matching.

Step 9: perform control testing

Use the control-specific test procedure to design a bounded test.

Record:

- Test objective.
- Test steps.
- Population and sample.
- Expected result.
- Pass criteria.
- Safety limits.
- Test date.
- Tester.
- Result.
- Exceptions.
- Retest date.

Do not conduct destructive production tests without explicit authorisation.

Step 10: raise findings

Raise a finding when:

- A question is No.
- A Gate fails.
- Evidence is rejected or materially insufficient.
- A test fails or produces an exception.
- Ownership is unresolved.
- The implementation differs from policy or contract.
- The control cannot be demonstrated in operation.
- A Critical control lacks required test coverage.

A useful finding contains:

- Clear title.
- Affected control and system.
- Condition observed.
- Expected requirement.
- Evidence.
- Root cause.
- Risk consequence.
- Severity.
- Recommendation.
- Owner.
- Target date.
- Status.

Step 11: record the assessor conclusion

Complete the structured conclusion:

- Design effectiveness.
- Implementation effectiveness.
- Operating effectiveness.
- Evidence conclusion.
- Residual risk.
- Assessor conclusion.
- Monitoring cadence.
- Reassessment triggers.
- Reviewer.
- Review date.

The conclusion should identify:

- What was reviewed.
- What was tested.
- What could not be verified.

- Open exceptions.
- Compensating controls.
- The basis for residual risk.
- The conditions under which the conclusion expires.

Step 12: review the derived result

Review:

- Conformance.
- Verified evidence.
- Narrative sufficiency.
- Ownership clarity.
- Audit readiness.
- Assurance percentage.
- Assurance score.
- Gate failures.
- Critical-control evidence and tests.
- Open findings.

Do not use the headline score without reading the underlying gaps and caps.

Step 13: use AI Assist

AI Assist can:

- Summarise the current validated scope.
- Identify evidence and testing gaps.
- Suggest bounded test procedures.
- Draft effectiveness and residual-risk reasoning.
- Identify monitoring and reassessment needs.

The assessor must:

- Confirm the displayed system scope.
- Check every claimed verified fact.
- Reject invented or unsupported conclusions.
- Decide whether to apply suggestions.
- Retain human accountability.

Step 14: perform quality review

Before sign-off, ask:

- Is the correct system selected?
- Is the authoritative Intake route complete and current?
- Has ACRS been reviewed where it changes required assurance depth?
- Do the scope and depth totals reconcile?
- Can the factual scope drivers be explained?
- Are all Mandatory, Triggered and Enhanced controls addressed?
- Are Critical and Gate controls prioritised?

- Are N/A decisions independently approved?
- Are Yes answers supported by linked evidence?
- Are Critical controls tested?
- Are failed tests and rejected evidence reflected in the conclusion?
- Are all findings linked and owned?
- Does residual risk reflect open exceptions?
- Are monitoring and reassessment triggers specific?
- Does the report use the correct system bucket?

Step 15: lock and report

Once reviewed:

- Lock the assessment according to your governance process.
- Generate the appropriate system or workspace report.
- Obtain formal review or approval.
- Record risk acceptance separately where required.
- Schedule the next review.

Locking prevents assessment changes and AI suggestions from being applied without reopening the governed workflow.

When reassessment is required

Reassess after:

- A new model or major model version.
- New training data or a changed retrieval corpus.
- A supplier or hosting change.
- New tools, plugins, memory or autonomy.
- Expanded privileges or system access.
- New personal or sensitive data.
- A move into a new jurisdiction.
- A serious incident, complaint or control failure.
- A material change in intended use.
- A failed control test.
- A regulatory or contractual change.
- Expiry of a N/A decision or evidence item.

Minimum sign-off record

A defensible sign-off should include:

68. Concepts, labels and terminology

Explain every GTSAF badge and status, including Critical, High, Gate, Mandatory, Triggered, Enhanced, Gap, Supported, Assured, Partial and N/A.

GTSAF deliberately shows several labels on one control because each label describes a different assurance dimension. Reading the badges from left to right gives a compact explanation of the control:

1. **How important is it?** Criticality.
2. **Can it block assurance?** Gate status.

3. **Why does it apply?** Applicability.
4. **What does the assessment demonstrate?** Assessment result.

A complete badge example

Suppose a control displays:

High · Gate · Enhanced · Gap

This should be read as:

- **High**: failure of the control could have a serious consequence.
- **Gate**: failure can block a positive assurance conclusion.
- **Enhanced**: the selected system requires deeper evidence and stronger testing for this control.
- **Gap**: at least one applicable assurance question is currently answered No.

The words are not one combined rating. They are independent labels.

Criticality labels

Criticality expresses the potential consequence of control failure. It influences prioritisation and the weighting of domain and overall assurance.

Label	Weight	Meaning	Assessment consequence
Critical	2.0	Failure could create severe safety, legal, security, rights, operational or systemic consequences.	Weak verified evidence caps assurance below a defensible level. A passing operating-effectiveness test is required before the control can reach Assured.
High	1.5	Failure could create a serious material consequence.	Weak evidence constrains the control's assurance result.
Medium	1.0	Material control expected in ordinary AI governance and security operation.	Normal weighting.
Low	0.75	Lower relative consequence within the GTSAF control population.	Lower roll-up weight, but still requires assessment when in scope.

The current library contains **70 Critical**, **241 High**, **43 Medium** and **4 Low** controls.

Criticality is not the same as the current assessment result. A Critical control can be Assured, and a Low control can still contain a Gap.

Gate labels

Gate

Gate is permanent control metadata. It means the control is important enough that a negative answer can prevent the system from receiving an unconstrained assurance conclusion.

Examples include controls governing approval authority, prohibited deployment conditions, identity boundaries, human intervention, evidence integrity or recovery capability.

Gate fail

Gate fail is an assessment result. It occurs when:

- The control is marked as a Gate control; and
- At least one applicable question is answered **No**.

A Gate fail:

- Overrides the ordinary status label.
- Caps calculated assurance at 25%.
- Produces a proxy assurance score of 1.

- Requires remediation, compensating controls, or an explicit decision not to proceed.
- Must not be hidden by strong performance elsewhere in the domain.

There are **71 Gate controls** in the current GTSAF library.

Applicability labels

Applicability answers: how and why does this control apply to this AI system?

Baseline or Mandatory baseline

The control sits inside the baseline expected for the selected system.

Baseline, Triggered and Enhanced are mutually exclusive assurance-depth labels. Together they make up the applicable-control population.

Mandatory does not mean every organisation implements the control identically. The architecture, shared-responsibility model and operating context determine how the required outcome is achieved.

Triggered

The system profile activates the control because a relevant feature or exposure is present.

Common triggers include:

- Personal or sensitive data.
- Training or fine-tuning.
- Retrieval-augmented generation.
- Public-facing interfaces.
- API serving.
- Agentic behaviour, tools or memory.
- Automated decisions.
- Material human impact.
- Third-party models or hosted services.
- Multi-tenant operation.
- Regulatory or jurisdictional exposure.
- Cultural, heritage, local-language or community-sensitive content.

Triggered means the system's characteristics activate the control and require more than baseline attention. It does not mean the control has failed.

Enhanced

Enhanced means the control is in scope and needs **deeper assurance** because the system's risk tier or exposure is elevated.

Enhanced assessment normally expects:

- More complete and current evidence.
- Stronger sampling.
- Independent or second-line challenge.
- A passing operating-effectiveness test.
- Tighter monitoring.
- Clearer exception governance.
- More frequent reassessment.
- Stronger evidence of supplier or shared-responsibility operation.

Enhanced is not an additional control duplicated on top of another count. It is the deepest of the three mutually exclusive assurance-depth states for an applicable control.

Not applicable or Out of scope

The complete system route does not contain a factual trigger for the conditional control.

The screen may use:

- **Not applicable**
- **Out of scope**

These describe routing, not an assessor-entered N/A answer. An assessor can still inspect an out-of-scope control and may choose to assess it where professional judgement requires.

Applicability pending

The routing facts are not complete or conflict-free enough to support inclusion or exclusion. Pending is not a pass or an exemption. The assessor must resolve the missing or conflicting facts.

Applicable controls, workload and intensity

- **Applicable controls** measure factual scope breadth.
- **Depth-weighted workload** applies transparent weights of 1 to Baseline, 2 to Triggered and 3 to Enhanced controls.
- **Assurance intensity** divides workload by the maximum possible depth for the applicable population.

These are scope-planning measures. They are separate from control effectiveness, conformance and the derived assurance score.

Assessment-result labels

Assessment result answers: what does the current assessment record demonstrate?

The labels follow this precedence:

1. N/A
2. Unassessed
3. Gate fail
4. Gap
5. Assured
6. Supported
7. Partial

Unassessed

There is no usable assessment response for the control.

Unassessed does not mean low risk, compliant or safe. Unassessed in-scope controls reduce coverage and therefore reduce conformance and roll-up assurance.

Gap

At least one applicable question is answered **No**.

A Gap indicates that the required outcome is not fully met. The assessor should:

- Identify which requirement is not met.
- Determine whether the weakness concerns design, implementation or operation.
- Link a finding where appropriate.
- Record the affected system and control.
- Set a remediation owner and target date.
- Decide whether interim compensating controls exist.
- Record residual risk and any deployment restriction.

If a control shows **Gap · Triggered**, the words describe two dimensions:

The system profile triggered the control, and the current assessment says at least one requirement is not met.

Assured

The control has:

- No applicable No answers; and
- Calculated assurance of at least 75%.

Assured indicates a strong assessment position supported by the current record. It does **not** mean certified, permanently effective or free of residual risk.

For Critical controls, Assured also requires sufficient evidence and at least one passing operating-effectiveness test because Critical controls are subject to additional caps.

Supported

At least one applicable question is answered **Yes**, there are no No answers, but assurance remains below 75%.

Typical reasons include:

- Incomplete coverage.
- Evidence that is uploaded but not reviewed or accepted.
- No passing control test.
- Weak or missing ownership.
- Insufficient implementation detail.
- Evidence that is stale, generic or not system-linked.

Supported is a positive direction of travel, but it is not yet strong enough for Assured.

Partial

Assessment activity exists, but the record contains neither affirmative Yes support nor a No that would make it a Gap, and it has not reached Assured.

This is commonly a provisional or incomplete state, for example:

- An N/A response has not completed the governed approval requirements.
- Information has been entered but the assessor has not made an affirmative determination.
- The available record is insufficient to classify the control as Supported.

Partial should normally prompt the assessor to complete the decision rather than leave it as the final conclusion.

N/A

The control's assessment rows are entirely excluded through approved N/A decisions.

N/A is only honoured when the governed N/A record is complete. Otherwise, the requirement remains in scope and is treated as not met.

Answer labels

Answer	Meaning
YES	The assessor states that the requirement is met for the selected scope. Evidence and testing are still required to support the claim.
NO	The requirement is not met. This creates a Gap and creates a Gate fail if the control is a Gate.
NA	The assessor proposes that the requirement does not apply. The exclusion is only honoured after the N/A governance fields are complete.

Governed N/A requirements

A valid assessor-entered N/A requires:

- A specific rationale of meaningful length.

- A decision category.
- A named decision owner.
- A separate independent approver.
- An approval date.
- A review date or reassessment trigger.
- Supporting evidence references where available.
- Compensating controls or boundary safeguards where relevant.

The decision owner and independent approver must not be the same person.

Shared-responsibility labels

Code	Meaning
MP	Model Provider
OSP	Orchestrated Service Provider
AP	Application Provider
AIC	AI Customer
CSP	Cloud Service Provider
Shared	Two or more named parties divide responsibility
ND	Not determined

ND is not a neutral long-term answer. It identifies an ownership gap.

AI Assist labels

Per-control AI Assist may use additional structured labels:

Applicability

- Applicable
- Potentially applicable
- Not applicable
- Insufficient information

Evidence assessment

- None
- Weak
- Partial
- Adequate
- Strong

Effectiveness

- Effective
- Partially effective
- Ineffective
- Not tested
- Insufficient evidence

Residual risk

- Low
- Moderate

- Elevated
- High
- Critical
- Unknown

These are AI recommendations for human review. They do not automatically modify routing, evidence, scores or the assessor conclusion.

A safe one-sentence explanation

When explaining a control badge to a stakeholder, use:

> Criticality tells us how important the control is; Gate tells us whether failure can block assurance; Mandatory, Triggered or Enhanced tells us why and how deeply it applies; and the final status tells us what the current evidence-led assessment demonstrates.

69. Domains and control library

Assessor field guide to all 17 GTSAF domains, including purpose, assessment focus, typical evidence, testing and common failure patterns.

GTSAF's 358 controls are grouped into 17 domains. This page explains how to approach each domain, what evidence normally matters and what failure patterns an assessor should look for.

The examples are not substitutes for the unique advisory attached to each individual control.

A - Governance, Strategy and Accountability

Purpose: Establish authority, policy, decision rights, ownership and oversight for the AI estate.

Assess:

- Board and executive authority.
- AI policy hierarchy.
- Governance forums.
- RACI and named accountability.
- Risk appetite.
- Exception and escalation routes.
- Management review.

Typical evidence:

- Board-approved AI charter.
- AI policy suite.
- Committee terms of reference and minutes.
- RACI.
- Decision logs.
- Risk-appetite statements.
- Management-review records.

Typical tests:

- Trace a material AI decision from initiation to approval.
- Confirm committees met at the stated cadence and quorum.
- Sample accountable owners against current organisation records.
- Verify a policy review was triggered after a material change.

Common failures:

- Governance exists only informally.
- Committees have no decision authority.
- Shared accountability with no single accountable owner.
- Policies are outdated or inaccessible.
- Risk appetite is expressed only as general principles.

B - Legal, Regulatory and Contractual Compliance

Purpose: Identify and operate the legal, regulatory, contractual and jurisdictional obligations that apply to the system.

Assess:

- Applicable-law inventory.
- Regulatory role and system classification.
- Prohibited or restricted use.
- Contractual obligations.
- Intellectual-property and licensing conditions.
- Data and cross-border obligations.
- Regulatory change management.

Typical evidence:

- Legal applicability assessment.
- Regulatory inventory.
- Contract clauses.
- Licence records.
- Data-processing agreements.
- Legal review and approval.
- Regulatory-change register.

Typical tests:

- Trace one obligation to an implemented control and evidence.
- Sample supplier and customer contracts for required AI clauses.
- Verify a regulatory change triggered reassessment.
- Confirm prohibited uses are technically or procedurally blocked.

Common failures:

- Generic legal memo with no system-specific conclusion.
- Unclear provider/deployer/customer role.
- Contract wording that does not match actual responsibility.
- Open-source or model licence obligations not monitored.
- No control for jurisdiction changes.

C - AI Use Case Intake, Approval and Risk Tiering

Purpose: Ensure every AI use case is registered, classified, approved and reassessed.

Assess:

- Intake completeness.
- System ownership.

- Risk tiering.
- ACRS.
- Approval gates.
- Exceptions.
- Material-change triggers.
- Retirement and status changes.

Typical evidence:

- AI system register.
- Intake form.
- ACRS record.
- Approval decision.
- Risk-tier rationale.
- Exception record.
- Change and reassessment history.

Typical tests:

- Select a live AI system and trace it to approved intake.
- Recalculate a sample ACRS route.
- Verify a system change caused reassessment.
- Identify unregistered or shadow-AI use.

Common failures:

- Intake performed after deployment.
- Draft ACRS treated as approved.
- Risk tier based on reputation rather than system facts.
- No link between intake and actual architecture.
- Changes do not trigger reassessment.

D - Data Governance, Lineage and Provenance

Purpose: Maintain trustworthy knowledge of where AI data comes from, how it changes and whether it is fit for use.

Assess:

- Data inventory.
- Source approval.
- Lineage.
- Provenance.
- Quality.
- Rights and permitted use.
- Retention.
- Dataset versioning.
- Retrieval corpus governance.

Typical evidence:

- Data catalogue.
- Lineage diagram.

- Source register.
- Dataset cards.
- Provenance metadata.
- Quality reports.
- Rights records.
- Retention schedule.

Typical tests:

- Trace a model output or retrieved document back to source.
- Sample dataset versions against approvals.
- Verify removed or expired sources no longer influence the system.
- Test lineage completeness after transformation.

Common failures:

- Unknown source origin.
- Approved dataset differs from deployed version.
- Retrieval documents bypass source governance.
- Quality checks cover format but not fitness for use.
- Retention rules do not reach embeddings or derived data.

E - Data Security and Privacy Engineering

Purpose: Protect sensitive data throughout AI collection, processing, storage, retrieval and output.

Assess:

- Data minimisation.
- Classification.
- Encryption.
- Segregation.
- Privacy engineering.
- Tokenisation and pseudonymisation.
- Tenant isolation.
- Data-loss prevention.
- Sensitive-output controls.

Typical evidence:

- Data-flow diagrams.
- Privacy assessment.
- Encryption configuration.
- Key-management records.
- DLP rules.
- Data classification.
- Tenant-isolation design.
- Redaction tests.

Typical tests:

- Attempt cross-tenant or unauthorised retrieval in a safe environment.

- Verify sensitive fields are redacted from prompts, logs and outputs.
- Inspect encryption and key rotation.
- Trace deletion across source, cache, vector and log stores.

Common failures:

- Sensitive data copied into prompts without need.
- Logs contain raw personal data.
- Deletion does not reach vector stores.
- Shared indexes permit cross-tenant exposure.
- Privacy controls rely entirely on provider statements.

F - Secure Data Acquisition and Annotation

Purpose: Protect data collection, labelling, annotation and preparation from manipulation, quality failure and unauthorised use.

Assess:

- Source acquisition.
- Consent and rights.
- Annotation instructions.
- Annotator access.
- Quality assurance.
- Poisoning resistance.
- Synthetic data.
- Dataset acceptance.

Typical evidence:

- Acquisition procedures.
- Annotation guidelines.
- Annotator training.
- Quality samples.
- Source approvals.
- Poisoning checks.
- Supplier records.
- Acceptance decisions.

Typical tests:

- Sample annotations for consistency and bias.
- Introduce bounded malformed or unauthorised records and confirm rejection.
- Verify annotator access and removal.
- Reproduce dataset acceptance criteria.

Common failures:

- Annotation quality measured only by volume.
- No segregation between raw and approved data.
- External annotators retain access.
- Poisoning and duplication checks absent.
- Synthetic data is undocumented.

G - Model Development, Validation and Robustness

Purpose: Ensure models are engineered, evaluated, approved and released against defined risk and performance criteria.

Assess:

- Development lifecycle.
- Version control.
- Validation independence.
- Performance and robustness.
- Bias and impact testing.
- Release criteria.
- Change approval.
- Reproducibility.

Typical evidence:

- Model cards.
- Experiment records.
- Validation reports.
- Benchmark results.
- Robustness tests.
- Bias analysis.
- Release approvals.
- Version history.

Typical tests:

- Reproduce a reported benchmark.
- Challenge performance under distribution shift.
- Verify the deployed version matches approval.
- Sample failed validation criteria and release decisions.

Common failures:

- Validation uses the training population only.
- Benchmark results cannot be reproduced.
- Model updates bypass approval.
- Business performance hides safety degradation.
- No defined minimum release criteria.

H - Prompt, Context and Retrieval Security

Purpose: Protect prompts, instructions, retrieved context and grounding from manipulation and untrusted influence.

Assess:

- System-prompt protection.
- Prompt injection.
- Context boundaries.
- Retrieval source approval.
- RAG poisoning.
- Grounding.

- Tool-output injection.
- Context leakage.

Typical evidence:

- Prompt architecture.
- Injection test results.
- Retrieval allowlists.
- Source-integrity controls.
- Context filtering.
- Grounding evaluation.
- Prompt version history.

Typical tests:

- Submit bounded direct and indirect injection patterns.
- Insert an unapproved retrieval document.
- Test whether retrieved instructions override policy.
- Verify sensitive context is not exposed across sessions.

Common failures:

- Retrieved text is treated as trusted instruction.
- System prompts contain secrets.
- No source-integrity check.
- Guardrails test only known phrases.
- Context persists beyond the intended session.

I - Inference, API and Runtime Security

Purpose: Protect the serving surface through authentication, authorisation, validation, session security and runtime enforcement.

Assess:

- API authentication.
- Authorisation.
- Rate limiting.
- Input and output validation.
- Session management.
- Abuse protection.
- Runtime policy.
- Error handling.

Typical evidence:

- API specifications.
- Gateway configuration.
- Authentication flows.
- Rate-limit rules.
- Session settings.
- Validation schemas.
- Runtime logs.

- Error-handling standards.

Typical tests:

- Call an endpoint without credentials.
- Attempt an unauthorised operation.
- Test rate limits.
- Submit malformed and adversarial inputs.
- Verify errors do not expose internals.

Common failures:

- Authentication without action-level authorisation.
- Unlimited model endpoints.
- Front-end validation with no server enforcement.
- Session state leaks between users.
- Provider errors expose sensitive details.

J - Identity, Access and NHI Security

Purpose: Control human and non-human identities, privileges, secrets and access lifecycle.

Assess:

- Identity inventory.
- Service accounts and agents.
- Least privilege.
- Segregation of duties.
- Secret storage.
- Token lifecycle.
- Access reviews.
- Revocation.

Typical evidence:

- Identity register.
- Role and permission exports.
- Access reviews.
- Secret-vault configuration.
- Token rotation.
- Joiner-mover-leaver records.
- Revocation logs.

Typical tests:

- Compare actual permissions with documented need.
- Attempt an unauthorised action.
- Revoke an identity and verify access stops.
- Sample dormant or orphaned non-human identities.

Common failures:

- Shared service accounts.
- Long-lived tokens.

- Agent permissions exceed workflow need.
- Revocation affects the UI but not backend sessions.
- Non-human identities omitted from access reviews.

K - Agentic AI and Autonomous Action Governance

Purpose: Govern tools, memory, delegation, approvals and autonomous action.

Assess:

- Autonomy boundaries.
- Tool permissions.
- Approval gates.
- Memory governance.
- Delegation.
- Multi-step plans.
- Kill switches.
- Human intervention.
- Action logging.

Typical evidence:

- Agent register.
- Tool catalogue.
- Access matrix.
- Approval-gate configuration.
- Memory policy.
- Workflow plans.
- Runtime logs.
- Kill-switch tests.

Typical tests:

- Attempt an action outside approved tool scope.
- Verify approval is required for a consequential action.
- Poison or manipulate memory in a safe test.
- Trigger the kill switch and measure containment.
- Verify an agent cannot self-escalate.

Common failures:

- Tool access inherited from a broad user account.
- Approval exists only in the user interface.
- Memory has no provenance or retention.
- Agent can create or approve its own plan.
- Kill switch does not revoke downstream authority.

L - Third-Party, Model and Software Supply Chain Assurance

Purpose: Govern external models, software, plugins, data providers and shared-responsibility dependencies.

Assess:

- Supplier due diligence.

- Contract controls.
- Model and component provenance.
- Vulnerability management.
- SBOM or inventory.
- Provider change.
- Exit and substitution.
- Assurance reports.

Typical evidence:

- Due-diligence assessment.
- Contracts and SLAs.
- Supplier assurance.
- Component inventory.
- SBOM.
- Vulnerability records.
- Provider-change notices.
- Exit plan.

Typical tests:

- Trace a deployed dependency to approved inventory.
- Verify a critical supplier issue enters risk governance.
- Test provider substitution or degraded operation.
- Sample contract obligations against actual service settings.

Common failures:

- Procurement approval treated as security assurance.
- Unknown transitive dependencies.
- Provider terms change without review.
- No exit path.
- Supplier reports are accepted without scope analysis.

M - Monitoring, Detection and AI Security Operations

Purpose: Detect harmful, anomalous or degrading AI behaviour and coordinate operational response.

Assess:

- Logging.
- Behavioural baselines.
- Detection rules.
- Alert handling.
- Model and data drift.
- Security operations.
- Incident classification.
- Metrics and escalation.

Typical evidence:

- Log inventory.

- Detection rules.
- Dashboards.
- Alert records.
- Drift reports.
- Incident tickets.
- Escalation records.
- Monitoring coverage.

Typical tests:

- Generate a safe detectable event.
- Verify alert delivery and triage.
- Trace an anomaly to investigation and closure.
- Confirm critical logs cannot be disabled without detection.

Common failures:

- Logs exist but nobody reviews them.
- Monitoring covers uptime but not AI behaviour.
- Alerts cannot be tied to a system or action.
- Drift thresholds have no response process.
- Security operations lack AI-specific playbooks.

N - Human Oversight, Transparency and Impact Management

Purpose: Ensure people can understand, challenge, oversee and intervene in consequential AI use.

Assess:

- Human decision authority.
- Meaningful review.
- Transparency notices.
- Explainability.
- Contestability and appeal.
- Fairness and impact.
- Accessibility.
- Cultural and community context.

Typical evidence:

- Oversight procedures.
- User notices.
- Decision explanations.
- Appeal records.
- Impact assessments.
- Fairness tests.
- Stakeholder consultation.
- Intervention logs.

Typical tests:

- Observe a human review and confirm genuine authority.

- Sample explanations against actual decision factors.
- Trace an appeal to resolution.
- Test whether automation bias is addressed in training and workflow.

Common failures:

- Human reviewer merely confirms the AI output.
- Notices are generic or hidden.
- No route to challenge a decision.
- Fairness testing excludes affected groups.
- Explanations do not reflect actual system logic.

O - Resilience, Continuity and Recovery

Purpose: Maintain safe service, containment and recovery when AI components fail or become untrustworthy.

Assess:

- Failover.
- Graceful degradation.
- Rollback.
- Backup and restoration.
- Provider failure.
- Crisis management.
- Recovery objectives.
- Rebuild integrity.

Typical evidence:

- Continuity plan.
- Recovery procedures.
- Backup records.
- Failover tests.
- Rollback records.
- Provider-substitution plan.
- Crisis exercise.

Typical tests:

- Fail a dependency in a controlled environment.
- Restore from backup.
- Roll back a model or configuration.
- Verify the system enters a lower-risk mode.
- Measure recovery time and data loss.

Common failures:

- Recovery covers infrastructure but not models or vector stores.
- Backups are never restored.
- Failover preserves service but removes controls.
- Provider outage creates uncontrolled manual workarounds.
- No evidence-preserving incident recovery.

P - Auditability, Evidence and Assurance

Purpose: Make the assessment reproducible, reviewable and independently challengeable.

Assess:

- Evidence governance.
- Control testing.
- Audit trails.
- Record integrity.
- Review independence.
- Findings.
- Assurance planning.
- Retention.

Typical evidence:

- Evidence register.
- Test plan.
- Workpapers.
- Audit log.
- Review sign-off.
- Findings register.
- Retention schedule.
- Assurance calendar.

Typical tests:

- Reproduce a control conclusion from its records.
- Verify audit-log immutability.
- Sample evidence expiry and review.
- Confirm reviewer independence.

Common failures:

- Conclusion cannot be traced to evidence.
- Testing is described but not recorded.
- Evidence is overwritten without history.
- Reviewer is also the control operator with no challenge.
- Findings close without validation.

Q - Infrastructure, Platform and Environment Security

Purpose: Protect the compute, cloud, network, storage and environment foundations supporting AI.

Assess:

- Cloud and platform configuration.
- Network segmentation.
- Compute isolation.
- Container and workload security.
- Secrets.
- Patch and vulnerability management.

- Environment separation.
- Capacity and resource controls.

Typical evidence:

- Cloud configuration.
- Network diagrams.
- Container policies.
- Vulnerability scans.
- Patch records.
- Environment inventory.
- Secrets configuration.
- Capacity controls.

Typical tests:

- Inspect configuration against secure baseline.
- Attempt prohibited network paths.
- Verify environment segregation.
- Sample critical vulnerabilities through remediation.
- Test resource exhaustion protections safely.

Common failures:

- AI workloads inherit generic cloud defaults.
- Development and production share credentials.
- Model endpoints bypass network controls.
- Unpatched libraries and images.
- GPU or accelerator access is not isolated.

Using the domain guide

For each domain:

1. Confirm it is in scope.
2. Review the domain reason.
3. Prioritise Critical and Gate controls.
4. Read each control's unique advisory.
5. Collect system-specific evidence.
6. Execute bounded tests.
7. Raise findings.
8. Record the assessor conclusion.

The domain guide helps you navigate. The individual control remains the authoritative assessment unit.

70. Evidence, testing and findings

How to collect, review, link and test GTSAF evidence, distinguish narrative from proof, and raise actionable findings.

GTSAF separates four things that are often incorrectly combined:

1. **Assessment answer**: the assessor's Yes, No or N/A determination.
2. **Implementation narrative**: the explanation of how the control is intended to work.

3. **Verified evidence**: reviewed artefacts or operating records supporting the claim.
4. **Control test**: a procedure used to determine whether the control works.

An assessment is defensible only when these elements agree.

Narrative is not evidence

A detailed implementation narrative is valuable because it explains:

- Control design.
- Operating process.
- Ownership.
- Frequency.
- Exceptions.
- Shared responsibility.

It does not become verified evidence because it contains words such as "policy", "reviewed", "approved", "tested" or "logged".

The GTSAF screen deliberately keeps **Narrative sufficiency** separate from **Verified evidence**.

What good evidence looks like

Evidence should be:

- **Relevant**: directly supports the control requirement.
- **Scoped**: applies to the selected system, environment or shared control.
- **Current**: reflects the assessed period and current implementation.
- **Authentic**: comes from a credible source and has not been improperly altered.
- **Complete**: covers the material parts of the requirement.
- **Traceable**: linked to the control, system, owner and period.
- **Reviewable**: another assessor can understand what it proves.
- **Consistent**: does not contradict tests, findings or other records.

Evidence types

Depending on the control, evidence may include:

- Policies and standards.
- Approved procedures.
- Architecture and data-flow diagrams.
- Configuration exports.
- Identity and access records.
- API gateway or runtime configuration.
- Model and dataset documentation.
- Data lineage and provenance records.
- Supplier contracts and assurance reports.
- Approval records and decision logs.
- Board or committee minutes.
- Monitoring dashboards.
- Alert and incident records.
- Change records.

- Test results.
- Access reviews.
- Exception registers.
- Training records.
- User notices and transparency material.
- Impact assessments.
- Recovery exercises.
- Red-team or adversarial test records.

The control's Required evidence advisory identifies the expected starting set.

Evidence linkage

Each record should be linked to:

- Workspace.
- **Framework:** GTSAF.
- Control ID.
- Selected AI system or intake.
- Evidence owner.
- Relevant review period.

System-scoped AI analysis uses operational records explicitly linked to the selected system or intake. Another system's evidence is not silently reused.

If an enterprise-wide artefact applies to several systems, document that relationship rather than assuming universal applicability.

Evidence workflow

Typical evidence states include:

State	Meaning
Requested	Evidence has been requested but not received
Received	The artefact is available for review
Reviewed	An assessor has evaluated it
Accepted	It sufficiently supports the relevant requirement for the stated scope
Rejected	It does not support the claim or is materially deficient
Expired	It is no longer current for the assessment period

Evidence quality may be recorded as:

- Poor
- Fair
- Good
- Strong

Status and quality are separate. A reviewed artefact may still be Poor.

Evidence review questions

For every important artefact, ask:

1. What precise requirement does this prove?
2. Does it apply to the selected system?

3. Is it current?
4. Is it approved where approval matters?
5. Does it show design, implementation or operation?
6. What period does it cover?
7. Who produced and reviewed it?
8. Is the population complete?
9. Does it contain exceptions?
10. Does it contradict another record?
11. Can another assessor reproduce the conclusion?

Evidence by effectiveness layer

Design evidence

Shows that a control has been designed to address the risk:

- Policy.
- Standard.
- Architecture.
- Procedure.
- Control specification.
- Contractual requirement.

Implementation evidence

Shows that the design has been put in place:

- Configuration.
- Role assignments.
- Deployed rule sets.
- Approved workflow.
- Training completion.
- System inventory.
- Supplier onboarding records.

Operating evidence

Shows that the control works over time:

- Logs.
- Review records.
- Alerts.
- Decisions.
- Incident handling.
- Samples of transactions.
- Periodic access reviews.
- Completed recovery exercises.
- Repeated monitoring results.

A policy alone usually supports design, not operating effectiveness.

Control testing

Testing answers:

> Does the control operate as designed under defined conditions?

Each GTSAF control includes a specific test procedure. The assessor can adapt the sample and safety limits to the system without changing the control objective.

A complete test record

Record:

- Test objective.
- Control and system.
- Preconditions.
- Population.
- Sampling method.
- Sample size.
- Test steps.
- Expected result.
- Pass criteria.
- Safety boundaries.
- Tester.
- Test date.
- Actual result.
- Exceptions.
- Evidence captured.
- Conclusion.
- Retest requirement.

Test-result labels

Common results include:

- Passed
- Pass
- Effective
- Failed
- Fail
- Ineffective
- Exception
- Pending
- Not tested

Passed, Pass and Effective are positive operating signals.

Failed, Fail, Ineffective and Exception are adverse signals. A failed test caps control assurance at 25% until the failure is resolved and retested.

Bounded and safe testing

GTSAF testing must be proportionate and authorised.

Do not:

- Run destructive tests against production without explicit approval.
- Expose personal or sensitive data unnecessarily.
- Attempt privilege escalation outside an agreed test boundary.
- Trigger harmful autonomous actions.
- Send live customer communications.
- Modify production models or retrieval data without rollback.
- Use real secrets in test prompts.

Use:

- Test environments.
- Synthetic or masked data.
- Read-only inspection.
- Rate limits.
- Approval gates.
- Rollback plans.
- Monitoring during execution.
- Predefined stop conditions.

Sampling

Choose a sample that reflects:

- Risk and criticality.
- Transaction volume.
- Environment diversity.
- User or customer groups.
- Time period.
- Supplier or model variants.
- Known exceptions.
- Recent changes.

Critical and Enhanced controls normally require stronger sampling than Low or ordinary baseline controls.

Findings

A finding is a governed record of a control weakness or assurance exception.

Raise a finding when:

- A requirement is No.
- A Gate fails.
- Evidence is missing, rejected or expired.
- A test fails.
- The control operates inconsistently.
- Ownership is unclear.
- A supplier obligation is not demonstrated.
- The system boundary differs from the documented design.
- Monitoring is insufficient.
- An assessor cannot support an effectiveness conclusion.

Finding anatomy

Field	Purpose
Title	Concise statement of the weakness
Condition	What was observed
Criteria	What GTSAF or the organisation requires
Evidence	Records supporting the finding
Cause	Why the weakness exists
Consequence	Risk created by the weakness
Severity	Priority based on impact and exposure
Recommendation	Required corrective outcome
Owner	Accountable remediation party
Target date	Agreed completion date
Status	Open, in progress, pending validation, resolved or closed

Findings versus remediation

A finding states the assurance problem.

A remediation item tracks the work required to resolve it.

One finding may require several remediation actions. A remediation action should not be closed until the relevant control has been retested or otherwise validated.

Evidence and findings in the final conclusion

The assessor conclusion must reconcile:

- Positive evidence.
- Missing evidence.
- Rejected evidence.
- Passing tests.
- Failed tests.
- Open findings.
- Compensating controls.
- Residual risk.

Do not write "effective" while leaving an unexplained failed test or open Critical finding.

Example: policy versus operation

Control requirement:

> Privileged non-human identities used by an AI agent must be scoped, reviewed and revocable.

Possible records:

- **Narrative**: "The agent uses least privilege."
- **Design evidence**: access-control standard and approved identity architecture.
- **Implementation evidence**: actual service-account role and permission export.
- **Operating evidence**: quarterly access review and revocation logs.
- **Test**: attempt an unauthorised action and confirm it is denied; revoke the identity and confirm subsequent calls fail.

The narrative alone is insufficient. The combined evidence and test support an operating conclusion.

Evidence checklist before Assured

Before accepting Assured, confirm:

71. Reference and glossary

Quick-reference tables for GTSAF domains, labels, roles, formulas, statuses, AI Assist output and assessor terminology.

Framework statistics

Item	Value
Version	1.3
Domains	17
Controls	358
Gate controls	71
Critical controls	70
High controls	241
Medium controls	43
Low controls	4

Domain reference

ID	Domain	Controls
A	Governance, Strategy and Accountability	11
B	Legal, Regulatory and Contractual Compliance	11
C	AI Use Case Intake, Approval and Risk Tiering	11
D	Data Governance, Lineage and Provenance	11
E	Data Security and Privacy Engineering	24
F	Secure Data Acquisition and Annotation	11
G	Model Development, Validation and Robustness	11
H	Prompt, Context and Retrieval Security	25
I	Inference, API and Runtime Security	25
J	Identity, Access and NHI Security	26
K	Agentic AI and Autonomous Action Governance	29
L	Third-Party, Model and Software Supply Chain Assurance	25
M	Monitoring, Detection and AI Security Operations	35
N	Human Oversight, Transparency and Impact Management	15
O	Resilience, Continuity and Recovery	36
P	Auditability, Evidence and Assurance	11
Q	Infrastructure, Platform and Environment Security	41

Badge dimensions

Dimension	Labels
Criticality	Critical, High, Medium, Low
Control type	Gate
Applicability	Mandatory, Triggered, Enhanced, Not applicable
Result	Unassessed, Partial, Supported, Assured, Gap, Gate fail, N/A

Criticality weights

Criticality	Weight
Critical	2.0
High	1.5
Medium	1.0
Low	0.75

Question answers

Answer	Meaning
Yes	Requirement is claimed as met
No	Requirement is not met
N/A	Proposed exclusion, subject to governance approval

Shared-responsibility roles

Code	Meaning
MP	Model Provider
OSP	Orchestrated Service Provider
AP	Application Provider
AIC	AI Customer
CSP	Cloud Service Provider
Shared	Responsibility divided between named parties
ND	Not determined

Assurance components

Component	Weight
Design effectiveness	15%
Implementation effectiveness	20%
Operating effectiveness	20%
Verified evidence	25%
Coverage	10%
Resilience	10%

Audit-readiness components

Component	Weight
Assurance	50%
Ownership clarity	15%
Verified evidence adjusted for coverage	35%

Assurance score

Score	Threshold
1	Gate fail or assurance at/below 25
2	Assurance above 25 and below 55
3	Assurance at least 55 and below 75
4	Assurance at least 75
5	Assurance at least 90 and audit readiness at least 85

Caps

Condition	Cap
Gate failure	25%
Critical control with evidence below 75	54%
Critical control without passing test	74%
High control with evidence below 50	54%
Failed, ineffective or exception test	25%

Result precedence

1. N/A
2. Unassessed
3. Gate fail
4. Gap
5. Assured
6. Supported
7. Partial

Evidence workflow

- Requested
- Received
- Reviewed
- Accepted
- Rejected
- Expired

Evidence quality

- Poor
- Fair
- Good
- Strong

Test results

Positive:

- Passed
- Pass
- Effective

Adverse:

- Failed
- Fail
- Ineffective
- Exception

Other:

- Pending

- Not tested

Effectiveness labels

- Effective
- Partially effective
- Ineffective
- Not tested
- Insufficient evidence

Residual-risk labels

- Low
- Moderate
- Elevated
- High
- Critical
- Unknown

N/A decision fields

- Rationale.
- Category.
- Decision owner.
- Independent approver.
- Approval date.
- Review date or reassessment trigger.
- Compensating controls.
- Evidence references.

Scope modes

Scope	Meaning
All GTSAF controls	Full canonical control population
Selected AI system	Per-system scope derived from a complete, conflict-free authoritative route

Scope calculations

Applicable controls

Controls factually relevant to the selected system. This is a breadth measure, not a risk, compliance or effectiveness score.

Applicability pending

Controls awaiting sufficient, conflict-free facts. Pending controls are not treated as passed or out of scope.

Out of scope

Conditional controls for which the complete route contains no factual trigger.

Baseline depth

Applicable controls requiring the ordinary assurance depth.

Depth-weighted workload

Planning measure calculated as Baseline × 1 + Triggered × 2 + Enhanced × 3.

Assurance intensity

Depth-weighted workload divided by Applicable controls × 3, expressed as a percentage.

Factual scope driver

A recorded system characteristic capable of activating one or more conditional controls. Driver counts may overlap.

AI Assist actions

Per control

- Applicability analysis.
- Evidence assessment.
- Bounded test plan.
- Effectiveness analysis.
- Residual-risk analysis.
- Recommendations.
- Monitoring.
- Cross-control dependencies.

Active scope

- Executive assurance summary.
- Material gaps.
- Evidence and testing gaps.
- Ownership issues.
- Residual-risk decisions.
- Monitoring needs.
- Prioritised remediation roadmap.

Glossary

ACRS

Agentic Capability Risk Score. Product-based risk classification using dependency, action autonomy, access scope and harm potential.

Applicable

A control or requirement included in the selected assessment boundary.

Assurance

The degree of confidence supported by assessment answers, evidence, testing, ownership, coverage and resilience.

Assured

GTSAF result for a control with no No answers and assurance of at least 75%.

Audit readiness

The degree to which another reviewer can reproduce and defend the conclusion from the record.

Compensating control

An alternative safeguard reducing risk where the primary requirement is not fully met.

Conformance

Yes answers divided by all applicable question rows, including unanswered rows.

Control

A required governance, security or assurance outcome represented by a canonical GTSAF ID.

Control owner

The party accountable for implementing or operating the control.

Evidence owner

The party responsible for producing and maintaining evidence.

Enhanced

An in-scope control requiring deeper evidence and stronger testing because of elevated risk or exposure.

Finding

A governed record describing a control weakness or assurance exception.

Gate

A control whose failure can block a positive assurance conclusion.

Gap

A result produced by at least one applicable No answer.

Mandatory baseline

Applicable controls assigned the ordinary assurance depth. Baseline, Triggered and Enhanced are mutually exclusive and reconcile to the applicable-control population.

Narrative sufficiency

Quality of implementation and customer-responsibility explanation; separate from evidence.

Operating effectiveness

Whether a control works in practice over the assessed period.

Residual risk

Risk remaining after considering implemented controls and compensating measures.

Scope

The named system, components, environments, data, users, suppliers, period and exclusions covered by the assessment.

Supported

A positive but below-Assured result: at least one Yes, no No, and assurance below 75%.

Triggered

An applicable control requiring additional assurance because a factual system feature or exposure is present. ACRS or governance weighting may also deepen the assurance requirement without changing factual applicability.

Verified evidence

Linked, reviewed and quality-evaluated artefacts or operating records supporting the assessment.

Crosswalk disclaimer

GTSAF references to external frameworks provide traceability. They do not establish automatic compliance, exact equivalence or certification.

72. Reporting and governance decisions

How GTSAF system and workspace reports use assurance exceptions, evidence, tests, findings, conclusions and residual-risk decisions.

GTSAF reporting converts the assessment record into an assurance decision package. Reports must preserve the selected scope and make adverse signals visible.

Report types

GTSAF can contribute to:

- Full assessment reports.
- Framework-focus reports.
- Executive summaries.
- Board packs.
- Evidence packs.
- Control-testing packs.
- Workpaper packs.
- Workspace portfolio reports.
- Selected-system reports.

The selected pack controls the level of detail, not the underlying assessment truth.

Selected-system reports

A selected-system report uses the GTSAF assessment bucket associated with the selected system's intake.

It includes that system's:

- Derived control scores.
- Answers.
- Ownership.
- Implementation and customer-responsibility records.
- N/A decisions.
- Assessor conclusions.
- Evidence.
- Evidence requests.
- Tests.
- Findings.
- AI analysis where included and correctly scoped.

If no matching GTSAF system bucket exists, the report fails closed to an unassessed position. It does not fall back to whichever workspace-level or browser-active bucket happens to exist.

Workspace reports

A workspace report provides an aggregate view of the AI estate.

It is useful for:

- Governance oversight.
- Portfolio prioritisation.
- Board reporting.
- Identifying repeated weaknesses.
- Monitoring assessment coverage.

It does not prove that every registered system shares the workspace average.

What counts as an action item

GTSAF does not use a generic "below 60%" rule to decide action items.

A control requires attention when one or more of these conditions exists:

- Gate failure.

- Failed, ineffective or exception test.
- Rejected or expired evidence.
- Open finding.
- Unassessed in-scope control.
- Critical control without required evidence.
- Critical control without a passing test.

  This is more defensible than treating every score below a single threshold as equally important.

Control-level report content

A full control table can include:

- Control ID and title.
- Domain.
- Requirement.
- Criticality.
- Gate status.
- Assurance score.
- Assessment status.
- Uploaded evidence count.
- Accepted/reviewed evidence count.
- Rejected/expired evidence count.
- Open findings.
- Test procedure.
- Passing and failed test records.
- Effectiveness conclusion.
- Residual risk.
- Assessor conclusion.
- Reviewer and review date.

System-level decision relevance

For a system report, the report should distinguish:

Monitor

No material GTSAF assurance exception currently requires action.

Action, evidence completion, remediation or risk decision required

One or more GTSAF attention conditions exists.

The decision may require:

- Additional evidence.
- Control implementation.
- Retesting.
- Finding remediation.
- Compensating controls.
- Deployment restriction.
- Formal residual-risk acceptance.

- Escalation to governance or board level.

Gate failures in reporting

A Gate failure should be:

- Explicitly named.
- Linked to the affected system.
- Connected to the failed requirement.
- Supported by evidence or assessment rationale.
- Assigned an owner.
- Given a target decision or remediation date.
- Reflected in deployment or use restrictions.

Do not bury a Gate failure in a domain average.

Critical controls in reporting

For Critical controls, report:

- Whether verified evidence is sufficient.
- Whether operating effectiveness has been tested.
- Whether tests passed.
- Whether evidence is current.
- Whether findings remain open.
- Whether residual risk is accepted and by whom.

An Assured label without a passing test should be treated as a defect because the scoring model prevents that outcome.

Evidence reporting

Evidence reporting should state:

- What was uploaded.
- What was reviewed.
- What was accepted.
- What was rejected or expired.
- Which controls lack evidence.
- Which systems the evidence applies to.
- Whether evidence supports design, implementation or operation.
- Whether the evidence period is current.

Avoid statements such as "evidence is complete" based solely on document count.

Testing reporting

Testing reporting should state:

- Controls tested.
- Test period.
- Sample.
- Passing tests.
- Failed tests.
- Exceptions.

- Retests required.
- Critical controls not tested.

Failed tests remain action items even when the weighted score appears strong.

Findings reporting

Summarise findings by:

- Severity.
- Status.
- System.
- Control.
- Owner.
- Target date.
- Overdue state.
- Root cause.
- Residual risk.

Board reporting should focus on:

- Gate failures.
- Critical findings.
- Repeated systemic weaknesses.
- Unaccepted high residual risk.
- Overdue remediation.
- Decisions requiring authority or funding.

Assessor conclusion in reports

The conclusion should explain:

- Scope.
- Work performed.
- Evidence reviewed.
- Testing completed.
- Limitations.
- Effectiveness.
- Exceptions.
- Residual risk.
- Monitoring.
- Reassessment.

It should not merely repeat the score.

Residual-risk decisions

A residual-risk decision should identify:

- Risk owner.
- Decision authority.
- Risk level.

- Open gaps.
- Compensating controls.
- Exposure period.
- Conditions of acceptance.
- Expiry or review date.
- Reassessment triggers.

Risk acceptance is separate from control assurance. A control may remain a Gap even when the organisation formally accepts the residual risk.

Cross-framework reporting

GTSAF mappings can identify related:

- EU AI Act obligations.
- ISO/IEC 42001 clauses and controls.
- ISO/IEC 42005 impact-assessment expectations.
- NIST AI RMF functions and categories.

Use the mappings to support traceability and evidence reuse.

Do not state:

> "GTSAF control passed, therefore the organisation is compliant with the mapped law or standard."

Instead state:

> "The evidence and testing for this GTSAF control are relevant to the mapped external requirement and should be reviewed as part of that separate determination."

Board explanation

A board-level GTSAF summary should answer:

1. What AI systems are in scope?
2. How much of the required control population is assessed?
3. Which Gate controls fail?
4. Which Critical controls lack evidence or testing?
5. What material findings remain open?
6. What residual risks require acceptance?
7. What decisions or investment are required?
8. How will the organisation monitor and reassess?

Quality checks before export

- Correct workspace and system selected.
- Correct framework and pack type.
- System-specific GTSAF bucket present.
- Scope shown clearly.
- No cross-system evidence leakage.
- Gate failures visible.
- Failed tests visible.
- Rejected evidence visible.
- Critical controls reviewed.

- Assessor conclusion complete.
- Residual risk recorded.
- Reviewer and date present.
- Draft or approval status correct.
- Crosswalk caveat included.

Report wording to avoid

Avoid:

- "Fully compliant."
- "Certified by GTSAF."
- "No risk."
- "All controls effective" where controls remain untested.
- "Evidence complete" based on upload count.
- "N/A" without the approved rationale.
- "AI verified" where AI only analysed the record.

Prefer:

73. Scoring and assurance model

Exact explanation of GTSAF conformance, evidence, ownership, audit readiness, assurance weighting, caps, status labels and the derived 1-to-5 score.

GTSAF is an evidence-led assurance assessment, not a self-declared maturity survey. Gamut derives several measures from assessment answers, ownership, implementation detail, evidence, tests and coverage.

> **Derived Gamut measures:** The percentage and 1-to-5 score described here are Gamut assurance measures used to support prioritisation and assessor judgement. They are not a native external certification grade.

Scope measures are not effectiveness scores

Before interpreting conformance or assurance, distinguish the scope measures shown above the assessment:

Scope measure	Purpose
Applicable controls	Factual breadth of the selected system's control population
Baseline depth	Applicable controls requiring ordinary assurance
Triggered depth	Applicable controls requiring additional attention because a condition is present
Enhanced depth	Applicable controls requiring the deepest evidence, testing and review
Depth-weighted workload	Baseline × 1, Triggered × 2 and Enhanced × 3
Assurance intensity	Workload as a percentage of maximum depth for the applicable population
Applicability pending	Controls awaiting sufficient, conflict-free routing facts
Out of scope	Conditional controls not triggered by the complete route

These measures describe **what must be assessed and how deeply**. They do not demonstrate that a control is designed, implemented or operating effectively.

The counts must reconcile:

> Baseline + Triggered + Enhanced = Applicable controls
>
> Applicable controls + Applicability pending + Out of scope = 358

See System scope, applicability and assurance depth for the full calculation and worked comparison.

The headline measures

Measure	What it means
Conformance	Whether applicable requirements are answered Yes, including unanswered requirements in the denominator
Narrative sufficiency	Whether implementation and customer-responsibility explanations are detailed enough to review
Verified evidence	Strength of linked evidence workflow and test signals
Ownership clarity	Whether applicable answers have a determined responsible party
Assurance	Weighted combination of design, implementation, operation, evidence, coverage and resilience
Audit readiness	Combined view of assurance, ownership and verified evidence
Assurance score	Derived 1-to-5 summary of assurance, subject to gates and caps

Conformance

Conformance is calculated over every in-scope, non-approved-N/A assessment row:

Conformance = Yes answers / applicable question rows

The denominator includes unanswered applicable rows.

This prevents a user from answering one easy question Yes and reporting 100% conformance while the rest of the scope is blank.

Gamut also retains an answered-only percentage as a secondary quality indicator, but the headline conformance measure is coverage-inclusive.

Example

An in-scope control has four questions:

- 2 Yes
- 0 No
- 0 approved N/A
- 2 unanswered

Headline conformance is:

2 / 4 = 50%

It is not 100%.

Approved and unapproved N/A

An approved N/A is removed from the conformance denominator.

An unapproved N/A remains in scope and counts as not met.

This prevents silent scope reduction.

Narrative sufficiency

Narrative sufficiency measures the quality of:

- Implementation detail.
- Customer or shared-responsibility detail.

Longer text does not automatically receive full credit. Repetition and low-information content can reduce the quality result.

Narrative sufficiency is useful because an assessor needs enough explanation to understand the control. It is not verified evidence.

Verified evidence

The verified-evidence measure uses linked record signals such as:

- Evidence uploaded.
- Evidence received.
- Evidence reviewed.
- Evidence accepted.
- Evidence quality rated Fair, Good or Strong.
- A passing or effective control test.
- Rejected or expired evidence.
- Poor evidence quality.
- A failed, ineffective or exception test.

Positive evidence progression raises the signal. Adverse evidence and failed testing cap it.

Representative progression:

Signal	Evidence position
Evidence uploaded	At least 25
Evidence received	At least 35
Evidence reviewed	At least 55
Evidence accepted	At least 70
Fair quality	At least 45
Good quality	At least 75
Strong quality	At least 90
Passing/effective test	At least 90
Poor or rejected evidence	Capped at 35
Failed/ineffective/exception test	Capped at 30

These signals do not mean every uploaded file is useful. The assessor must still evaluate relevance, scope, currency, integrity and sufficiency.

Ownership clarity

Ownership clarity is the percentage of applicable answered rows with a valid ownership selection.

For a Yes answer, ND is not valid ownership.

Shared ownership should be supported by an explanation of responsibility division.

Assurance components

The control assurance percentage combines:

Component	Weight
Design effectiveness	15%
Implementation effectiveness	20%
Operating effectiveness	20%
Verified evidence	25%
Coverage	10%
Resilience	10%

Design effectiveness

Does the control design address the stated risk?

Implementation effectiveness

Is the designed control actually implemented across the selected scope?

Operating effectiveness

Does the control work in practice?

Operating effectiveness requires a passing test or credible operating records. Narrative alone does not establish it.

Coverage

How much of the applicable control has been affirmatively demonstrated?

Resilience

Does the control remain dependable through change, failure, escalation and operational pressure?

What happens when a No exists

If at least one question is No:

- Design is reduced.
- Implementation is constrained by the weaker of narrative and evidence.
- Operating effectiveness is zero.
- Coverage reflects the reduced conformance.
- Resilience is constrained.
- The status becomes Gap, or Gate fail for a Gate control.

This prevents strong evidence for one part of a control from hiding an explicit failed requirement.

Assurance caps

Caps express rules that a weighted average must not override.

Gate failure

If a Gate control contains a No:

- Assurance is capped at 25%.
- The assurance score is 1.

Critical control with weak evidence

If a Critical control has verified evidence below 75:

- Assurance is capped at 54%.

It cannot reach Assured.

Critical control without a passing test

If a Critical control has no passing operating-effectiveness test:

- Assurance is capped at 74%.

It remains below the 75% Assured threshold.

High control with weak evidence

If a High control has verified evidence below 50:

- Assurance is capped at 54%.

Failed test

If a linked test is Failed, Ineffective or Exception:

- Assurance is capped at 25%.

The failure must be resolved or explicitly reflected in the risk decision.

Audit readiness

Audit readiness combines:

Component	Weight
Assurance	50%
Ownership clarity	15%
Verified evidence adjusted for coverage	35%

Audit readiness asks whether a reviewer could reproduce and defend the conclusion, not merely whether the control appears designed.

Derived assurance score

Gamut converts the assurance position to a 1-to-5 score:

Score	Rule	Practical interpretation
1	Gate fail, or assurance at or below 25	Material weakness or failed assurance condition
2	Assurance above 25 but below 55	Limited support; significant improvement required
3	Assurance at least 55 but below 75	Developing assurance; partially defensible
4	Assurance at least 75	Assured under the current record
5	Assurance at least 90 and audit readiness at least 85	Strong, well-evidenced and audit-ready assurance position

A score of 4 or 5 does not remove the need to review residual risk, findings, scope and evidence currency.

Assessment-result logic

Result	Logic
N/A	No score and the entire control is covered by approved N/A decisions
Unassessed	No derived score
Gate fail	Gate control with at least one No
Gap	Non-gate control with at least one No
Assured	No No answers and assurance at least 75
Supported	At least one Yes, no No, assurance below 75
Partial	A scored/provisional record exists without affirmative Yes support or a No

Criticality weighting in roll-ups

Domain and overall figures are weighted:

Criticality	Weight
Critical	2.0
High	1.5
Medium	1.0
Low	0.75

This means a weak Critical control affects the domain and overall posture more than a weak Low control.

Weighting does not permit a strong domain average to erase a Gate failure. Gate failures remain visible separately.

How to interpret the numbers

Always interpret the result in this order:

1. Selected system and scope.
2. Gate failures.
3. Failed tests.

4. Rejected evidence.
5. Critical control gaps.
6. Open findings.
7. Coverage and unanswered controls.
8. Residual risk.
9. Headline assurance figures.

Do not begin and end with the average score.

Common misunderstandings

"We answered Yes, so why are we not Assured?"

Because Yes is a claim. Assured also depends on coverage, ownership, verified evidence and operating effectiveness.

"We uploaded a policy, so why is evidence weak?"

The policy may not prove implementation or operation. It may also be unreviewed, stale, generic or not linked to the selected system.

"The domain average is strong, so can we ignore one Gate fail?"

No. Gate failures are explicit blockers and are not averaged away.

"Can a Critical control be Assured without testing?"

No. The no-passing-test cap keeps Critical assurance below 75.

"Does a score of 5 mean certified?"

No. It means the current Gamut record shows assurance of at least 90% and audit readiness of at least 85%, subject to the documented scope and human conclusion.

74. System scope, applicability and assurance depth

How GTSAF derives a system-specific control population, explains factual scope drivers, and applies ACRS and governance weighting without confusing breadth with risk.

GTSAF is assessed **per AI system**. The AI System Scope selector determines which system's applicability, answers, ownership, implementation narrative, evidence, tests, findings, assessor conclusions and AI analysis are active.

System scoping prevents a control record for one system from being mistaken for assurance over another system.

The governing principle

GTSAF separates three questions:

1. **Applicability:** Which controls are relevant to the system's actual characteristics?
2. **Assurance depth:** How rigorously must each applicable control be assessed?
3. **Effectiveness:** What do the answers, evidence and tests demonstrate?

These questions must not be collapsed into one score.

- The linked System Record and Use Case Intake determine factual applicability.
- ACRS and the Governance Weighting Profile may increase assurance depth, review cadence and escalation.
- Assessment answers, evidence and tests determine the demonstrated control position.

ACRS and governance weighting do not invent a factual system feature and cannot add or remove a control merely because a system has a higher or lower risk score.

Before a system can be selected

A system becomes selectable when its authoritative route is complete and conflict-free. This means the linked System Record and Intake provide enough information to make control applicability decidable.

ACRS confirmation is **not** an applicability gate. If ACRS has not been confirmed, GTSAF can still show the system's intake-derived control scope. The screen states that ACRS depth is unavailable so the assessor understands that no confirmed ACRS depth overlay is being used.

Systems with missing or conflicting routing facts remain visible but unavailable for system-specific exclusions. Complete or reconcile those facts in Routing and applicability.

The two assessment views

All GTSAF controls

This is the unscoped workspace view of the full canonical population of **358 controls**.

Use it when:

- Establishing a complete framework baseline.
- Inspecting the whole library.
- Routing facts are not yet complete.
- Professional judgement requires review beyond a system's calculated scope.

This view does not claim that all 358 controls apply to one system.

Selected AI system

This is the preferred view for a formal system assessment.

It combines:

- The linked System Record.
- The complete Use Case Intake route.
- Explicit per-control applicability conditions.
- The current governance tier.
- The selected system's ACRS position, where available.

The selected system has its own assessment bucket. Changing the selection changes the active answers, ownership, narrative, governed N/A decisions, conclusions and AI outputs. Data from the previously selected system is not reused as evidence for the new one.

Factual scope drivers

Controls are activated by facts such as:

Driver	Typical GTSAF effect
Governed data dependency	Data governance, quality, provenance, security and lifecycle controls
Personal or sensitive data	Privacy, rights, access, impact and records controls
Training or tuning	Data acquisition, annotation, model-development and validation controls
Retrieval or RAG	Retrieval, grounding, ingestion, context and prompt-injection controls
Public-facing interaction	Transparency, abuse protection, monitoring and human-impact controls
API or serving surface	Authentication, sessions, inference and runtime-security controls
Agentic execution	Identity, permissions, approvals, action monitoring and recovery controls
Automated decisioning	Oversight, contestability, explainability and impact controls
Consequential human or service impact	Stronger impact, governance and operational controls
External supplier or component	Supplier assurance, contracts, software supply chain and shared responsibility
Multi-tenant operation	Isolation, access, infrastructure and privacy controls
Regulated context	Legal accountability, records, oversight and assurance controls
Cultural or community impact	Context, participation, harm and representation controls

The screen shows each active driver with the number of conditional controls for which it is a valid trigger. **Driver counts can overlap.** A control may be triggered by both supplier dependency and RAG, for example, so driver counts must not be added together.

Reading the scope calculation

The scope cards reconcile the complete control population:

Applicable controls + Applicability pending + Out of scope = 358

Applicable controls are divided into three mutually exclusive assurance depths:

Baseline + Triggered + Enhanced = Applicable controls

Applicable controls

The number of controls factually relevant to the selected system. This measures **scope breadth**, not risk or compliance.

A lower-risk system can have a broader applicable population than a higher-risk system. For example, an externally hosted content tool may activate many supplier controls that do not apply to an internally developed fraud model.

Baseline depth

Applicable controls requiring the ordinary level of assessment evidence and review.

Triggered depth

Applicable controls activated by a relevant system feature or exposure and requiring additional attention beyond the baseline.

Enhanced depth

Applicable controls requiring the strongest evidence, testing, challenge, monitoring or review because their factual conditions, governance tier or ACRS depth justify heightened assurance.

Enhanced does not mean the control has failed. It describes required assessment rigour.

Applicability pending

Controls for which the available route cannot yet support inclusion or exclusion. Pending controls remain visible and are excluded from scope-based scores and workload calculations until the facts are resolved.

Out of scope

Controls for which the complete route provides a factual basis for exclusion. Out of scope is a routing conclusion, not an assessor claim that the control is satisfied.

Depth-weighted workload

Applicable-control count alone cannot show assessment stringency. GTSAF therefore presents a depth-weighted workload:

Workload = (Baseline × 1) + (Triggered × 2) + (Enhanced × 3)

This is a transparent planning measure. It expresses the relative assurance work implied by the active scope; it is not a compliance score, cost estimate or certification grade.

Worked calculation

If the screen shows:

- **Baseline:** 21
- **Triggered:** 251
- **Enhanced:** 15

Then:

(21 × 1) + (251 × 2) + (15 × 3) = **568 workload points**

Assurance intensity

Assurance intensity normalises the weighted workload against the maximum possible depth for the applicable population:

Assurance intensity = Workload ÷ (Applicable controls × 3) × 100

For 287 applicable controls and 568 workload points:

568 ÷ (287 × 3) × 100 = **65.97%**, displayed as **66%**

Assurance intensity makes systems with differently sized scopes easier to compare. It indicates how deeply the applicable population must be assessed, not whether the controls are effective.

Breadth is not stringency

Consider two systems:

System	Applicable	Baseline	Triggered	Enhanced	Workload	Intensity
External content generator	263	21	227	15	520	66%
Consequential fraud monitor	259	21	3	235	732	94%

The content generator has four more applicable controls because its external component activates supplier-specific requirements. The fraud monitor is nevertheless substantially more stringent: most of its applicable controls require Enhanced assurance, producing much greater workload and intensity.

Use **applicable controls** to explain breadth. Use **workload and assurance intensity** to explain required rigour.

How ACRS affects GTSAF

ACRS assesses:

Dimension	Question
Operational dependency	How dependent are people or operations on the system?
Action autonomy	How independently can it recommend, decide or act?
Access scope	What data, systems, tools, credentials or privileges can it reach?
Harm potential	How serious could failure, misuse or compromise be?

ACRS may deepen an already applicable control from Baseline to Triggered or Enhanced. It may also inform review cadence and escalation. It cannot make a supplier control applicable where no supplier exists, or remove a human-impact control where consequential impact is evidenced.

How governance weighting affects GTSAF

The Governance Weighting Profile expresses the organisation's approved risk policy. A higher governance tier may require stronger evidence, independent review, testing, monitoring or escalation for applicable controls.

Governance weighting changes **depth**, not factual applicability. This prevents a policy score from overriding the recorded architecture, data, suppliers, users and impacts of the system.

What happens when the route changes

If a material System Record or Intake fact changes after confirmation, the route becomes **Reassessment required**.

The assessor should:

1. Review the changed System Record and Intake facts.
2. Resolve any missing information or conflicts.
3. Review the four ACRS dimensions.
4. Adjust ACRS where the evidence requires it, or approve the current ACRS position.
5. Review changed framework routes, factual scope drivers and GTSAF control scope.
6. Revisit affected assessments, evidence, tests, findings and conclusions.

The green route panel means the route is complete and calculable. It does not mean that no action is required. An accompanying reassessment warning means the previous confirmation no longer represents the current facts.

Evidence and operational records

System-scoped operational records should be linked to the selected system or Intake. This applies to:

- Evidence.
- Evidence requests.
- Control tests.
- Findings.

An unlinked record is not automatically treated as system evidence. Where an organisation-wide artefact supports several systems, document and maintain the relationship to each relevant scope.

System assessment versus workspace roll-up

The selected-system assessment answers:

> How well is this named AI system controlled and assured?

The workspace roll-up answers:

> What is the aggregate assurance posture across the assessed AI estate?

The roll-up summarises system-level records. It does not transfer answers, evidence, tests, findings, conclusions or AI output between systems and must not be used to claim that every system has the same control effectiveness.

When returning to a workspace, confirm the AI System Scope selector before entering answers or running AI Assist. The assessment screen identifies the active system so the assessor can verify the boundary.

When professional judgement expands review

The route is a disciplined starting point. The assessor should review or add controls when:

- Architecture or system boundaries change.
- A model, hosting arrangement or supplier changes.
- Retrieval, tools, plugins, memory or delegated action are introduced.
- Autonomy, access or privileges increase.
- New data or affected populations enter scope.
- The system becomes public-facing.
- Jurisdiction or regulatory role changes.
- An incident, complaint or control failure reveals an unrecorded exposure.

Do not manipulate Intake facts to obtain a preferred control count. Record the system accurately, then document any additional professional-judgement review.

Explaining the result to an auditor

Use this statement:

> GTSAF is assessed per AI system. Complete System Record and Use Case Intake facts determine control applicability. ACRS and the approved governance weighting profile determine proportionate assurance depth but cannot invent or remove factual applicability. Baseline, Triggered and Enhanced controls reconcile to the applicable population. Depth-weighted workload and assurance intensity explain required rigour, while assessment answers, evidence and tests establish the demonstrated control position.

75. Worked example

A complete GTSAF example showing system routing, High, Gate, Enhanced, Gap, evidence, testing, AI Assist, remediation and final assurance.

This example shows how the GTSAF labels and workflow fit together.

Scenario

An insurer deploys **Claims Triage Assistant**, a GenAI system that:

- Summarises customer claims.
- Retrieves policy and claims information through RAG.
- Recommends triage priority.
- Processes personal and health-related information.
- Is hosted using a third-party foundation model.
- Exposes an internal API.
- Does not autonomously approve or reject claims.

Intake, applicability and ACRS

The intake records:

- **Sensitive personal data:** Yes.
- **RAG:** Yes.
- **Third-party model:** Yes.
- **API serving:** Yes.
- **Automated decision support:** Yes.
- **Public-facing:** No.
- **Agentic action:** No.

The complete authoritative route makes the system selectable for system-scoped assessment. ACRS is reviewed separately and indicates Moderate capability risk; it may deepen assurance but does not create or remove factual control applicability.

Routing result

The baseline includes governance, intake, monitoring, assurance and infrastructure controls.

The system characteristics trigger:

- Data governance and lineage.
- Data security and privacy.
- Prompt, context and retrieval security.
- Inference and API security.
- Third-party supply-chain assurance.
- Human oversight and transparency.

Higher-risk privacy and oversight controls receive Enhanced depth.

Reading the scope cards

Assume the selected-system view shows:

- **Applicable controls:** 287
- **Baseline:** 21
- **Triggered:** 251

- **Enhanced:** 15
- **Pending:** 0
- **Out of scope:** 71

The assessor first checks:

> 21 + 251 + 15 = 287 applicable controls

and:

> 287 + 0 + 71 = 358 total controls

The depth-weighted workload is:

> (21 × 1) + (251 × 2) + (15 × 3) = 568

Assurance intensity is:

> 568 ÷ (287 × 3) × 100 = 65.97%, displayed as 66%

This does not mean the system is 66% compliant. It means the applicable population requires 66% of the maximum possible assessment depth. Control answers, evidence and tests determine effectiveness.

The factual scope drivers may show governed data dependency, sensitive data, RAG, external supplier, API serving and regulated context. Driver counts can overlap and are not added together.

Example control badge

Assume the assessor opens a control for RAG source integrity and sees:

> **High · Gate · Enhanced · Gap**

High

Poisoned or unauthorised retrieval content could influence claims triage and expose sensitive data.

Gate

The system should not proceed to an unrestricted assurance conclusion if retrieval integrity is not controlled.

Enhanced

The system processes sensitive claims information and uses RAG in a consequential workflow. Stronger evidence and testing are required.

Gap

One applicable requirement is answered No.

Assessment questions

Question	Answer	Explanation
Are approved retrieval sources inventoried and owned?	Yes	A source register exists with owners and approval status.
Are retrieved documents integrity-checked before indexing?	No	The ingestion pipeline validates file type but not source signature or approved checksum.
Are retrieval permissions aligned to claims-handler access?	Yes	Index queries use the authenticated employee's access group.

The No creates:

- Gap status.
- Gate fail because the control is a Gate.
- Assurance cap of 25%.
- Proxy score of 1.

Ownership

The organisation records **Shared** responsibility:

- **AIC:** approves sources and user access.
- **AP:** implements ingestion controls and retrieval filtering.
- **CSP:** protects the hosted storage and platform.
- **MP:** has no direct responsibility for the customer's retrieval source approval.

The implementation narrative explains the division and references the service agreement.

Evidence collected

Accepted

- Retrieval source register.
- Access-control matrix.
- RAG architecture diagram.
- Quarterly access-review record.

Rejected

- Generic vendor security brochure.

Reason:

> The brochure describes platform security but does not demonstrate integrity checks in the customer's ingestion pipeline.

Test performed

Test objective:

> Determine whether an unapproved or modified document can enter the production retrieval index.

Bounded procedure:

1. Use the test environment.
2. Create a synthetic claims-guidance document.
3. Do not add it to the approved source register.
4. Submit it through the ingestion path.
5. Confirm whether indexing is blocked.
6. Repeat using an approved document modified after approval.
7. Capture logs and alerts.

Expected result:

- Both documents are rejected.
- An alert identifies the reason.
- The event is logged.

Actual result:

- Both documents are indexed.

Test result:

> Failed

The failed test independently caps assurance at 25%.

Finding

Title:

> RAG ingestion accepts unapproved and modified source documents

Condition:

> The test environment indexed both an unapproved synthetic document and a modified approved document.

Risk:

> Untrusted retrieval content could manipulate triage recommendations, expose claims information or produce inconsistent customer outcomes.

Recommendation:

- Enforce approved-source allowlisting.
- Validate integrity through signed manifests or approved checksums.
- Quarantine failed documents.
- Alert the security and claims-data owners.
- Retest before production approval.

Owner:

> Application Security Lead

AI Assist result

The per-control AI Assist panel shows:

> **Scope:** AI system: Claims Triage Assistant

It identifies:

- Applicable with High confidence.
- Evidence position Partial.
- Verified facts limited to the linked source register, access matrix and test record.
- Operating effectiveness Ineffective.
- Residual risk High.
- Recommended bounded retest after remediation.
- Monitoring for rejected ingestion attempts and integrity failures.

The assessor reviews the output but does not apply any claim that is unsupported.

Assessor conclusion before remediation

- **Design effectiveness:** Partially effective.
- **Implementation effectiveness:** Ineffective.
- **Operating effectiveness:** Ineffective.
- **Evidence conclusion:** Partial.
- **Residual risk:** High.
- **Decision:** Production approval restricted until remediation and successful retest.

Remediation

The application team implements:

- Signed source manifests.
- Approved checksum validation.
- Source-owner approval workflow.
- Quarantine storage.
- Rejection alerts.

- Immutable ingestion logs.

Retest

The assessor repeats the original procedure.

Result:

- Unapproved document rejected.
- Modified approved document rejected.
- Both attempts logged.
- Alerts sent.
- Approved unchanged document indexed successfully.

Test result:

> Passed

Evidence after remediation

- Updated ingestion design.
- Configuration export.
- Source signing procedure.
- Alert rule.
- Retest record.
- Change approval.

Final result

The No answer is changed to Yes only after the assessor confirms remediation.

The control now has:

- No No answers.
- Accepted evidence.
- Passing test.
- Determined ownership.
- Complete coverage.

Its status may move to **Assured** once calculated assurance reaches at least 75%.

The label sequence becomes:

> **High · Gate · Enhanced · Assured**

This does not mean the risk disappears. It means the current assessment record provides a defensible assurance position under the stated scope and review period.

What this example demonstrates

76. AI Assist & security

How ISO/IEC 42001 per-requirement and whole-AIMS AI Assist use saved scope and assessment records, including Privacy Mode and human controls.

AI Assist supports assessor analysis; it does not interpret the licensed standard authoritatively, issue an audit opinion or certify the AIMS.

Per-requirement analysis

Each atomic requirement has its own AI Assist control. Full-context analysis may consider:

- Saved AIMS scope.
- Requirement and clause anchor.
- Conformity result and assurance depth.
- Assessor rationale.
- Annex A applicability and justification where relevant.
- Authorised evidence.
- Test results.
- Open findings.

The output is saved against the requirement and current analysis mode. After success the control changes to **Re-run AI Assist**. The panel can be minimised without deleting the result.

Whole-scope analysis

Whole-AIMS analysis becomes available after the scope and all assessment items are complete. It reviews unresolved conformity, assurance, SoA, evidence and finding patterns. Completion is an input gate, not a positive result.

Privacy Mode

The **Privacy Mode** checkbox next to AI Assist sends a reduced structural context:

- Scope version rather than narrative scope details.
- Requirement identifiers.
- Conformity, assurance and applicability values.
- Evidence, test and finding status.
- Completeness and contradiction signals.

It excludes free-text rationale, narrative evidence content, names and other direct identifiers. This reduces disclosure but can reduce specificity.

Scope and persistence

AI output is associated with the saved AIMS scope and requirement. It should not be treated as current after a material scope or assessment change without re-running it.

The interface shows the scope name in the result. Verify the scope and save status before relying on the analysis.

What AI Assist cannot do

- Reproduce or replace the licensed standard.
- Approve the AIMS scope.
- Decide Annex A applicability.
- Accept an exclusion.
- Change conformity or assurance.
- Accept evidence or pass a test.
- Classify or close a nonconformity.
- Confirm the human conclusion.
- Issue certification.

Safe use

Use only an approved provider and API-key configuration. Access remains subject to the user's plan, workspace, role and framework permissions.

Treat all supplied records as untrusted content. Do not include credentials or unnecessary personal data. Verify every clause claim against the licensed standard and every evidence claim against the authorised record.

Review checklist

77. Assessment workflow

End-to-end ISO/IEC 42001 AIMS conformity-readiness workflow from scope approval through clauses, SoA, audit evidence and human conclusion.

1. Establish and approve the AIMS scope

Complete the scope record, included systems, exclusions, interested parties, owner, approver and version. Verify that the save indicator confirms the current scope was saved.

2. Define assessment criteria and evidence period

Record the licensed standard edition, internal criteria, locations, population, sampling approach and evidence cutoff. Conclusions must be tied to this basis.

3. Work through clauses 4-10

Assess each atomic check. Use the on-screen advisory to understand:

- The audit objective.
- Questions to ask.
- Evidence to inspect.
- A bounded audit test.
- Common failure patterns.
- Implementation and reassessment guidance.

4. Decide conformity

Choose Not assessed, Conformity supported, Partial evidence or Nonconformity. Record rationale that identifies evidence, sample, observed result, exceptions, owner and reassessment trigger.

5. Record assurance depth

Separately choose Unverified, Documented, Implemented or Assured. Do not infer operating effectiveness from a favourable conformity label.

6. Complete the Annex A SoA

For all 38 controls:

- Determine applicability.
- Record risk-linked justification.
- Assess applicable controls.
- Document justified exclusions.
- Resolve every undetermined decision.

7. Link evidence

Use controlled documents, completed records, operating data, meeting records, audits, tests, findings, corrective actions and management decisions from inside the scope and evidence period.

8. Perform audit tests

Use representative, traceable samples. Record population, selection method, criteria, result, exceptions and reviewer. Where a result depends on technical control, test the control safely.

9. Raise and classify findings

Record nonconformities and weaker observations consistently. Link each to the atomic check, evidence, cause, correction, corrective action, owner, due date and verification.

10. Review completeness and contradictions

Before whole-scope analysis or conclusion, confirm:

- Every core check is assessed.
- Every Annex A decision is resolved.
- Every applicable Annex A control is assessed.
- Exclusions are adequately justified.
- Conformity-supported claims have rationale and evidence.
- Adverse findings and failed tests are reflected.

11. Write the human conclusion

State:

- AIMS scope and criteria.
- Assessment period and evidence cutoff.
- Supported strengths.
- Material nonconformities.
- SoA status.
- Evidence and sampling limitations.
- Corrective-action priorities.
- Confidence and next review.

Confirmation records the assessor's conclusion only. It is not certification.

12. Maintain and improve

Update the assessment after material scope or system change, audit findings, incidents, new obligations, management review decisions and corrective-action verification.

Completion checklist

78. Conformity, assurance & SoA

Exact meaning of ISO/IEC 42001 conformity labels, assurance depth, readiness gates and Statement of Applicability decisions in Gamut.

Conformity labels

Label	Meaning
Not assessed	No assessor conclusion has been recorded
Conformity supported	Current evidence and testing support the atomic requirement within scope
Partial evidence	Some relevant practice or evidence exists, but the requirement is not fully supported
Nonconformity	The requirement is not met or current adverse evidence contradicts conformity

"Partial evidence" is not a favourable substitute for nonconformity. Use it where evidence is incomplete or mixed and record the resulting gap. Where the requirement itself is not fulfilled, record the nonconformity.

Assurance depth

Depth	Meaning
0 - Unverified	No objective evidence has been evaluated
1 - Documented	Approved documented information has been inspected
2 - Implemented	Implementation has been sampled in operation
3 - Assured	Operating effectiveness has been independently verified

Conformity answers **what the assessor concludes**. Assurance answers **how strongly that conclusion has been verified**.

Examples:

- **Conformity supported + Documented:** design appears suitable, but operation has not been sampled.
- **Conformity supported + Assured:** suitable evidence and independent effectiveness verification.
- **Nonconformity + Assured:** strong evidence confirms the requirement is not met.

Do not lower assurance to hide an adverse outcome.

Statement of Applicability labels

Label	Meaning
Applicability undetermined	The Annex A control has not received a defensible decision
Applicable	The control is selected as necessary treatment or otherwise required
Not applicable - justification required	A documented, risk-linked basis supports exclusion

For a not-applicable decision, explain:

- The exact scope and risk basis.
- Why the control is unnecessary.
- Relevant legal, contractual and interested-party requirements.
- Any alternative treatment.
- Decision owner and review trigger.

An unsupported exclusion remains a readiness gap.

Coverage and readiness

Coverage shows how much of the assessment has a recorded decision. It does not show that the AIMS conforms.

A whole-scope conclusion requires, at minimum:

- Complete assessment coverage.
- Complete Annex A applicability decisions.
- Assessment of every applicable Annex A control.
- Valid exclusion justifications.
- Rationale for requirement decisions.

Readiness can still be negative after completeness is achieved. A complete assessment with nonconformities is more defensible than an incomplete assessment with a high headline percentage.

Decision examples

Supported and implemented

The approved AI policy applies across the AIMS scope; sampled teams use the current version; communications and awareness records are current; exceptions are controlled. Record Conformity supported and at least Implemented, subject to

the strength of verification.

Documented but not operating

A risk-treatment process exists but sampled projects did not use it. Record Nonconformity or Partial evidence according to the exact atomic requirement, and do not exceed Documented assurance.

Annex A exclusion

A control is outside the approved treatment because the relevant capability does not exist within scope. Record the technical and risk basis, alternative safeguards if relevant, owner and change trigger. Reassess if the capability is introduced.

79. Evidence, audit & findings

Evidence design, sampling, audit testing, nonconformity and corrective action for ISO/IEC 42001 conformity readiness.

Evidence principles

Evidence should be:

- Within the approved AIMS scope.
- Relevant to the atomic requirement.
- Current for the evidence period.
- Authentic and controlled.
- Sufficient to support the conclusion.
- Representative of actual operation.
- Traceable to owner, system, site or process.

Generic governance material may establish organisation-wide design but cannot prove that every included process operates effectively.

Common evidence categories

Area	Examples
Context and scope	Context analysis, interested-party register, scope statement, process map
Leadership	Policy approval, objectives, resourcing, role assignments, management communications
Planning	Risk criteria, completed assessments, treatment plans, impact assessments, change plans
Support	Competence records, communications, document control, resource decisions
Operation	Lifecycle records, approvals, supplier controls, operational risk and impact work
Performance	Metrics, monitoring, internal audit programme and reports, management review minutes
Improvement	Nonconformity records, cause analysis, corrective action and effectiveness verification
Annex A	Control designs, operating records, tests and SoA decisions

Sampling

Define the population before selecting a sample. Consider:

- Sites and business units.
- AI roles and lifecycle stages.
- Internal and third-party systems.
- Risk and impact levels.

- New, changed and long-running processes.
- Successful and exceptional cases.
- Time periods and seasonal variation.

Record why the sample is sufficient. Convenience samples alone rarely support an organisation-level conclusion.

Audit test record

Capture:

- Requirement and objective.
- Population and selection method.
- Documents or systems inspected.
- Procedure performed.
- Expected criteria.
- Observed result.
- Exceptions and contrary evidence.
- Reviewer, date and independence.
- Conclusion and linked finding.

Findings

A finding should distinguish:

- **Nonconformity** - a requirement is not fulfilled.
- **Observation or improvement opportunity** - no demonstrated failure, but weakness may undermine future effectiveness.

Where an audit programme distinguishes major and minor nonconformity, apply the programme's defined criteria consistently. Do not downgrade a systemic failure merely to improve readiness reporting.

Corrective action

Correction addresses the immediate problem. Corrective action addresses the cause and prevents recurrence.

A complete record includes:

1. Requirement and objective evidence.
2. Scope and impact.
3. Immediate correction or containment.
4. Cause analysis.
5. Corrective action.
6. Owner and due date.
7. Verification method.
8. Effectiveness result.
9. Related risks and changes.

Closing an action without verifying effectiveness does not demonstrate improvement.

Audit-readiness checklist

80. Reference & glossary

ISO/IEC 42001 assessment labels, definitions, frequently asked questions and safe explanatory language.

Glossary

Term	Meaning in the assessment
AIMS	AI management system
AIMS scope	Approved organisational boundary to which the conclusion applies
Atomic check	Assessable part of a compound clause requirement
Conformity	Fulfilment of a requirement within the stated scope
Assurance depth	Strength of verification supporting the assessor conclusion
SoA	Statement of Applicability for Annex A reference controls
Nonconformity	Non-fulfilment of a requirement
Correction	Action addressing a detected problem
Corrective action	Action addressing cause to prevent recurrence
Evidence cutoff	Latest date covered by the conclusion
Certification	Independent decision by an appropriate certification body; not issued by Gamut

Frequently asked questions

Is every Annex A control mandatory?

The organisation must determine necessary controls through the standard's risk-treatment process and document Annex A applicability in the SoA. An exclusion needs a defensible basis.

Why 211 assessment items?

Gamut represents 173 atomic checks across clauses 4-10 and 38 individual Annex A controls. Atomic separation prevents partial fulfilment of a compound requirement from being reported as full support.

Is "Conformity supported" the same as certification?

No. It is an assessor result for one atomic check. Certification is a separate independent process.

Can a requirement be supported with Unverified assurance?

It can be recorded, but the claim lacks evaluated objective evidence and should not support a strong readiness conclusion.

Does Privacy Mode change scoring?

No. It only reduces the context sent for AI assistance.

What happens if the scope save fails?

The visible status shows the failure. Retry and confirm a successful save before leaving the page or running scope-dependent analysis.

Official resources

- ISO/IEC 42001:2023 official page
- ISO explanation of AI management systems
- ISO/IEC 42005:2025 official page

Safe statement

> Gamut supports an ISO/IEC 42001:2023 conformity-readiness assessment. It does not reproduce the standard or issue certification.

81. Reporting & certification readiness

How to report ISO/IEC 42001 conformity readiness, prepare management review and certification activity, and state conclusions safely.

Readiness report contents

Include:

- AIMS scope name, version, owner and approval.
- Standard edition and assessment criteria.
- Evidence period and cutoff.
- Coverage by clauses 4-10.
- Annex A SoA completion and applicable-control position.
- Conformity and assurance distributions.
- Material nonconformities and corrective actions.
- Evidence and sampling limitations.
- Management decisions and next review.

Human conclusion

A defensible conclusion states:

1. The exact scope.
2. The criteria used.
3. Work performed.
4. Evidence cutoff.
5. Supported areas.
6. Material nonconformities.
7. SoA position.
8. Limitations.
9. Readiness decision.
10. Owner and review trigger.

Safe example:

> The assessment covered the approved AIMS scope version 3 against ISO/IEC 42001:2023 readiness criteria through 30 June. Most core requirements were supported at implemented assurance; readiness remains conditional on closure and effectiveness verification of the stated nonconformities. This is an internal conformity-readiness conclusion, not certification.

Management review

Provide top management with enough information to decide on:

- Changes in context and interested-party needs.
- AIMS performance and objective progress.
- Risk, impact and opportunity trends.
- Audit results and nonconformities.
- Supplier and resource issues.
- Incidents and complaints.
- Corrective-action effectiveness.
- Improvement and scope changes.

Record decisions and assigned actions, not only meeting attendance.

Preparing for certification

Where certification is sought:

- Maintain a licensed copy and defined criteria.
- Establish the certification scope.
- Operate the AIMS long enough to produce evidence.
- Complete internal audit and management review.
- Maintain a coherent SoA.
- Address nonconformities.
- Prepare traceable evidence without over-collecting sensitive data.
- Engage an appropriate independent certification body.

Gamut supports evidence organisation and readiness assessment; the certification body controls its own audit plan, sampling and decision.

Reporting cautions

Do not:

- Use a percentage as a certification claim.
- Hide unresolved nonconformities in an average.
- Describe mapped framework work as automatic conformity.
- Omit the scope or evidence cutoff.
- State that Annex A controls are universally mandatory without the SoA analysis.
- Publish sensitive evidence merely to show readiness.

Continual improvement

Track trends in nonconformity recurrence, overdue action, test failure, evidence age, objectives, incident causes and management decisions. Improvement is demonstrated by changed and effective operation, not by rewriting documentation alone.

82. Scope, clauses & Annex A

How to define and approve the AIMS scope and understand Gamut's atomic coverage of ISO/IEC 42001 clauses 4-10 and Annex A.

Organisation-level scope

ISO/IEC 42001 is a management-system standard. Its scope is not simply the AI system selected on an intake form. Define the organisation, functions, sites, activities, technologies and lifecycle responsibilities covered by the AIMS.

Record:

- Scope name and version.
- Organisational and physical boundaries.
- Functions, business units and sites.
- Activities across development, provision, procurement and use.
- Included AI systems or system classes.
- Interfaces and dependencies.
- Interested parties and relevant requirements.

- Exclusions and why they do not affect the ability to achieve intended AIMS outcomes.
- Accountable owner and approver.
- Approval date and change triggers.

The scope is saved as it is edited. The visible save status distinguishes an unsaved change, saving in progress, successful save and a save failure that needs retry. Do not leave the page while a failure remains unresolved.

Clause structure

Clause 4 - Context

Establish internal and external issues, interested parties, the AIMS scope and the processes needed to establish, implement, maintain and improve it.

Clause 5 - Leadership

Assess top-management leadership, integration into business processes, resources, policy, responsibilities and reporting.

Clause 6 - Planning

Assess risks and opportunities, AI risk criteria, risk assessment and treatment, AI system impact assessment, objectives, plans and controlled change.

Clause 7 - Support

Assess resources, competence, awareness, communication and the creation, control, availability, protection, retention and change of documented information.

Clause 8 - Operation

Assess operational planning and control and the repeated performance of AI risk assessment, treatment and impact-assessment processes in the operating lifecycle.

Clause 9 - Performance evaluation

Assess monitoring, measurement, analysis, evaluation, internal audit and management review. The focus is whether the AIMS is suitable, adequate and effective, not just whether activities occurred.

Clause 10 - Improvement

Assess nonconformity handling, corrective action and continual improvement, including whether causes are addressed and remediation effectiveness is verified.

Atomic coverage

Gamut separates compound clause statements into **173 atomic checks**. This matters because a clause can contain several distinct duties-for example to define criteria, perform an activity, retain documented information and review results.

For each atomic check:

- Read its clause anchor and advisory.
- Determine the evidence population.
- Inspect design and operating evidence.
- Test a representative sample.
- Record conformity and assurance separately.
- Raise a finding where adverse evidence remains.

The wording in Gamut is assessment guidance. The licensed standard remains the normative source.

Annex A

Annex A provides a reference set of AI controls organised around themes including:

- AI policy.
- Internal organisation.
- Resources for AI systems.

- Impact assessment.
- AI system lifecycle.
- Data.
- Information for interested parties.
- Use of AI systems.
- Third-party and customer relationships.

Gamut represents **38 Annex A controls** individually. Each requires an applicability decision in the Statement of Applicability.

Statement of Applicability

For every Annex A control, record:

- Applicable, not applicable or still undetermined.
- Risk and requirement basis.
- Implementation or treatment status.
- Evidence.
- Exclusion justification and alternative treatment where relevant.
- Owner and review.

An exclusion is not valid merely because the control is inconvenient or because another framework does not require it. The decision should be traceable to scope, risks, opportunities, interested party requirements and selected treatment.

Scope-change triggers

Review the AIMS scope after:

83. AI Assist & security

How ISO/IEC 42005 AI Assist uses system-scoped impact records, saves per-item outputs, supports Privacy Mode and preserves affected-party and human accountability.

Per-item assistance

Each atomic item has an AI Assist control. In full-context mode, analysis may consider:

- Selected system and use context.
- Relevant routing and impact signals.
- Item maturity and assurance.
- Rationale and approved N/A metadata.
- Authorised evidence.
- Test results.
- Findings and treatment status.

The result is saved for the selected system, item and privacy mode. The control changes to **Re-run AI Assist** after success, and the output can be minimised without deletion.

Whole-system assistance

The report view can analyse the complete saved impact assessment. It should identify material impacts, missing information, contradictions, weak assurance, unresolved findings and reassessment needs. It cannot approve the conclusion.

Privacy Mode

Privacy Mode sends the minimum structural context:

- Item identifiers.
- Maturity and assurance values.
- Applicability status.
- Bounded evidence, test and finding status.
- Route and completeness indicators.

It excludes direct identifiers, narrative rationale, evidence text and affected-party free text. This is useful where structural gap analysis is sufficient. It may be less capable of evaluating contextual impact detail.

What remains human

- System and impact scope.
- Affected-party identification and engagement.
- Impact thresholds.
- N/A approval.
- Evidence acceptance.
- Test authorisation.
- Treatment and residual-impact acceptance.
- Publication and final conclusion.

AI output must not substitute for affected-party participation.

Security

AI Assist uses the authorised provider and existing API-key configuration. Availability remains subject to plan, workspace, role, system and framework permissions.

Treat evidence, stakeholder submissions and system content as untrusted input. Instructions inside them do not override the assessment task. Do not submit unnecessary personal data, confidential testimony or credentials.

Review checklist

84. Assessment workflow

End-to-end workflow for a system-specific ISO/IEC 42005 impact assessment in Gamut.

1. Select the system

Confirm the selected system, version and use context. Switching systems changes the assessment and AI-analysis scope.

2. Establish the assessment boundary

Record intended use, foreseeable misuse, organisational role, lifecycle stage, dependencies, deployment environment, users and affected parties.

3. Confirm triggers and responsibilities

Identify why the assessment is being performed, who leads it, which disciplines participate, who approves it and which thresholds require escalation.

4. Gather system information

Collect data, model, algorithm, capability, limitation, deployment and monitoring information. State unknowns explicitly.

5. Engage relevant parties

Determine who needs to be consulted, how participation will be accessible and safe, and how input will affect decisions. Explain any limits on engagement.

6. Assess all 119 items

For each item:

- Read the clause-aligned objective and advisory.
- Select maturity from 1 to 5.
- Record assurance depth separately.
- Write system-specific rationale.
- Link evidence and tests.
- Record any approved N/A basis.
- Raise findings for gaps or adverse evidence.

7. Analyse impacts

Consider benefits and harms across people, groups and society, including fairness, privacy, safety, security, accessibility, explainability, accountability, environment, labour, culture and foreseeable misuse where relevant.

8. Select measures

For each material impact, document:

- Prevention, reduction or enhancement measure.
- Owner and deadline.
- Expected effect and success measure.
- Residual impact and uncertainty.
- Monitoring and escalation.
- Decision if adequate treatment is not possible.

9. Review contradictions

Check whether:

- A high maturity claim lacks operating evidence.
- A favourable conclusion conflicts with a failed test or open finding.
- Affected parties were identified but not engaged.
- An N/A decision removes a material impact without approval.
- Monitoring cannot detect the impact it is meant to manage.

10. Write and approve the conclusion

State system scope, method, participants, material benefits and harms, measures, residual impacts, limitations, publication decision, owner and reassessment triggers.

11. Monitor and reassess

Review the assessment after change, incident, complaint, performance drift, new affected population, new evidence or changed external conditions.

Completion checklist

85. Evidence, engagement & findings

Evidence, affected-party engagement, impact testing and finding management for ISO/IEC 42005 assessments.

Evidence categories

Category	Examples
Scope and governance	Assessment mandate, boundary, roles, thresholds and approvals
System	System card, architecture, capabilities, limitations, intended and prohibited uses
Data and model	Provenance, quality, representativeness, evaluation, drift and change records
Deployment	Geography, culture, language, accessibility, workflow and integration evidence
Interested parties	Stakeholder mapping, engagement plan, participant input and response
Impacts	Scenarios, metrics, complaints, incidents, research, monitoring and distribution analysis
Measures	Treatment design, implementation, tests, owner, success criteria and residual impact

Affected-party engagement

Engagement should be proportionate, inclusive and capable of influencing the assessment.

Consider:

- People directly subject to decisions.
- People indirectly affected.
- Vulnerable and marginalised groups.
- Workers and operators.
- Domain experts and civil society.
- Suppliers and downstream organisations.
- Public authorities where relevant.

Record participant selection, accessibility, language, compensation, safeguarding, questions, feedback, disagreements and how decisions changed. Consultation theatre-collecting views after the decision is fixed-does not provide meaningful evidence.

Impact tests and analysis

Use methods suited to the impact:

- Disaggregated performance evaluation.
- Accessibility and usability testing.
- Scenario and misuse analysis.
- Human-factors and oversight testing.
- Privacy and security testing.
- Contestability and explanation review.
- Monitoring and incident trend analysis.
- Environmental or resource measurement.

Define the population, benchmark, threshold and limitation before claiming effectiveness.

Findings

Raise a finding where:

- Scope omits a material affected group or use.

- Required information is unavailable.
- Engagement is absent or non-representative.
- A material impact has no measure or owner.
- A test fails.
- Monitoring cannot detect a relevant harm.
- A high maturity score lacks evidence.
- Residual impact exceeds an approved threshold.
- N/A is being used as a gap substitute.

Treatment and residual impact

Treatment may avoid, reduce, share, monitor or-in a documented accountable decision-accept residual impact. For benefits, measures may increase reach or equitable distribution.

Record whether the measure transfers burden to another group or creates a new impact. Re-evaluate the complete system, not only the corrected metric.

Evidence checklist

86. Maturity, assurance & N/A

Exact meaning of ISO/IEC 42005 maturity scores, assurance depth, coverage and structured not-applicable decisions in Gamut.

Maturity scale

Score	Label	Meaning
1	**Not Addressed**	No process, documentation or controls address the item
2	**Initial**	Work is ad hoc, incomplete or weakly documented
3	**Defined**	A documented process exists and is followed
4	**Managed**	Practice is measured, reviewed and consistently applied
5	**Optimised**	Practice is integrated, evidence-rich and continually improved

Maturity describes how established the practice is. It does not state whether the resulting impact is acceptable.

Assurance depth

Depth	Meaning
Unverified	No objective evidence evaluated
Documented	Approved documentation inspected
Implemented	Practice sampled in operation
Assured	Operating effectiveness independently verified

A maturity 4 claim at Unverified assurance is an unsupported assertion. A maturity 1 result can be Assured where strong evidence confirms that the practice is absent.

Rationale

Explain:

- System-specific facts.
- Evidence and sample.

- Observed result.
- Affected-party perspective.
- Exceptions and uncertainty.
- Treatment or decision.
- Owner and reassessment trigger.

Generic statements such as "policy exists" are insufficient.

Not applicable

An item should be marked N/A only where the selected system and assessment context make the guidance genuinely irrelevant. Record:

- Detailed reason.
- Clause and contextual basis.
- Information limitation or alternative assessment where relevant.
- Decision owner.
- Review date.
- Human approval.

Until these fields are complete and approved, N/A does not remove the item from scope.

Do not use N/A because:

- Evidence is missing.
- Work has not started.
- The organisation relies on a supplier.
- An impact is difficult to measure.
- Affected parties were not available.
- Another framework covers similar work.

Coverage

Coverage includes each assessed or validly excluded atomic item. Unassessed items remain visible and prevent an incomplete assessment from appearing mature.

Read section averages alongside:

- Assessment coverage.
- Assurance depth.
- N/A count and basis.
- Failed tests.
- Findings.
- Impact severity and residual uncertainty.

Example interpretations

87. Reference & glossary

ISO/IEC 42005 assessment terms, labels, frequently asked questions and safe public explanation.

Glossary

Term	Meaning
AI system impact assessment	Structured evaluation of how an AI system and foreseeable applications may affect individuals, groups or society
Affected party	Person, group or community experiencing direct or indirect effects
Interested party	Person or organisation that can affect, be affected by, or perceive itself affected by the assessment
Actual impact	Impact already observed
Reasonably foreseeable impact	Plausible impact supported by context and evidence, not remote speculation
Measure	Technical or management action addressing harm or benefit
Residual impact	Impact remaining after measures
Maturity	How established and managed an assessment practice is
Assurance	How strongly the assessment result has been verified
Approved N/A	Human-approved decision that an item is irrelevant to the stated system context

Frequently asked questions

Is ISO/IEC 42005 certifiable?

It is guidance for AI system impact assessment. Gamut does not issue certification.

Why are there 119 items?

The module separates the process and documentation guidance into unique atomic checks so assessors can identify exact gaps rather than score broad paragraphs.

Is high maturity the same as low impact?

No. A mature process may identify a severe residual impact and decide not to proceed.

Does stakeholder engagement determine the decision?

It informs accountable judgement. The assessment should record differing views and explain the decision.

Can missing evidence be N/A?

No. Missing information is uncertainty or a gap.

Does Privacy Mode remove affected-party data from AI prompts?

It excludes direct identifiers and free-text content, sending only bounded structural state.

Official resource

- ISO/IEC 42005:2025 official page

Safe statement

Gamut provides a system-specific assessment aligned to ISO/IEC 42005:2025 guidance. The result is an accountable impact-assessment record, not certification or a guarantee that all impacts have been eliminated.

88. Reporting & reassessment

How to report ISO/IEC 42005 alignment, communicate impacts proportionately and maintain the assessment after change.

Report structure

Include:

- System, version, purpose and assessment boundary.
- Organisational role and lifecycle stage.
- Method, participants and evidence period.
- Intended use and foreseeable misuse.
- Affected people, groups and society.
- Material benefits and harms.
- Distribution, severity, likelihood, duration and uncertainty.
- Measures, owners and expected effectiveness.
- Residual impacts and accountable decisions.
- Limitations, dissent and unverified assumptions.
- Monitoring and reassessment triggers.

Reporting maturity

Show maturity with coverage and assurance. A single average cannot communicate whether a severe impact remains untreated.

Separate:

- Process maturity.
- Evidence assurance.
- Impact significance.
- Treatment status.
- Residual impact.

Proportionate disclosure

Determine which results should be shared with affected parties, customers, governance bodies or public authorities. Balance transparency with privacy, security, legal privilege and confidential information.

Redaction should protect legitimate interests without concealing material impacts or undermining accountability.

Safe conclusion language

Prefer:

> The selected system was assessed against the organisation's ISO/IEC 42005:2025-aligned method. The assessment identified the stated impacts and treatment measures, with the listed limitations and residual uncertainties. Human approval applies to this system version and use context.

Avoid:

- "ISO 42005 certified."
- "No impacts."
- "Fully safe."
- "Compliant with all impact laws."

Reassessment triggers

Reassess after:

- Purpose or use change.
- New user or affected population.
- New market, language or deployment environment.
- Model, data, supplier or architecture change.
- Increased autonomy or decision impact.
- Incident, complaint, appeal or unexpected outcome.
- Performance, fairness or drift trend.
- New research or stakeholder evidence.
- Treatment failure or threshold breach.
- Legal or policy change.

Monitoring plan

For each material impact define:

- Indicator and data source.
- Population and disaggregation.
- Threshold.
- Review frequency.
- Owner.
- Escalation and stop criteria.
- Link to reassessment.

Monitoring should be capable of detecting the impact in time to act.

89. Scope, process & catalogue

How ISO/IEC 42005 system scope, impact-assessment process and Gamut's 119 atomic items fit together.

System-specific scope

Select the registered AI system before assessing. Confirm:

- System and version.
- Purpose and intended uses.
- Foreseeable unintended use and misuse.
- Lifecycle stage.
- Organisational role.
- System and assessment boundaries.
- Models, data, interfaces, suppliers and dependencies.
- Deployment environment and jurisdictions.
- Users, operators and decision subjects.
- Directly and indirectly affected people, groups and society.

The same model in two use contexts may have different impacts and needs separate system-scoped reasoning.

When an impact assessment is needed

Use impact screening throughout the lifecycle. Relevant triggers include:

- A new AI system or materially different use.
- Consequential or legally significant decisions.
- Vulnerable or marginalised populations.
- Biometrics, emotion recognition or synthetic content.
- Critical services or public-sector functions.
- Training or deployment in a new population or geography.
- Significant model, data, supplier or autonomy change.
- Incident, complaint, monitoring trend or newly identified impact.

Incomplete material facts lead to screening; they do not support exclusion.

Clause 5 - the process

The process portion addresses:

- Establishing and documenting a consistent method.
- Integrating it with risk, privacy, security, human-rights and management processes.
- Defining timing, scope and responsibility.
- Establishing sensitive-use and impact thresholds.
- Performing and analysing the assessment.
- Recording, reporting and approval.
- Monitoring, review and improvement.

Clause 6 - assessment content

The content portion addresses:

- Scope and system information.
- Functionality, capability, purpose and use.
- Data and quality.
- Algorithms, models and evaluation.
- Deployment environment.
- Relevant interested parties.
- Actual and reasonably foreseeable impacts.
- Measures addressing harms and benefits.

Atomic catalogue

Gamut represents the guidance as **119 unique atomic items**:

Section	Items
Process Foundation	18
Governance & Triggers	25
Execution & Review	25
Documentation Content	29
Impacts & Measures	22

Atomic separation prevents a broad answer from obscuring missing scope, evidence, consultation, approval, monitoring or treatment work.

Impact dimensions to consider

For each plausible impact, examine:

- Affected people or groups.
- Benefit or harm.
- Direct, indirect and cumulative effects.
- Likelihood and magnitude.
- Duration and reversibility.
- Distribution and differential impact.
- Scale and geographic reach.
- Vulnerability and ability to contest.
- Uncertainty and evidence limitations.
- Interaction with other systems and social conditions.

  Avoid relying only on aggregate model metrics. A technically accurate system can still cause unfair, unsafe, inaccessible or socially harmful outcomes.

Relationship with routing

Authoritative routing decides whether the system should enter impact screening. The ISO/IEC 42005 assessment then establishes detailed scope and item-level applicability. Routing does not decide maturity, approve N/A or accept an impact.

Relationship with other work

- ISO/IEC 42001 integrates impact assessment into the AIMS.
- EU AI Act may require legal risk or fundamental-rights work.
- NIST AI RMF provides broader risk-management outcomes.
- NAGF applies Nigerian legal and policy routes.
- GTSAF provides wider assurance controls.

  Evidence can be reused only after scope and criteria are checked.

90. AI Assist and security

Public guide to MAESTRO per-threat AI assistance, system scope, information considered, Privacy Mode, saved outputs and human threat-modelling accountability.

MAESTRO AI Assist analyses one canonical threat at a time for the selected AI system. Per-threat placement keeps the model focused on the layer, architecture and attack path being assessed.

It cannot decide applicability, approve likelihood or impact, accept evidence, pass a test, close a finding or accept residual risk.

System and threat scope

The panel identifies:

- Selected system.
- Threat ID and layer.
- Current architecture pattern.
- Current likelihood, impact and severity where assessed.

  Switching systems changes the analysis scope. A result generated for one system is not shown as the current result for another.

Full-context information considered

Subject to authorised access, AI Assist may consider:

- Canonical threat definition and layer-specific advisory.

- Selected system, purpose and relevant operating context.
- Components, trust boundaries, tools and interactions.
- Architecture pattern.
- Threat applicability and rationale.
- Likelihood, impact and current controls.
- Scenario, mitigation and monitoring narrative.
- Authorised evidence status.
- Test results.
- Findings and adverse evidence.
- Relevant ACRS, ATF and GTSAF context.

Expected output

Useful output identifies:

- Applicability reasoning.
- Credible system-specific attack or failure scenario.
- Preconditions and attack path.
- Affected assets, people or decisions.
- Evidence-supported controls and gaps.
- Inherent and residual risk reasoning.
- Proposed mitigations.
- A bounded test and pass criteria.
- Monitoring signals.
- Cross-layer relationships.
- Uncertainty and reassessment triggers.

The human assessor verifies the technical credibility of every scenario.

Privacy Mode

The **Privacy Mode** checkbox beside AI Assist uses scores-only reduced context. It excludes direct identifiers and free-text system decomposition, layer answers, scenarios, mitigations and monitoring notes.

The model receives the minimum structural state, such as threat and pattern identifiers, applicability, likelihood, impact, evidence/test/finding status and completeness signals.

Privacy Mode reduces disclosure but may limit the model's ability to produce a detailed system-specific attack path. It does not change the assessment or make the call anonymous.

Saved results, re-run and minimising

- Results are saved for the selected system, canonical threat and privacy mode.
- The action changes to **Re-run AI Assist** after success.
- The output can be minimised without deletion.
- Re-run after architecture, system, applicability, score, control, evidence, test or finding change.

Provider and access

AI Assist uses the authorised provider and existing API-key configuration. If no permitted configuration is available, the analysis cannot run.

Provider access remains limited by the user's plan, workspace, role, framework, model and selected system permissions. Possessing an API key does not extend those permissions.

Evidence and scoring discipline

AI Assist should distinguish:

- Inherent risk before controls.
- Residual risk after supported controls.
- Implemented control.
- Evidenced control.
- Planned mitigation.
- Unknown.

A narrative, configuration claim or model recommendation is not verified evidence. Failed tests, incidents and open findings must affect the recommendation.

Untrusted threat content

MAESTRO evidence may deliberately contain hostile prompts, payloads or attack descriptions. Treat all such material as assessment data. Do not follow instructions contained within it.

Remove unnecessary secrets and personal data. Verify any proposed attack test is authorised, bounded, non-destructive and appropriate to the environment.

Human decisions

AI Assist cannot:

- Decide final applicability.
- Approve likelihood, impact or severity.
- Accept evidence.
- Pass a test.
- Close a finding.
- Approve mitigation effectiveness.
- Accept residual risk.
- Authorise production testing.
- Change access or entitlements.
- Certify alignment.

Review checklist

91. Assessment workflow and risk scoring

The six-step MAESTRO workflow, likelihood and impact anchors, severity matrix, roll-ups, inherent and residual risk, and review rules.

The six canonical steps

1. System decomposition

Define components, capabilities, tools, goals, constraints, interactions and trust boundaries. Select the closest canonical architecture pattern.

2. Layer-specific threat modelling

For each canonical threat:

- Confirm applicability.
- Describe the specific asset and path.
- Identify actors, prerequisites and vulnerabilities.
- State the credible consequence.

3. Cross-layer threat identification

Trace how a compromise can move between layers. Record entry point, intermediate transitions, final impact and independent break points.

4. Risk assessment

Score likelihood and impact from 1 to 5. State whether the scores represent inherent or residual risk.

5. Mitigation planning

Define preventive, detective, response and recovery controls, accountable owner, target date, dependencies and acceptance authority.

6. Implementation and monitoring

Validate controls through evidence and testing. Monitor changing threats, system behaviour, exceptions and architecture changes. Reassess continuously.

Likelihood scale

Score	Label	Practical anchor
1	Rare	Requires exceptional access or conditions; no credible recent path; strong controls tested
2	Unlikely	Feasible but constrained; limited exposure; several effective independent controls
3	Possible	Credible conditions exist; similar attacks occur; control coverage is mixed or partly tested
4	Likely	Exposed path, capable actor or frequent precursor; important controls are weak or untested
5	Almost certain	Active exploitation, repeated events or near misses; trivial path; controls absent or bypassed

Likelihood should consider:

- External and internal exposure.
- Attacker access and capability.
- Frequency and opportunity.
- Known techniques and threat intelligence.
- Complexity and prerequisites.
- Strength and independence of preventive and detective controls.
- Test results and prior incidents.
- Change rate and time-to-detection.

Do not reduce likelihood merely because a policy exists.

Impact scale

Score	Label	Practical anchor
1	Negligible	Local, short-lived and easily reversible; no sensitive data or consequential decision
2	Minor	Limited operational or data effect; manageable recovery; low external consequence

Score	Label	Practical anchor
3	Moderate	Material disruption, privacy issue, financial loss or affected individuals; containable
4	Major	Serious legal, customer, security, safety or business harm; broad or difficult recovery
5	Severe	Catastrophic, irreversible or systemic harm; critical infrastructure, life safety or fundamental rights

Assess impact across:

- Confidentiality, integrity and availability.
- Privacy and fundamental rights.
- Safety and physical consequences.
- Financial loss and fraud.
- Legal, regulatory and contractual duties.
- Customer and public harm.
- Mission and operational continuity.
- Reputation and market integrity.
- Blast radius, persistence, detectability and reversibility.

Use the highest credible consequence, not an average of unrelated outcomes.

Risk calculation

```
Raw risk = likelihood × impact
```

Raw product	Severity	Gamut band
1-4	1	Low
5-9	2	Moderate
10-14	3	Elevated
15-19	4	High
20-25	5	Critical

Examples:

- Likelihood 2 × impact 4 = 8 -> **Moderate**.
- Likelihood 3 × impact 4 = 12 -> **Elevated**.
- Likelihood 4 × impact 5 = 20 -> **Critical**.

Inherent versus residual risk

- **Inherent risk** assumes current controls have not yet reduced the threat.
- **Residual risk** considers controls only where their design and operation are evidenced.

Record which basis is being used. Do not mix inherent likelihood with residual impact or call a control-maturity score "risk".

Roll-up rules

Gamut uses maxima for MAESTRO roll-up:

- A section's risk is its **highest scored threat**, not its average.
- A system's posture is its **highest scored threat**.
- The workspace view reports the **worst system** and distribution of each assessed system's worst-case band.

This prevents one Critical threat from being diluted by many Low threats.

Unscored threats do not count as Low. Coverage is displayed separately.

Treatment expectations by band

Band	Expected response
Low	Confirm baseline controls and monitor material change
Moderate	Document proportionate treatment, owner and review date
Elevated	Actively reduce risk and validate prevention, detection, response and recovery
High	Treat before broader deployment unless formally accepted by authorised risk ownership
Critical	Contain or pause the capability pending effective treatment or exceptional executive acceptance

Scoring quality checks

Before accepting a score, confirm:

1. The scenario names real components.
2. Likelihood cites exposure and control evidence.
3. Impact describes a credible outcome.
4. The score basis is inherent or residual.
5. Missing information has not been converted into a low score.
6. A failed test or open finding is reflected.
7. Cross-layer amplification is considered.
8. The risk owner can understand and challenge the rationale.

When to rescore

Rescore after:

92. Evidence, testing and treatment

How to evidence, safely test, record findings, mitigate and continuously monitor every MAESTRO threat.

The assessor playbook

Every canonical threat in Gamut has a unique playbook containing:

1. Layer scope.
2. Applicability decision.
3. Threat mechanism and attack path.
4. Assessment method.
5. Questions to ask.
6. Evidence to request and acceptance criteria.
7. Audit test, safety limits and pass condition.
8. Red flags.
9. Expected control set.
10. Monitoring and reassessment.
11. Likelihood anchor.
12. Impact anchor.
13. Required assessment record.

The playbook is an assessor aid. The real architecture and evidence remain authoritative.

Evidence hierarchy

Stronger evidence combines design and operating records:

Evidence type	Examples	What it establishes
Architecture	Data flows, trust boundaries, dependency graph, permission graph	The path and scope
Configuration	Policies, IAM, admission rules, filters, quotas, network controls	Control design and implementation
Provenance	Signatures, hashes, SBOM, model and data lineage, build attestations	Integrity and origin
Operating records	Logs, alerts, approvals, exceptions, revocations, rollback events	Actual operation
Test records	Test plan, inputs, expected result, observed result, exceptions	Effectiveness under defined conditions
Independent review	Red-team, audit, supplier assessment, specialist evaluation	Challenge and corroboration
Governance	Owners, treatment decisions, risk acceptance, target dates	Accountability

A policy alone rarely proves that a technical threat is controlled.

Evidence acceptance criteria

Evidence should be:

- Bound to the selected system and threat.
- Current for the assessed version and environment.
- Attributable to a named owner or trusted source.
- Complete enough to reconstruct the relevant control.
- Protected from unauthorised change.
- Representative of real operation.
- Consistent with scores, tests and findings.

Reject or qualify:

- Generic corporate evidence with no AI-system link.
- Screenshots without date, environment or source.
- Model cards that do not match the deployed version.
- Logs that omit the relevant identity, action or outcome.
- Supplier claims without scope or independent support.
- Test summaries with no procedure or pass criteria.

Safe testing standard

Every test should define:

- Objective.
- Authorised environment and owner.
- Preconditions.
- Synthetic or approved test data.
- Procedure.
- Expected result.

- Pass criteria.
- Stop conditions.
- Maximum volume, cost, time and blast radius.
- Containment and rollback.
- Evidence to capture.
- Treatment of partial passes and exceptions.

Never recommend or execute adversarial testing against production without explicit authorisation. Prefer sandbox, test tenants, synthetic identities, canary data, test agents and tabletop or purple-team exercises.

Test families

Layer	Representative safe tests
Foundation Models	Adversarial corpus, bounded extraction, trigger differential, membership inference, compute stress
Data Operations	Poison quarantine, identity-aware retrieval, canary exfiltration, integrity change, malicious RAG document
Agent Frameworks	Unsigned dependency, parser injection, API load, alternate-path policy bypass
Deployment	Unsigned image admission, low-privilege orchestration action, insecure IaC change, east-west connection
Observability	Metric alteration, low-privilege rule change, evaluator failover, threshold evasion, log canary leakage
Security & Compliance	Evasive security sample, authority revocation, cohort outcome comparison, decision reconstruction
Agent Ecosystem	Spoofed agent message, revoked credential, unauthorised tool call, registry modification, capability verification
Cross-layer	Tabletop or controlled attack graph exercise with independent break points

Pass, partial pass and fail

- **Pass:** the expected control prevents or contains the action, produces attributable telemetry and supports recovery within the objective.
- **Partial pass:** one element succeeds but detection, attribution, containment or recovery is weak.
- **Fail:** the threat path succeeds beyond the approved condition, or the expected control does not operate.
- **Not tested:** no valid conclusion about operating effectiveness.

Partial passes and bypasses should become findings, not be rounded up to a pass.

Control design

For each threat, identify:

- Preventive controls.
- Detective controls.
- Response controls.
- Recovery controls.
- Human decision gates.
- Independent controls at different trust boundaries.
- Shared dependencies and common-mode failures.

Good treatment reduces one or more of:

- Exposure.
- Attacker access.

- Available privilege.
- Success probability.
- Time undetected.
- Blast radius.
- Persistence.
- Irreversibility.
- Recovery time.

Treatment record

A complete treatment entry should include:

- Threat ID and system-specific scenario.
- Current inherent or residual score.
- **Chosen treatment:** avoid, reduce, transfer, accept or monitor.
- Planned controls and expected risk reduction.
- Accountable owner.
- Evidence owner.
- Target date.
- Dependencies.
- Validation test.
- Residual-risk target.
- Risk-acceptance authority and expiry, if accepted.
- Monitoring signals and cadence.
- Reassessment triggers.

Findings and escalation

Raise a finding when:

- An applicable threat lacks a credible control.
- Required evidence is missing or rejected.
- A test fails or partially passes.
- Control ownership is unclear.
- A High or Critical risk lacks time-bound treatment.
- A cross-layer path has no independent break point.
- Monitoring cannot detect the attack or control failure.
- Risk acceptance lacks authority, rationale or expiry.

Critical exposure should normally trigger containment or pause. High exposure should normally be treated before broader deployment unless explicitly accepted.

Monitoring design

Define both leading and lagging indicators:

- Attempts and precursor behaviour.
- Control denials and bypass attempts.
- Integrity or provenance changes.
- Privilege, model, data, tool and dependency changes.

- Drift in attack success or false-negative rates.
- Unusual cost, latency, query or resource patterns.
- Cross-layer identity and event correlation.
- Findings ageing and target-date breaches.
- Mean time to detect, contain, revoke and recover.

Monitoring must be protected against the Layer 5 threats that can corrupt or remove it.

Closing a threat treatment

Do not close treatment merely because a control was deployed. Require:

93. Layers and threat catalogue

Complete reference for all seven MAESTRO layers, 50 layer-specific threats and five cross-layer threats implemented in Gamut.

This is the complete `CSA-2025-02-06-v1` catalogue implemented in Gamut. Stable `MAE-` identifiers are Gamut assessment IDs; the layer and threat names follow the canonical CSA source.

Layer 1 - Foundation Models

The model at the core of the agent. Assess model provenance, privacy, robustness, capability, availability and the consequences of model behaviour being trusted by downstream agents.

ID	Threat	What the assessor must establish
`MAE-L1-01`	Adversarial Examples	Whether attacker-crafted text, image, audio, embedding or tool output can cause an unsafe model decision that a downstream agent trusts or acts upon
`MAE-L1-02`	Model Stealing	Whether repeated queries, outputs, probabilities, embeddings, timing or explanations permit useful extraction of protected model behaviour or parameters
`MAE-L1-03`	Backdoor Attacks	Whether models, adapters or fine-tunes contain hidden triggers that activate attacker-chosen behaviour
`MAE-L1-04`	Membership Inference Attacks	Whether responses reveal that a person, event or record was present in training or fine-tuning data
`MAE-L1-05`	Data Poisoning (Training Phase)	Whether malicious training, preference, feedback or fine-tuning records can survive ingestion and alter model behaviour
`MAE-L1-06`	Reprogramming Attacks	Whether a general model can be repurposed for prohibited or harmful tasks beyond its authorised intent
`MAE-L1-07`	Denial of Service (DoS) Attacks	Whether expensive inputs, long context, recursion, concurrency or sponge attacks exhaust model compute, latency, quota or cost

Key evidence includes model provenance, supplier attestations, hashes, evaluation sets, privacy and extraction testing, query controls, promotion records, rollback capability, rate limits and model-specific monitoring.

Key mistake: testing only average model quality. MAESTRO requires adversarial, targeted and downstream-impact analysis.

Layer 2 - Data Operations

Data processing, preparation and storage for agents, including operational databases, vector stores, retrieval pipelines, feedback, memory and training data.

ID	Threat	What the assessor must establish
MAE-L2-01	Data Poisoning	Whether untrusted operational, retrieval, feedback, memory or fine-tuning data can be accepted as authoritative and steer behaviour
MAE-L2-02	Data Exfiltration	Whether direct access, retrieval, prompts, tools, memory or iterative queries can remove sensitive data
MAE-L2-03	Denial of Service on Data Infrastructure	Whether overload, malformed queries, lock contention, deletion or dependency failure can remove or stale required data services
MAE-L2-04	Data Tampering	Whether data or metadata can be altered at rest, in transit, during transformation or within an index without authorised attribution
MAE-L2-05	Compromised RAG Pipelines	Whether malicious documents, connectors, parsers, metadata or retrieved instructions can manipulate context, ranking or agent tools

Key evidence includes source approval, lineage, trust tiers, identity-aware retrieval, tenant isolation, integrity protection, write permissions, DLP, content scanning, connector scope, index rebuild and poison-purge capability.

Key mistake: applying output filtering while leaving retrieval permissions and source trust uncontrolled.

Layer 3 - Agent Frameworks

Frameworks, libraries, modules and APIs that build and operate the agent. This layer includes orchestration logic and enforcement points supplied by the framework.

ID	Threat	What the assessor must establish
MAE-L3-01	Compromised Framework Components	Whether malicious or vulnerable modules can alter agent planning, execution or control enforcement
MAE-L3-02	Backdoor Attacks	Whether hidden framework functions or triggers grant unauthorised control
MAE-L3-03	Input Validation Attacks	Whether untrusted inputs cross parsers, templates, APIs or execution boundaries and reach code or privileged actions
MAE-L3-04	Supply Chain Attacks	Whether direct or transitive dependencies can be compromised between supplier and production
MAE-L3-05	Denial of Service on Framework APIs	Whether framework API overload can block planning, tool use, messaging or state management
MAE-L3-06	Framework Evasion	Whether agents or attackers can bypass policy, validation, approvals or safety enforcement through alternate routes or encodings

Key evidence includes SBOMs, lockfiles, signatures, build attestations, version policy, dependency alerts, API schemas, rejected-input logs, enforcement-point inventory, bypass tests, circuit breakers and recovery evidence.

Key mistake: assuming the framework's advertised safety feature is enforced consistently across every invocation path.

Layer 4 - Deployment & Infrastructure

Cloud, on-premise, edge, container, orchestration, network and compute infrastructure supporting agent operation.

ID	Threat	What the assessor must establish
MAE-L4-01	Compromised Container Images	Whether images can be replaced, poisoned or built from untrusted content and still reach production
MAE-L4-02	Orchestration Attacks	Whether control-plane, workload, secret, scheduling or admission weaknesses permit takeover or disruption
MAE-L4-03	Infrastructure-as-Code (IaC) Manipulation	Whether tampered infrastructure code can provision insecure or attacker-controlled AI resources

ID	Threat	What the assessor must establish
MAE-L4-04	Denial of Service (DoS) Attacks	Whether volumetric or resource-exhaustion attacks defeat availability or cost controls
MAE-L4-05	Resource Hijacking	Whether compromised compute can be diverted to cryptomining or unauthorised workloads
MAE-L4-06	Lateral Movement	Whether a compromised AI workload can reach metadata, credentials, networks or sensitive services

Key evidence includes image provenance and signatures, admission policy, cluster RBAC, network policy, infrastructure change protection, policy-as-code, drift detection, capacity and cost limits, east-west telemetry, service identity scopes and tested isolation.

Key mistake: treating an AI workload as an ordinary low-privilege application when its service identity or tool connections provide a much larger blast radius.

Layer 5 - Evaluation & Observability

Evaluation and monitoring processes used to track performance, behaviour and anomalies. These systems are security-sensitive because attackers may corrupt the evidence used to trust the agent.

ID	Threat	What the assessor must establish
MAE-L5-01	Manipulation of Evaluation Metrics	Whether benchmarks, labels, test cases or aggregation can be altered to misstate performance
MAE-L5-02	Compromised Observability Tools	Whether collectors, agents, dashboards, alerting systems or monitoring dependencies can be taken over
MAE-L5-03	Denial of Service on Evaluation Infrastructure	Whether attackers can remove evaluation capacity or visibility during release and operation
MAE-L5-04	Evasion of Detection	Whether agents or attackers can remain below alert thresholds or shape telemetry to avoid detection
MAE-L5-05	Data Leakage through Observability	Whether prompts, outputs, identities, secrets or model data leak into logs, traces, metrics or dashboards
MAE-L5-06	Poisoning Observability Data	Whether false or altered telemetry can hide incidents or misdirect people and automated response

Key evidence includes benchmark provenance, evaluator separation, telemetry schemas, redaction, collector and dashboard hardening, rule-change logs, source authentication, integrity protection, reconciliation, false-negative review and fail-closed release behaviour.

Key mistake: trusting observability as objective evidence without protecting its confidentiality, integrity and availability.

Layer 6 - Security & Compliance

Layer 6 is a **vertical layer** across the architecture. The canonical threats focus on AI agents used for security, compliance, detection or enforcement because compromise can disable the controls intended to protect everything else.

ID	Threat	What the assessor must establish
MAE-L6-01	Security Agent Data Poisoning	Whether poisoning security data causes false negatives, false positives or unsafe enforcement
MAE-L6-02	Evasion of Security AI Agents	Whether adversarial techniques bypass or mislead AI security agents
MAE-L6-03	Compromised Security AI Agents	Whether takeover permits misuse of privileged visibility or enforcement authority
MAE-L6-04	Regulatory Non-Compliance by AI Security Agents	Whether security-agent processing or decisions violate privacy, legal or regulatory duties
MAE-L6-05	Bias in Security AI Agents	Whether people, systems or groups receive systematically unequal protection or enforcement

ID	Threat	What the assessor must establish
MAE-L6-06	Lack of Explainability in Security AI Agents	Whether consequential security decisions can be reconstructed, audited and challenged
MAE-L6-07	Model Extraction of AI Security Agents	Whether querying the security agent reveals enough of its model to enable replication or evasion

Key evidence includes security-data lineage, adversarial detection testing, privileged agent identity, kill switches, authority revocation, false-positive and false-negative analysis, privacy records, cohort testing, decision provenance and extraction controls.

Key mistake: granting a security agent broad privileges because its purpose is protective. Zero trust applies equally to defensive agents.

Layer 7 - Agent Ecosystem

The wider marketplace and operating ecosystem where agents interact with users, other agents, registries, discovery mechanisms, tools, suppliers and business applications.

ID	Threat	What the assessor must establish
MAE-L7-01	Compromised Agents	Whether malicious or taken-over agents can enter trusted workflows and act
MAE-L7-02	Agent Impersonation	Whether an attacker can convincingly pose as a legitimate agent to people or peer agents
MAE-L7-03	Agent Identity Attack	Whether credentials, identifiers, authorisation bindings or identity lifecycle can be compromised
MAE-L7-04	Agent Tool Misuse	Whether manipulated intent produces out-of-scope or harmful tool use
MAE-L7-05	Agent Goal Manipulation	Whether user, peer, memory or tool channels can alter the intended goal or constraints
MAE-L7-06	Marketplace Manipulation	Whether ratings, reviews, recommendations or reputation can be manipulated
MAE-L7-07	Integration Risks	Whether APIs, SDKs, protocols or connectors introduce exploitable weaknesses
MAE-L7-08	Horizontal/Vertical Solution Vulnerabilities	Whether industry- or function-specific design creates unique abuse paths
MAE-L7-09	Repudiation	Whether insufficient attribution lets an agent or operator deny an action
MAE-L7-10	Compromised Agent Registry	Whether agent identities, metadata or trusted listings can be injected or changed
MAE-L7-11	Malicious Agent Discovery	Whether discovery and ranking can promote malicious agents or hide legitimate ones
MAE-L7-12	Agent Pricing Model Manipulation	Whether pricing, metering, billing or incentives can be abused for loss or unfair advantage
MAE-L7-13	Inaccurate Agent Capability Description	Whether declared capabilities, limits and permissions differ from actual behaviour

Key evidence includes agent inventory, workload identity, signed messages, freshness controls, credential lifecycle, tool policy, goal and delegation records, marketplace integrity controls, registry change protection, integration testing, immutable action logs and capability acceptance tests.

Key mistake: trusting a declared agent identity or capability description without runtime verification.

Cross-layer threats

Cross-layer analysis models propagation and common-mode failure. It should describe the complete chain, not merely list several layer names.

ID	Threat	Required analysis
MAE-X-01	Supply Chain Attacks	Trace compromise of a model, dataset, library, service or supplier through every affected layer

ID	Threat	Required analysis
MAE-X-02	Lateral Movement	Map how access in one layer enables movement into data, models, frameworks, infrastructure or agents
MAE-X-03	Privilege Escalation	Identify conversions between infrastructure, data, framework, model and agent-action authority
MAE-X-04	Data Leakage	Trace sensitive data through storage, retrieval, prompts, memory, tools, outputs, agents and telemetry
MAE-X-05	Goal Misalignment Cascades	Model how an altered goal propagates through data, orchestration, models, agents and ecosystem interactions

For each chain record:

- Entry point.
- Affected layers.
- Identities and privileges used.
- Trust transitions.
- Intermediate assets.
- Final impact.
- Preventive break points.
- Detective signals that correlate the full path.
- Containment, revocation and rollback.
- Multi-owner treatment plan.

Threats with similar names

The same mechanism can occur in different layers:

- **Backdoor Attacks** in Layer 1 concern model weights or fine-tunes; in Layer 3 they concern framework code or hidden functions.
- **Data Poisoning** in Layer 1 concerns training-phase model influence; in Layer 2 it includes operational, retrieval, feedback and memory data; in Layer 6 it targets security-agent data.
- **Denial of Service** appears at model, data, framework, infrastructure and evaluation layers.
- **Lateral Movement** in Layer 4 is infrastructure-originating; `MAE-X-02` traces movement across architectural layers.
- **Model Extraction** in Layer 1 concerns the foundation model; Layer 6 concerns security agents whose extraction may enable defensive evasion.

Assess the layer-specific asset and consequence. Do not merge distinct canonical items into one generic risk.

94. Method, concepts and terminology

MAESTRO's purpose, principles, boundaries, labels, threat concepts and relationship to control and risk frameworks.

Why MAESTRO exists

Traditional threat methods remain useful, but agentic AI adds risks they do not always model explicitly: autonomous decision-making, goal manipulation, adversarial machine learning, non-deterministic behaviour, agent-to-agent interaction, self-learning, tool use and cascading effects across models, data, infrastructure and business ecosystems.

MAESTRO supplements, rather than invalidates, methods such as STRIDE, PASTA, LINDDUN, OCTAVE, Trike and VAST. It provides an AI-agent-specific reference architecture and threat landscape.

Canonical principles

The CSA method is based on:

- **Extended security categories:** retain traditional cybersecurity concerns while adding AI-specific attack and failure modes.
- **Multi-agent and environment focus:** include peer agents, people, tools, suppliers, registries, markets and the physical or digital environment.
- **Layered security:** examine every architectural layer and the dependencies between them.
- **AI-specific threats:** model adversarial ML, data poisoning, model extraction, autonomy and goal-related risk.
- **Risk-based prioritisation:** use likelihood and impact in the system's actual context.
- **Continuous monitoring and adaptation:** update the threat model as systems and adversaries change.

Threat, vulnerability, risk and control

Term	Meaning in a MAESTRO assessment
Asset	**Something of value:** model, dataset, identity, tool, decision, service, reputation or safety outcome
Threat	A canonical type of harmful event, attack or failure
Threat actor	A malicious user, insider, supplier, compromised agent, peer agent or automated process
Vulnerability	A weakness or missing condition that makes the threat feasible
Attack path	The sequence from entry point through trust transitions to harm
Control	A preventive, detective, responsive or recovery measure that breaks or constrains the path
Inherent risk	Risk before considering the effectiveness of current controls
Residual risk	Risk remaining after evidenced controls are considered
Finding	A documented control weakness, failed test, exception or unacceptable exposure
Treatment	Avoid, reduce, transfer, accept or monitor the risk under authorised governance

A threat is not a control failure by itself. A vulnerability is not the same as business impact. A control existing on paper does not establish that the threat is well managed.

Canonical versus Gamut-authored material

The following are canonical CSA concepts:

- The MAESTRO name and purpose.
- Seven-layer architecture.
- Layer descriptions.
- 50 layer-specific threat names.
- Five cross-layer threat names.
- Eight architecture patterns.
- Six-step workflow.
- Likelihood-and-impact risk approach.

Gamut adds operational assessment machinery:

- Stable IDs such as `MAE-L2-05`.
- System-scoped storage.
- Detailed unique assessor playbooks.
- Evidence, testing and findings panels.
- A defined five-band risk matrix.
- Reporting and cross-framework links.

- Structured AI assistance.

Gamut-authored guidance makes the CSA method assessable; it should not be represented as additional CSA text or CSA endorsement.

Labels on the screen

Label	Meaning
Layer 1-7	The canonical architectural area in which a threat originates
Cross-layer / X	A threat path spanning two or more layers
Vertical	Layer 6 applies across the full architecture
Threats assessed	Both likelihood and impact have been recorded for that number of threats
Low / Moderate / Elevated / High / Critical	Derived risk severity; higher is worse
Highest threat risk	Maximum severity among scored threats in that section or system
Workspace	Assessment is not tied to one registered AI system
Assessing system	The selected registered system whose MAESTRO bucket is active
AI-assisted assessment	Model-generated advisory output grounded in validated assessment context
Human approval required	Suggestions do not become approved assessment conclusions automatically

Applicability

A canonical threat is applicable when the required asset, interface, actor or pathway exists or is reasonably foreseeable. Applicability can be:

- **Applicable:** the required path exists.
- **Potentially applicable:** it may exist, but architecture or ownership is incomplete.
- **Not applicable:** the required asset or pathway demonstrably does not exist.
- **Insufficient information:** the assessor cannot make a defensible decision.

Do not score an unexamined threat as Low. If it is not applicable, record the architectural reason. If information is missing, leave it unscored and request the missing information.

Defence-in-depth view

For each threat, consider independent controls at several points:

1. Prevent initial access or manipulation.
2. Validate identity, provenance and authorisation.
3. Constrain privileges, tools and reachable assets.
4. Detect abnormal behaviour or integrity loss.
5. Stop or isolate the affected component.
6. Revoke credentials and delegated authority.
7. Roll back data, configuration, models or agent state.
8. Recover service and preserve forensic evidence.

Controls that share the same dependency or enforcement point may fail together and should not be counted as independent layers of defence.

What "aligned" means

A MAESTRO-aligned Gamut assessment should:

- Use the canonical layer ordering and threat names.
- Include the cross-layer threats.

- Begin with system decomposition.
- Tailor threats to the real architecture.
- Assess likelihood and impact.
- Plan mitigations.
- Implement and monitor them continuously.
- Preserve provenance and framework attribution.

Alignment does not mean every canonical threat must receive a numeric score. It means every threat is considered and applicable threats are assessed rigorously.

Relationship to other Gamut frameworks

Framework	Primary question
MAESTRO	What can attack, misuse or destabilise this system, and through which paths?
GTSAF	Which safeguards should be implemented, evidenced and tested?
ACRS	How much risk is created by dependency, autonomy, access and harm potential?
ATF	What trust controls and autonomy ceiling should govern the agent?

MAESTRO identifies the attack paths. GTSAF provides a broad control baseline. ACRS supplies capability risk context. ATF governs agent trust and action authority.

95. Reference and glossary

Quick reference for MAESTRO identifiers, counts, formulas, labels, architecture patterns, workflow and assessor terminology.

Quick facts

Item	Value
Source	Cloud Security Alliance article, 6 February 2025
Lead author	Ken Huang
Gamut catalogue	`CSA-2025-02-06-v1`
Layers	7
Layer threats	50
Cross-layer threats	5
Total threats	55
Architecture patterns	8
Workflow steps	6
Risk formula	Likelihood × impact

Layer counts

ID	Name	Threats
L1	Foundation Models	7
L2	Data Operations	5
L3	Agent Frameworks	6
L4	Deployment & Infrastructure	6
L5	Evaluation & Observability	6
L6	Security & Compliance	7
L7	Agent Ecosystem	13

ID	Name	Threats
X	Cross-Layer Threats	5

Identifier format

- `MAE-L1-01`: first threat in Layer 1.
- `MAE-L7-13`: thirteenth threat in Layer 7.
- `MAE-X-05`: fifth cross-layer threat.
- `#L` and `#I` are internal persisted score suffixes for likelihood and impact.

Likelihood

1	2	3	4	5
Rare	Unlikely	Possible	Likely	Almost certain

Impact

1	2	3	4	5
Negligible	Minor	Moderate	Major	Severe

Risk bands

Product	Severity	Label
1-4	1	Low
5-9	2	Moderate
10-14	3	Elevated
15-19	4	High
20-25	5	Critical

Higher is worse.

Architecture patterns

1. Single-Agent Pattern.
2. Multi-Agent Pattern.
3. Unconstrained Conversational Autonomy.
4. Task-Oriented Agent Pattern.
5. Hierarchical Agent Pattern.
6. Distributed Agent Ecosystem.
7. Human-in-the-Loop Collaboration.
8. Self-Learning and Adaptive Agents.

Workflow

1. System Decomposition.
2. Layer-Specific Threat Modeling.
3. Cross-Layer Threat Identification.
4. Risk Assessment.
5. Mitigation Planning.
6. Implementation and Monitoring.

Glossary

Agent ecosystem The network of agents, users, tools, suppliers, registries, marketplaces and business applications within which an agent operates.

Applicability Whether the assets and pathways required for a canonical threat exist in the selected system.

Attack path The sequence from actor or failure through entry point, weakness and trust transitions to harm.

Canonical threat A threat name and layer placement from the CSA MAESTRO catalogue.

Cross-layer threat A threat that exploits relationships between two or more architectural layers.

Defence in depth Multiple independent preventive, detective, response and recovery controls across a threat path.

Evidence An artefact or operating record supporting a claim about architecture or control operation.

Finding A recorded weakness, exception, failed test or unacceptable risk condition.

Highest threat risk The maximum severity among scored threats in the relevant section or system.

Inherent risk Risk before credit is given for current controls.

Layer A functional area of the agentic architecture used to organise threat analysis.

Residual risk Risk remaining after evidenced and effective controls are considered.

Risk severity The Low-to-Critical band derived from likelihood × impact.

System decomposition The description of components, capabilities, goals, constraints, interactions and trust boundaries.

Threat scenario The canonical threat tailored to a real actor, path, asset and consequence in the selected system.

Treatment The governed decision to avoid, reduce, transfer, accept or monitor risk.

Vertical layer Layer 6, Security & Compliance, which applies across the rest of the architecture.

Workspace roll-up Portfolio summary across registered systems; not a substitute for a system-level assessment.

Zero trust The principle that identity, device, workload, agent, data and action claims are continuously verified and constrained by least privilege.

Common questions

Is Low the same as fully mitigated?

No. Low is a risk conclusion from likelihood and impact. It is not a control-maturity score.

Does every threat need a score?

Every threat must be considered. Applicable threats need assessment. Demonstrably non-applicable threats should record the architectural reason rather than receive an artificial Low score.

Why does a layer use the highest threat?

To prevent material exposure from being diluted by averaging.

Is the cross-layer section Layer 8?

No. It is a dedicated section for attack chains spanning the seven layers.

Is Security & Compliance Layer 5 or Layer 6?

In the canonical catalogue implemented by Gamut, **Layer 5 is Evaluation & Observability** and **Layer 6 is Security & Compliance**, with Layer 6 treated as vertical.

Does MAESTRO replace GTSAF?

No. MAESTRO models threats; GTSAF assesses safeguards. They complement each other.

Does AI Assist approve the assessment?

No. Human approval is required for applicability, scores, testing, evidence and risk decisions.

Does an API key unlock MAESTRO?

No. Framework, role, model, quota and workspace entitlements are enforced separately.

Does a completed assessment prove CSA certification?

No. Gamut provides a CSA-aligned workflow and is not CSA-endorsed.

96. Reporting and governance

How MAESTRO results are rolled up, reported, governed and connected to GTSAF, ACRS, ATF, risk treatment and decision-making.

What the report should communicate

A useful MAESTRO report explains:

- The selected system and architecture pattern.
- Assessment coverage.
- Highest-risk threats.
- Layer and cross-layer attack paths.
- Existing controls and evidence quality.
- Failed or missing tests.
- Treatment owners and dates.
- Monitoring and reassessment.
- Residual risk and decision required.

The number alone is not the conclusion.

System report

For the selected system, report:

- Threats assessed out of 55.
- Highest threat severity.
- High and Critical threats.
- Unscored applicable threats.
- Cross-layer chains.
- Open findings.
- Tests passed, partially passed, failed or not run.
- Treatment status.
- Accepted risks and expiry dates.

Workspace roll-up

The workspace view shows:

- Registered systems.
- Systems with a MAESTRO assessment.
- Worst-case system risk.
- Distribution of assessed systems by each system's highest threat band.

This is a portfolio triage view. It does not prove that an unassessed system is Low risk.

AI-generated threat report

The generated narrative is instructed to cover:

1 Executive threat summary.

2 High and Critical exposures.

3 Mitigation and monitoring adequacy.

4 At least three cross-layer attack chains.

5 Supporting ACRS context.

6 MAESTRO-to-GTSAF mitigation mapping.

7 Five red-team scenarios.

8 Ownership, target-date and reassessment gaps.

It should distinguish recorded facts from recommendations and should not call a threat well controlled solely because its score is low.

Decision gates

Condition	Normal governance response
Critical risk	Contain or pause, urgent executive decision, remediation and adversarial validation
High risk	Treat before broader deployment or obtain explicit authorised acceptance
Failed test	Open finding, analyse root cause, remediate and retest
Missing owner	Assign accountable owner before accepting treatment
Cross-layer path with one control	Add independent break points and correlated detection
Unscored applicable threat	Complete assessment before claiming coverage
Expired acceptance	Reassess and renew or remediate

Relationship to ACRS

ACRS provides supporting capability-risk context:

- Operational dependency.
- Action autonomy.
- Access scope.
- Harm potential.

High autonomy, broad access and severe harm potential may increase MAESTRO likelihood, impact or treatment urgency, but ACRS does not calculate the MAESTRO score.

Relationship to GTSAF

MAESTRO threats should drive control verification:

- Foundation-model threats commonly connect to model governance, validation, robustness and supplier controls.
- Data threats connect to lineage, provenance, privacy, access and RAG security.
- Framework threats connect to secure development, dependencies, input validation and enforcement.
- Deployment threats connect to infrastructure, identity, segmentation and resilience.
- Observability threats connect to monitoring, logging, evidence integrity and incident response.
- Ecosystem threats connect to agent identity, tools, supply chain, contracts and non-repudiation.

Crosswalks identify candidate control relationships. Direct threat evidence and testing are still required.

Relationship to ATF

ATF governs agent trust and autonomy. MAESTRO findings may require:

- Lower autonomy.
- Stronger identity.
- Narrower tool scopes.

- Additional approval gates.
- Behavioural monitoring.
- Segmentation.
- Kill switch, revocation or rollback.
- Delayed promotion to a higher ATF level.

Risk-register integration

Create a risk-register entry where the threat:

- Is High or Critical.
- Requires multi-team treatment.
- Exceeds appetite.
- Has material legal, safety or financial consequence.
- Is formally accepted or transferred.
- Depends on a supplier.
- Cannot be resolved within the assessment workflow.

The risk record should reference the MAESTRO threat ID, selected system, scenario, inherent and residual risk, controls, findings, owner, decision and review date.

Reporting cautions

Avoid:

- Averaging away a Critical threat.
- Treating unscored threats as Low.
- Reporting workspace results as a system conclusion.
- Calling mitigation plans implemented.
- Calling a passed design review operating effectiveness.
- Claiming CSA certification or endorsement.
- Claiming regulatory compliance from a threat model.
- Publishing sensitive attack detail beyond the intended audience.

Board-level explanation

MAESTRO decomposes each agentic AI system into seven architectural layers and models both layer-specific and cascading threats. Each applicable threat is scored by likelihood and impact. The reported posture uses the highest risk, not an average, so material exposure remains visible. Management decisions are supported by evidence, safe testing, treatment ownership and monitoring, with Critical and High risks requiring explicit action or authorised acceptance.

Reassessment governance

Define:

- Routine review cadence.
- Event-driven triggers.
- Threat-intelligence owner.
- Model and dependency-change notification.
- Risk-acceptance expiry.
- Test recurrence.
- Report recipient and classification.

- Independent review expectations.

MAESTRO is iterative. A report is a time-bound view of one architecture and threat environment.

97. System decomposition and architecture patterns

How to define MAESTRO scope, document components and trust boundaries, select a canonical pattern and identify attack paths.

Threat scoring without a clear system model produces generic results. MAESTRO therefore begins with decomposition.

The linked System Record and Intake first establish whether architecture, model, data, tools, orchestration, ecosystem or human-interaction exposure requires MAESTRO screening. Unknown material architecture facts remain screening work. The decomposition below then supplies the layer-specific threat model; routing does not set threat applicability or risk scores. See Routing and applicability.

Choose the assessment subject

The **Assessing system** selector determines the active MAESTRO score bucket:

- Select a registered AI system for a system-level threat model.
- Select workspace scope only for a deliberate shared-platform or pre-registration review.
- Switching systems changes the active likelihood, impact and severity records.
- Workspace roll-up shows coverage and worst-case exposure without replacing system-level analysis.

Do not combine unrelated systems into one threat model merely because they share a workspace.

Four required decomposition fields

Components and trust boundaries

Record:

- Foundation model and version.
- Fine-tunes, adapters and model endpoints.
- Data sources, vector stores, memory and feedback stores.
- Agent framework, orchestrator and plugins.
- Tools, APIs, queues and business systems.
- Workload, container, network and cloud boundaries.
- Human approval points.
- External agents, suppliers, registries and marketplaces.
- Identity and credential boundaries.
- Entry and egress points.

Agent capabilities and tools

State what the system can actually do:

- Read, create, modify or delete data.
- Send messages or publish content.
- Execute code or commands.
- Make payments or commercial commitments.
- Change infrastructure or security configuration.
- Create or delegate work to agents.
- Learn from interactions or update persistent memory.

- Reach production, regulated or safety-relevant environments.

Goals and immutable constraints

Record:

- Intended objective.
- Prohibited objectives.
- System and policy instructions.
- Approval requirements.
- Transaction, time, volume and cost limits.
- Data-use and jurisdiction restrictions.
- Stop conditions.
- Whether the agent can alter, reinterpret or delegate its goal.

Interactions

Map:

- User-to-agent interactions.
- Agent-to-tool calls.
- Agent-to-agent messaging.
- Data ingestion and retrieval.
- Model invocation.
- Human feedback and approval.
- External environment signals.
- Supplier and marketplace interactions.
- Monitoring and response loops.

Eight canonical architecture patterns

Pattern	Description	Canonical threat focus
Single-Agent Pattern	One agent independently pursues a goal	Goal manipulation
Multi-Agent Pattern	Multiple agents communicate and collaborate	Communication-channel and identity attacks
Unconstrained Conversational Autonomy	Broad conversational inputs with limited task constraints	Prompt injection and jailbreaking
Task-Oriented Agent Pattern	An agent performs a defined task, often through APIs	Denial of service through overload
Hierarchical Agent Pattern	Higher-level agents direct subordinate agents	Compromise of a higher-level agent
Distributed Agent Ecosystem	Decentralised agents operate in a shared environment	Sybil and agent-impersonation attacks
Human-in-the-Loop Collaboration	Humans and agents iteratively influence decisions	Manipulation of human input or feedback
Self-Learning and Adaptive Agents	The agent changes from interactions or new data	Poisoning and backdoor-trigger injection

Select the closest primary pattern, then document secondary characteristics. A system may combine patterns; the selector is an analytical starting point, not a forced taxonomy.

Trust-boundary worksheet

For every boundary, capture:

Field	Question
Source	Who or what initiates the flow?
Destination	Which component or agent receives it?

Field	Question
Data	What content, command, identity or authority crosses?
Authentication	How is the source verified?
Authorisation	Why is this action permitted?
Integrity	How is tampering detected?
Confidentiality	How is disclosure prevented?
Validation	How is untrusted content handled?
Logging	Can the action be attributed and reconstructed?
Failure mode	What happens when verification is unavailable or ambiguous?

Attack-path construction

A useful scenario is structured as:

> Actor or failure -> entry point -> exploited weakness -> trust or privilege transition -> affected asset -> agent action or system behaviour -> business, safety, privacy or mission harm.

Example:

> A malicious supplier document enters an automatically indexed knowledge base, embeds hidden instructions, is retrieved by a privileged support agent, causes an unauthorised account action and leaks customer data through an external tool.

This path links Layer 2, Layer 3, Layer 7 and the cross-layer Data Leakage threat.

Minimum viable architecture evidence

Before scoring a complex agentic system, obtain:

- Current architecture or data-flow diagram.
- Component and dependency inventory.
- Model and model-provider details.
- Tool and permission inventory.
- Identity and service-account map.
- Data classification and lineage.
- Network and deployment diagram.
- Agent goal and policy definitions.
- Logging and monitoring architecture.
- Supplier and external-agent inventory.

Missing architecture information should increase uncertainty, not be silently interpreted as a safe design.

Reassessment triggers

Repeat decomposition when:

98. Worked example

A complete MAESTRO assessment example for a customer-support agent using RAG, tools and external services.

This example shows how the records connect. It is illustrative and not a substitute for the assessor playbook.

System

Customer Resolution Agent

- Public chat interface.

- Hosted foundation model.
- RAG over customer records and internal procedures.
- Tools to read and update CRM cases, issue limited refunds and send email.
- Human approval required above a refund threshold.
- Persistent case memory.
- External email and identity providers.

Primary pattern: Task-Oriented Agent Pattern, with Human-in-the-Loop and conversational characteristics.

Decomposition

Components and boundaries

- Browser -> application API.
- Application -> agent framework.
- Agent framework -> hosted model.
- Agent framework -> vector store.
- Agent framework -> CRM and refund API.
- Agent -> approval service.
- Agent -> email provider.
- All components -> observability platform.

Capabilities

- Retrieve customer records.
- Update a case.
- Propose and execute refunds up to a limit.
- Send customer email.
- Store case memory.

Goals and constraints

- Resolve verified customer issues.
- Never reveal another customer's data.
- Never follow instructions retrieved from customer-controlled content.
- Require approval above the defined refund limit.
- Never change identity, payment or security settings.

Layer-specific assessment

`MAE-L2-05` Compromised RAG Pipelines

Scenario:

> A customer uploads a document containing hidden instructions. The document is indexed without sanitisation, retrieved during a later case and instructs the agent to reveal records and call the refund tool.

Likelihood: 4 - Likely

- Public content enters the pipeline.
- Indexing is automatic.
- Indirect prompt-injection testing is incomplete.

Impact: 4 - Major

- Customer-data disclosure.

- Fraudulent refunds.
- Multi-customer exposure through persistent indexing.

Raw risk: `4 × 4 = 16` -> **High**.

Evidence requested:

- Connector and ingestion configuration.
- Document scanning rules.
- Source trust and provenance records.
- Retrieval permission tests.
- Prompt-injection red-team results.
- Tool policy and approval logs.

Safe test:

1. Use a non-production tenant.
2. Upload a synthetic document with a hidden instruction and canary identifier.
3. Trigger a query likely to retrieve it.
4. Confirm retrieved content is treated as untrusted data.
5. Confirm no unauthorised CRM read, refund or email occurs.
6. Confirm telemetry records the retrieval and block.
7. Purge the document and verify derivative removal.

Pass criteria:

- Hidden instruction cannot override system policy.
- Retrieval remains identity-aware.
- Tool policy independently blocks unauthorised action.
- Alert is attributable.
- Poisoned content can be removed quickly.

Treatment:

- Approved-source and file-type policy.
- Content sanitisation.
- Instruction/data separation.
- Retrieval provenance shown to the agent and reviewer.
- Independent tool authorisation.
- Poison-detection monitoring.
- **Owner:** AI Platform Lead.
- **Target:** before public launch.

`MAE-L7-04` Agent Tool Misuse

Scenario:

> A user persuades the agent to reinterpret a billing dispute as an approved refund and call the refund API repeatedly below the per-action approval threshold.

Likelihood: 3 - Possible.

Impact: 4 - Major.

Risk: `3 × 4 = 12` -> **Elevated**.

Required controls:

- Per-action and cumulative refund limits.
- Identity-bound customer and case context.
- Policy enforcement outside the model.
- Step-up approval.
- Idempotency and duplicate detection.
- Behavioural monitoring and circuit breaker.

`MAE-L5-05` Data Leakage through Observability

Scenario:

> Full prompts, retrieved customer records and tool responses are sent to a broadly accessible observability platform.

Likelihood: 4 - Likely.

Impact: 3 - Moderate.

Risk: `4 × 3 = 12` -> **Elevated**.

Controls:

- Telemetry schema minimisation.
- Redaction before export.
- Restricted dashboard access.
- Short retention for sensitive fields.
- Canary leakage scans.

Cross-layer chain

`MAE-X-04` Data Leakage

1. Layer 2 accepts a malicious document.
2. Layer 3 includes it in model context.
3. Layer 7 permits a tool call with excessive customer scope.
4. Layer 5 logs the sensitive response without redaction.
5. An attacker accesses the monitoring dashboard.

Independent break points:

- Ingestion sanitisation.
- Identity-aware retrieval.
- Instruction/data isolation.
- Tool authorisation.
- Output DLP.
- Telemetry redaction.
- Dashboard least privilege.

The chain remains material even if one individual threat is scored only Moderate.

Reassessment after treatment

After controls are deployed:

- Run the bounded RAG and tool tests.
- Review accepted evidence.
- Confirm findings are remediated.
- Rescore likelihood based on tested controls.

- Do not reduce impact merely because likelihood fell.

  If likelihood falls from 4 to 2 while impact remains 4:

  `2 × 4 = 8` -> **Moderate residual risk**.

  The risk owner may accept that residual risk with a review date, monitoring and conditions.

AI Assist use

AI Assist may:

- Suggest that `MAE-L2-05` is applicable.
- Draft the scenario from recorded architecture.
- Identify missing provenance and RAG test evidence.
- Propose a bounded test.
- Suggest likelihood and impact.
- Link the cross-layer Data Leakage threat.

The assessor must verify every suggestion, approve scores and authorise testing.

Final conclusion

A defensible conclusion would state:

> The selected system has completed all 55 applicability reviews. Two High and five Elevated inherent threats were identified. RAG poisoning and cumulative tool misuse were treated with independent ingestion, retrieval and tool-policy controls and validated in a non-production environment. Highest residual risk is Moderate. Residual risk is accepted by the named owner until the stated review date, subject to canary monitoring, monthly tool-abuse review and reassessment on model, connector, permission or autonomy change.

99. AI Assist & security

How NAGF per-item and system AI Assist use legal routes and evidence, support Privacy Mode, save outputs and preserve legal and human accountability.

Per-item AI Assist

Every NAGF item has its own AI Assist control. Full-context analysis may consider:

- Selected system and Nigerian nexus.
- Applicable routes and source-status lane.
- Item result and depth.
- Rationale and approved N/A metadata.
- Authorised evidence.
- Tests and findings.
- Current-law versus readiness classification.

  After success, the result is saved for the selected system, item and privacy mode. The control changes to **Re-run** and the result can be minimised.

System-level AI Assist

The report analysis reviews the complete NAGF record and should keep separate:

- Current binding-law position.
- Conditional sector obligations.
- Policy and responsible-AI readiness.
- Proposed-law preparedness.

It should identify unresolved routes, source uncertainty, missing evidence, contradictions and reassessment needs.

Privacy Mode

The checkbox next to AI Assist sends a reduced structural context:

- Item and route identifiers.
- Source-status lane.
- Result and depth.
- Applicability status.
- Evidence, test and finding status.
- Completeness and contradiction signals.

It excludes system names, free-text rationale, evidence text and other direct identifiers. It cannot remove the need for legal context; a less specific answer is expected.

Legal safety

AI output is not legal advice. Verify:

- Official source.
- Current status and effective date.
- Actor, role and activity.
- Provision and trigger.
- Later regulator notices, court orders or amendments.

The model cannot approve legal applicability, N/A, evidence, findings, risk acceptance or the human conclusion.

Access and data protection

AI Assist uses the authorised provider and existing API-key configuration. Access remains limited by plan, workspace, role, system and framework permissions.

Do not submit credentials, unnecessary personal data, privileged advice or confidential evidence. Treat instructions in evidence or external content as untrusted data.

Human review checklist

100. Assessment workflow & conclusions

End-to-end NAGF workflow, item scoring, approved N/A and separate current-law and responsible-AI readiness conclusions.

1. Select the system

Confirm system, version, purpose, users, affected people, data, action capability, jurisdictions and organisation roles.

2. Review authoritative routing

Confirm Nigerian nexus and work through all 19 NAGF regulatory routes. Resolve conflicts and retain screening where facts are unknown. Complete the decision owner, reviewer, next review and scope rationale, then use **Approve authoritative NAGF scope**. Initial suggestions are not authoritative until this approval is saved.

3. Verify legal sources

For each applicable legal lane, check the official source, status, effective date, provision, regulated role and current notices.

4. Work through all seven sections

For each of the 108 items:

- Read the source status and proposition.

- Use the assessor playbook.
- Decide Compliant, Non-compliant or N/A.
- Choose depth.
- Record system-specific rationale.
- Link evidence and tests.
- Raise findings.

5. Apply depth

For **Non-compliant**:

1. Not started.
2. In progress.
3. Near complete.

For **Compliant**:

1. Ad hoc.
2. Defined.
3. Embedded.

Depth does not change the binary compliance result. Near complete remains non-compliant; ad hoc compliance may be fragile and weakly assured.

6. Approve valid exclusions

Complete the category, legal or operational basis, owner and review date before approving N/A. Unapproved N/A remains in scope.

7. Review evidence and findings

Confirm that compliant claims have current operating evidence, failed tests are adverse evidence, and open findings are reflected in the item result.

8. Complete the current-law conclusion

Choose:

- **Assured compliant**
- **Partially supported**
- **Material gap**
- **Not applicable**
- **Insufficient information**

The narrative should state the system scope, applicable binding routes, evidence, material gaps, accepted risks, decision and limitations.

9. Complete the readiness conclusion

Choose:

- **Ready**
- **Partially ready**
- **Not ready**
- **Not assessed**

This separate narrative covers policy strategy, responsible-AI practice, best practice and proposed legislation. It must not be blended into the current-law result.

10. Confirm

Confirmation requires:

- Complete binding assessment gates.
- Named assessor.
- Named legal or compliance reviewer.
- Both status selections and narratives.
- Evidence cutoff.
- Next review.
- Reassessment triggers.

When all displayed gates are satisfied, use **Confirm human assessor conclusion**. Confirmation saves the assessor's current-law and readiness decisions as the human conclusion for the selected system. The AI Assist output remains advisory and does not perform this confirmation.

11. Maintain

Reassess after changes to the model, data, purpose, geography, sector, licence, role, supplier, incident, law, official guidance or regulator status.

Completion checklist

101. Evidence, testing & findings

Nigerian source verification, system evidence, bounded testing and finding management for NAGF assessments.

Evidence model

Use three connected evidence layers:

1. **Authority evidence** - official source, status, provision, effective date and applicability.
2. **Control evidence** - policy, procedure, contract, configuration, record and accountable owner.
3. **Outcome evidence** - test, monitoring, complaint, incident, decision sample and affected-party result.

A legal source proves the obligation, not compliance. A policy proves design intent, not operation.

Common evidence by route

Route	Useful evidence
Data protection	Processing record, lawful-basis analysis, notices, rights handling, impact assessment, security and processor records
Automated decisions	Decision logic, significance analysis, human review, explanations, challenge and outcome monitoring
Children/vulnerability	Age and vulnerability analysis, safeguarding, consent where applicable, accessible redress
Public sector	Procurement, approval, project clearance, architecture, records and public-law review
Identity/biometrics	Identity authority, necessity, data protection, matching evaluation, security and redress
Finance/insurance/pensions	Licence and product scope, suitability, decisions, overrides, complaints, model validation and regulator records
Telecoms/critical service	Licence context, resilience, service quality, incident and continuity evidence
Health	Clinical purpose, SaMD status, validation, safety, professional oversight and patient rights
Content/election/IP	Rights clearance, consent, provenance, labelling, moderation, complaints and escalation

Bounded tests

Tests should establish whether the specific obligation or control works. Examples:

- Trace a data-subject request through an AI system's data and downstream processors.
- Reperform a significant automated decision and test human review and explanation.
- Compare model outcomes across relevant populations or languages.
- Trace a procurement from applicability decision through approval and acceptance.
- Sample regulated advice, underwriting, claims or credit decisions.
- Test incident-route selection against actual facts and notification triggers.
- Trace synthetic content through rights, provenance, label and complaint handling.

Obtain authorisation and avoid live harm.

Findings

Raise a finding where:

- Applicable law or sector route was missed.
- Source status is wrong or unverified.
- A proposed obligation is reported as current law.
- A compliant claim lacks objective evidence.
- An item has failed outcome testing.
- A significant decision lacks meaningful review or redress.
- Affected populations show material disparity.
- A supplier prevents required access, evidence or control.
- An N/A decision lacks a specific basis.
- A regulator-status change invalidates the conclusion.

Severity

Consider:

- Binding status and enforcement exposure.
- Scale and vulnerability of affected people.
- Significance and reversibility of the outcome.
- Data sensitivity.
- Licence or critical-service implications.
- Detectability and ability to provide redress.
- Recurrence and systemic scope.

Do not assign severity from the framework section alone.

Evidence checklist

102. Legal landscape & sources

How NAGF distinguishes Nigerian enacted law, directives, sector rules, regulator status, policy, proposed legislation and good practice.

NAGF uses a source hierarchy so assessors do not collapse binding law, sector conditions, policy and future readiness into one "compliance" claim.

Documentation source check: 21 July 2026. This records the public-documentation review, not a permanent statement that every source remains unchanged.

Source lanes

Lane	Meaning	How to assess
Enacted law	Legislation currently in force	Determine scope, role and facts; assess compliance
Binding directive or regulatory code	Instrument issued under relevant authority	Confirm addressee, commencement, scope and current status
Sector rule or guidance	Rule, licence condition, guideline or official sector expectation	Apply only when the actor, activity, product and trigger match
Regulator statement or status notice	Current official interpretation, enforcement or status information	Record date and check for later notices or court action
Policy strategy	National strategy or policy direction	Assess readiness and alignment, not statutory breach
Proposed bill	Draft legislation without an effective date	Assess future readiness; do not report current non-compliance
Best practice	Responsible-AI or assurance practice	Use as readiness guidance, not binding legal authority
Inferred mapping	Cross-framework relationship	Use for navigation only; verify the primary obligation

Core cross-sector sources

The assessment includes routes anchored to official sources such as:

- Nigeria Data Protection Act 2023.
- Applicable Nigeria Data Protection Commission directives and guidance.
- National Artificial Intelligence Strategy 2025.
- Cybercrimes Act resources, including the amended Act and critical-infrastructure context where applicable.
- Copyright Act resources.
- NITDA regulations and guidelines register.
- Nigeria Startup Act.

The item advisory identifies its particular source. Follow that source rather than assuming a general AI rule.

Sector authorities represented

Depending on route, the module includes source-led checks involving:

- NDPC for data protection.
- NITDA and public authorities for digital-government and technology governance.
- NIMC for digital identity.
- CBN for regulated financial services and payments.
- NAICOM for insurance.
- PenCom for pensions.
- SEC Nigeria for capital markets and robo-advice.

- NCC for telecommunications.
- NAFDAC, NHIA and health authorities for health-related systems.
- NCAA for civil aviation and remotely piloted aircraft systems.
- FCCPC for consumer protection and relevant digital lending status.
- INEC and relevant media authorities for election and media contexts.
- ngCERT, ONSA and sector authorities for cyber and critical-infrastructure routes.

Regulator names are not substitutes for an applicability analysis. Identify the licensed or regulated actor, activity, product, system function and trigger.

Time-sensitive status

Some matters require explicit current-status checking. Examples include:

- Proposed legislation with no commencement date.
- Rules affected by litigation or an official enforcement suspension.
- Harmonisation or policy directions awaiting detailed implementation.
- Regulator registers and circulars that change over time.

Record:

- Source title and issuing authority.
- Publication and effective date where available.
- Retrieval or verification date.
- Current status.
- Exact provision or route relied upon.
- Reviewer and next legal-watch date.

Avoiding common legal overstatement

Do not:

- Treat the National AI Strategy as an enacted AI Act.
- Treat a proposed digital-economy bill as current law.
- Infer a universal data-localisation rule from the NDPA.
- Infer a universal AI-incident notification to every digital regulator.
- Describe every synthetic-content risk as subject to one general deepfake prohibition.
- Apply a financial, health, telecoms or aviation requirement without the sector trigger.
- Rely on a superseded secondary summary over a current official source.

Legal-review checklist

103. Reference & glossary

NAGF labels, route terminology, frequently asked questions and safe language for Nigerian AI governance reporting.

Key labels

Label	Meaning
Nigerian nexus	Facts connecting the system, actor, users, affected people, data, outputs or activity to Nigeria
Current law	Enacted and commenced law or binding instrument applicable to the facts

Label	Meaning
Conditional sector route	Obligation that applies only to a regulated actor, activity, licence, product or trigger
Policy readiness	Alignment with strategy or policy that is not reported as statutory compliance
Proposed law	Bill or draft without current binding effect
Compliant	Applicable obligation is fully met on current evidence
Non-compliant	Applicable obligation is not fully met
N/A	Human-approved exclusion with specific scope basis
Depth	Progress for a gap or establishment of a compliant practice
Gateway item	Item that routes the assessor to more specific atomic obligations

Frequently asked questions

Is NAGF an official Nigerian government framework?

NAGF is Gamut's consolidated assessment model. It uses official Nigerian sources but is not itself an enacted instrument or government endorsement.

Are all 108 items binding?

No. Items are status-labelled across current law, conditional sector requirements, policy, proposed-law readiness and good practice.

Why is NAGF system-specific?

Applicability depends on the system's purpose, people, data, organisation role, sector and Nigerian nexus. A portfolio conclusion cannot replace that analysis.

Does a gateway marked compliant mean the sector is compliant?

No. It means routing was performed. The downstream atomic items still need assessment.

Can an anticipated obligation create current non-compliance?

No. Proposed-law preparedness belongs in the separate readiness conclusion.

Does Privacy Mode send system notes?

No. It excludes free text and direct identifiers and uses bounded structural assessment state.

Official starting points

- Nigeria Data Protection Commission
- NITDA
- NCAIR
- Central Bank of Nigeria
- SEC Nigeria
- Nigerian Communications Commission
- FCCPC

Use the item-specific primary source and verify its current status.

Safe statement

NAGF is Gamut's system-specific Nigerian AI governance assessment. Its status-labelled items separate current legal and sector obligations from policy, proposed-law and responsible-AI readiness. It is not legal advice or regulator certification.

104. Reporting & legal change

How to report NAGF results without legal overstatement and maintain conclusions as Nigerian law, regulator status and system facts change.

Required report separation

Present at least two clear conclusions:

Current-law conclusion

Include applicable enacted law, binding directives and triggered sector rules. State system, role, routes, evidence, material gaps, accepted risks, limitations and reviewer.

Readiness conclusion

Include policy strategy, responsible-AI practice, best practice and proposed legislation. State what is preparatory rather than legally required.

Never combine the two into a single "Nigeria compliant" percentage.

Report contents

- System and Nigerian nexus.
- Organisation roles.
- Route decisions and official sources.
- Status and effective date.
- Item results and depth.
- Approved exclusions.
- Evidence and tests.
- Findings and treatment.
- Current-law status and narrative.
- Readiness status and narrative.
- Evidence cutoff, next review and reassessment triggers.

Safe conclusion example

> For the selected system and the identified Nigerian activities, the current-law assessment covers the listed data-protection and sector routes as verified at the evidence cutoff. The stated gaps prevent an assured-compliant conclusion. Separate readiness work addresses policy and proposed requirements and is not reported as current legal compliance.

Legal-change monitoring

Monitor:

- Legislation and commencement.
- NDPC directives and guidance.
- NITDA instruments and official registers.
- Sector regulator rules, circulars and licence conditions.
- Court decisions and enforcement suspensions.
- National strategy implementation.
- New elections, platform, online-safety and synthetic-content measures.

Use official sources and record verification dates.

Reassessment triggers

Reassess after:

- New or amended law.
- Regulator notice, circular, code or status change.
- Court order affecting enforcement.
- New regulated product, licence or organisation role.
- New Nigerian users, affected people or outputs.
- Material model, data, supplier or purpose change.
- Incident, complaint, rights request or adverse outcome.
- New evidence of language, accessibility or fairness impact.

Portfolio reporting

Show system-level results, route coverage, legal-status lanes, material gaps and overdue reviews. Avoid averaging a critical regulated system with low-impact readiness items.

Publication caution

Protect personal data, security-sensitive evidence, trade secrets and legal privilege. Publish enough to support accountability without disclosing protected information or implying regulator approval.

105. Scope & regulatory routing

How NAGF establishes Nigerian nexus, applies 19 legal and sector routes, handles unknowns and approves item-level exclusions.

Nigerian nexus

NAGF may require screening where the selected system:

- Is deployed or offered in Nigeria.
- Has users or affected people in Nigeria.
- Produces decisions or outputs used in Nigeria.
- Processes data subject to Nigerian requirements.
- Is developed, provided or operated by an organisation with relevant Nigerian obligations.
- Supports a Nigerian regulated activity, licence or public function.

Use the central routing facts for jurisdictions and roles. NAGF then performs its detailed legal and sector scope analysis. No Nigeria-only fields are added to the generic intake solely for NAGF.

The 19 routes

Route	Key trigger question
NDPA / GAID	Does the system process personal data within the relevant Nigerian data-protection scope?
DCPMI registration / CAR	Does the organisation or processing meet applicable NDPC registration or compliance-audit triggers?
Significant automated decisions	Does the system make or materially support decisions with significant effects?
Children or legally incapable persons	Are children or people requiring special protection affected?
Public authority / procurement	Is a public body procuring or deploying the system?
Digital identity / biometrics	Does it use identity systems, identity data or biometrics?
Employment	Does it affect recruitment, work allocation, monitoring, discipline or termination?
Intellectual property	Are protected works used in training, retrieval, generation or distribution?

Route	Key trigger question
Electoral / media	Does it create or distribute election, political or regulated media content?
Critical infrastructure	Does it support critical infrastructure or an essential service?
Financial services	Is a CBN-regulated actor, product or function involved?
Insurance	Is an insurer, intermediary or insurance function involved?
Pensions	Is a pension operator or pension-administration function involved?
Capital markets	Does it support advice, trading, market activity or digital assets?
Telecommunications	Is an NCC licensee or telecoms service involved?
Healthcare	Is it clinical, health-insurance or software-as-medical-device related?
Aviation / RPAS	Does it support civil aviation or remotely piloted aircraft?
Consumer	Does it affect consumer information, fairness, lending, complaints or redress?
Platform / content	Does it operate as an online platform or generate/distribute relevant content?

One system can trigger several routes.

Route decision

For each route:

- Determine whether it applies.
- State the official source and trigger.
- Record the organisation's role.
- Identify affected NAGF items.
- Name a decision owner.
- Set review conditions.

Unknown facts remain unresolved; they do not become N/A.

Approving the authoritative NAGF scope

The initial route suggestions are not authoritative. The assessor must complete the scope record for the selected system, including:

- Nigerian applicability;
- decision owner;
- reviewer;
- next review date;
- all legal and regulatory route decisions;
- a sufficiently detailed scope rationale.

Use **Approve authoritative NAGF scope** only after resolving the displayed scope gates. Approval saves the reviewed scope against the selected system and establishes the route basis used by the NAGF assessment. If material system or legal facts change, revisit and approve the scope again before relying on the conclusion.

Scope approval determines which obligations require assessment. It does not mark any item compliant.

Gateway items and atomic items

Some NAGF items are routing gateways. They ask whether the assessor has identified all triggered atomic obligations. A gateway marked compliant does not prove the downstream obligations are met.

For example, a financial-services gateway should lead to the exact credit, AML/CFT, open-banking, payments, identity and consumer items that apply to the function.

Item-level N/A

N/A requires:

- Detailed rationale.
- Applicability category.
- Specific legal, sector, role, technical or operational basis.
- Decision owner.
- Review date.
- Human approval.

An unapproved N/A remains in scope and affects completeness. Missing evidence is not an N/A basis.

Relationship with scoring

Routing controls which obligations are in scope. It does not mark them compliant.

- Binding items contribute to current-law conclusions.
- Conditional sector items contribute only when their trigger applies.
- Policy and proposed-law items contribute to the separate readiness conclusion.
- Crosswalk mappings never create compliance automatically.

Scope-review checklist

106. Sections & item catalogue

Assessor orientation to all seven NAGF sections and their 108 system-specific items.

AI Governance Framework - 16 items

Assesses the organisation's AI governance mandate, accountability, policy, inventory, roles, risk and impact governance, incidents, internal challenge, reporting and legal-change control.

Ask whether governance can stop or change a deployment, not only whether a committee exists.

NDPC Data Protection Compliance for AI - 20 items

Assesses relevant NDPA and NDPC requirements, including lawful basis, transparency, data-subject rights, automated decisions, children, data-protection impact, security, processors, transfers, breach routes, registration or audit obligations, identity and biometric context.

Do not assume that the original collection basis automatically covers a new AI use.

Responsible and Ethical AI - 19 items

Assesses:

- Fairness and non-discrimination.
- Accessible transparency and explanations.
- Privacy and data ethics.
- Human-centred design.
- Disability inclusion.
- Nigerian language and cultural performance.
- Ethics governance.
- Workforce and community impact.
- Environmental and sustainability concerns.
- Synthetic content, copyright, election and online-harm routes.

Measure impact across relevant Nigerian populations and languages rather than relying only on an aggregate metric.

AI Adoption and Sector Compliance - 15 items

Assesses responsible adoption, lifecycle controls, procurement, public-sector deployment, cross-sector legal duties, incident and critical-infrastructure readiness, and route gateways for regulated activity.

This section connects a broad use case to the exact atomic duties in other sections.

Sector-Specific AI Obligations - 24 items

Assesses conditional obligations and governance for:

- CBN-regulated finance, credit, payments and open banking.
- Insurance.
- Pensions.
- Capital markets and robo-advice.
- Telecommunications.
- Healthcare, health insurance and medical software.
- Aviation and RPAS.
- Consumer-facing and digital-lending activity.
- Digital identity and critical services.

Sector items apply only after actor, licence, activity, product and AI function are established.

AI Infrastructure Readiness - 6 items

Assesses infrastructure availability, resilience, cloud and data dependencies, cybersecurity, capacity and sustainability. Readiness items support reliable and sovereign decision-making but should not be presented as enacted AI-specific obligations unless anchored to a current rule.

AI Ecosystem and Talent - 8 items

Assesses competence, local research and innovation, skills, inclusion, local language capability, partnership and ecosystem development. These are important national and organisational readiness themes, distinct from current-law compliance.

Using the catalogue

Each item provides:

- Source and status.
- Legal or policy proposition.
- Auditor advisory.
- Evidence to inspect.
- Bounded audit test.
- Decision rule.
- Implementation guidance.
- Cross-framework references where relevant.

Read the item advisory before selecting a result. It is designed to let the assessor complete most of the work from the assessment screen while still opening the cited primary source for legal verification.

Catalogue cautions

107. AI-assisted assessment

Public guide to using system-scoped AI assistance for NIST AI RMF assessment while preserving evidence discipline and human accountability.

AI assistance helps an assessor analyse the selected system's NIST AI RMF record. It is advisory. It does not set final outcomes, accept evidence, pass tests, close findings or confirm the Profile.

Analysis modes

Whole-system analysis

Reviews the selected system across relevant Core outcomes and, where applicable, the Generative AI Profile.

Useful for:

- Executive summary.
- Cross-function gaps.
- Evidence and testing priorities.
- Target Profile planning.
- Reassessment triggers.

Single-outcome analysis

Focuses on one Core outcome.

Useful for:

- Interview questions.
- Evidence requests.
- Bounded test ideas.
- Pass criteria.
- Gap rationale.
- Current and Target Profile suggestions.

Information considered

Subject to the user's authorised access and configured privacy preference, analysis may consider:

- Selected system identity and purpose.
- Lifecycle and deployment context.
- Users and affected people.
- Data and model characteristics.
- Human oversight and autonomy.
- Suppliers and integrations.
- Relevant Profile priorities.
- Current and Target outcomes.
- Assurance depth and assessor notes.
- Linked evidence metadata.
- Evidence requests.
- Test records and results.
- Findings.
- Generative-AI risk signals.

The model does not directly inspect or operate the live AI system.

System scope

The selected system must be visible before analysis is run.

When switching systems:

- Confirm the new system name.
- Regenerate analysis.
- Recheck evidence, tests and findings.
- Do not reuse the prior system's conclusion.

The analysis panel can be minimised to reduce clutter. Minimising changes presentation only.

Evidence discipline

AI assistance should distinguish:

- Recorded fact.
- Assumption.
- Evidence gap.
- Adverse evidence.
- Recommendation.

It should not:

- Invent an artefact.
- Claim a test passed without a passing record.
- Treat narrative as accepted evidence.
- Ignore failed tests or open findings.
- Use evidence from another system.
- Describe a recommendation as implemented.

Structured recommendations

An AI result may include:

- Supported, partially supported, unsupported or insufficient-information conclusion.
- Verified facts.
- Assumptions.
- Adverse evidence.
- Evidence gaps.
- Suggested evidence requests.
- Suggested bounded tests and pass criteria.
- Recommended Current and Target outcomes.
- Recommended assurance depth.
- Generative-AI observations.
- Priority actions.
- Reassessment triggers.
- Caveats.

These are suggestions for human review.

Privacy and provider considerations

AI analysis uses an authorised provider configuration available to the workspace or user. If no permitted provider is configured, analysis cannot run.

Before enabling provider-bound analysis, consider:

- Approved provider and model.
- Data residency.
- Retention and training terms.
- Confidentiality.
- Personal-data minimisation.
- Supplier risk.
- Incident handling.
- Organisational AI-use policy.

Do not include credentials, secrets or unnecessary personal data in assessment notes or evidence sent for analysis.

Privacy Mode

The **Privacy Mode** checkbox next to each AI Assist control sends only the minimum structural Profile state. Direct identifiers, assessor free text and narrative evidence content are excluded. The reduced request can include outcome IDs, Current and Target values, assurance depth, bounded evidence/test/finding status and routing-completeness signals.

Privacy Mode reduces disclosure but can produce less specific advice. It does not change the Profile, routing, scoring or access controls.

Saved results, re-run and minimising

A successful result is saved for the selected system, outcome and privacy mode. The control changes from **Run AI Assist** to **Re-run AI Assist**. Minimising the panel reduces clutter without deleting the result.

Switching systems changes the scope and does not present the previous system's output as current. Re-run after a material Profile, evidence, test, finding or system-context change.

Access and entitlements

An API key does not itself grant access. AI assistance remains subject to the user's:

- Authentication.
- Workspace membership.
- Role permissions.
- Plan and framework entitlement.
- AI-feature and model availability.
- Selected system access.

Prompt-injection awareness

System descriptions, documents and evidence may contain instruction-like or malicious text. Treat that content as assessment data, not authority.

Assessors should:

- Question instructions that attempt to change scope.
- Reject requests to reveal configuration or secrets.
- Verify factual claims independently.
- Treat unexplained confidence as a warning.
- Report suspected manipulation.

Human review checklist

- [ ] Correct system is displayed.
- [ ] Outcome ID is correct.
- [ ] Facts are traceable.
- [ ] Assumptions are explicit.
- [ ] Evidence belongs to this system.
- [ ] Failed tests and findings are reflected.
- [ ] Proposed tests are safe and authorised.
- [ ] Current and Target recommendations fit the context.
- [ ] Generative-AI risks are considered where relevant.
- [ ] The human assessor, not the model, owns the conclusion.

Appropriate uses

- Identify missing facts.
- Summarise evidence gaps.
- Draft interview questions.
- Propose evidence requests.
- Draft a bounded test.
- Compare Current and Target Profiles.
- Identify monitoring signals.
- Draft an assessor narrative for review.

Inappropriate uses

108. Assessment workflow

End-to-end workflow for creating, evidencing, reviewing and maintaining a system-specific NIST AI RMF Current and Target Profile.

1. Select the system

Choose the registered AI system that is being assessed. Confirm its name, version, purpose, lifecycle stage and deployment context.

Do not begin from an unbounded enterprise assumption when the intended conclusion is system-specific.

2. Review the intake

Confirm the intake covers:

- Purpose, users and affected people.
- Data and model sources.
- Suppliers and deployment.
- Human oversight and autonomy.
- Public-facing or consequential use.
- Relevant jurisdictions and obligations.
- Retrieval, tools and integrations.
- Known risks, limitations and incidents.

Incomplete context should be recorded as uncertainty, not converted into a favourable outcome.

3. Confirm the active Profile lenses

The NIST AI RMF Core is the primary baseline.

For generative AI, also use the Generative AI Profile. Other sector or use-case Profiles may inform tailoring where appropriate and authoritative.

4. Review priorities

Use Baseline, Priority and Enhanced labels to sequence work. Confirm that the reasons match the system facts.

Priority affects assessment effort; it does not decide the outcome.

5. Work function by function

A common sequence is:

1. **GOVERN** - confirm policy, roles, accountability, culture and supplier governance.
2. **MAP** - establish context, purpose, system boundaries, actors, impacts, benefits and risks.
3. **MEASURE** - inspect metrics, evaluations, tests, monitoring and stakeholder feedback.
4. **MANAGE** - review treatment, proceed decisions, incidents, recovery and continual improvement.

The process is iterative. A MEASURE result may reveal new MAP risks. A MANAGE decision may require new governance or measurement.

6. Read the assessor advisory

For each outcome, use the on-screen guidance to:

- Understand the outcome being assessed.
- Ask accountable owners targeted questions.
- Identify appropriate evidence.
- Design a bounded test or review.
- Understand pass criteria.
- Recognise common failure patterns.
- Define monitoring and reassessment.

The advisory supports judgement. The official NIST publication remains authoritative.

7. Record the Current Profile

Choose:

- Not assessed.
- Not achieved.
- Partially achieved.
- Achieved.

The label must reflect the selected system, not merely an organisation-wide policy.

8. Record assurance depth

Separately state whether the conclusion is:

- Unverified.
- Documented.
- Implemented.
- Assured.

Outcome and assurance answer different questions:

Outcome: what is the current position?

Assurance: how strongly is that position supported?

9. Set the Target Profile

Choose the intended target for the outcome. Consider:

- Risk tolerance.
- Intended use and impact.
- Legal and contractual duties.
- Trustworthiness characteristics.
- Supplier and operating constraints.
- Stakeholder needs.
- Proportionality and available resources.

An Achieved target does not mean every Playbook suggestion must be implemented. It means the organisation intends to demonstrate the outcome appropriately for this context.

10. Record rationale

The note should explain:

- The system-specific facts.
- What exists today.
- What is missing.
- Evidence inspected.
- Test result.
- Owner.
- Treatment or accepted limitation.
- Target date or review trigger.

For Not achieved or Partially achieved outcomes, document the gap and treatment direction.

11. Link evidence

Attach or reference:

- Approved policies and procedures.
- System cards and architecture records.
- Risk and impact assessments.
- Data and model documentation.
- Supplier records.
- Evaluation and monitoring results.
- Training and oversight records.
- Incident, appeal and change records.

See Evidence, testing and findings.

12. Test effectiveness

Define:

- Objective.
- Procedure.

- Sample and environment.
- Expected result.
- Pass criteria.
- Safety limits.
- Actual result.
- Exceptions.
- Reviewer and date.

Tests should be authorised, proportionate and non-destructive.

13. Raise findings

Create a finding when:

- The Current outcome is unsupported.
- Evidence is missing, stale or rejected.
- A test fails.
- Practice differs from policy.
- Ownership is unclear.
- A supplier dependency is unresolved.
- The Target gap creates material exposure.

14. Review the Profile as a whole

Check for:

- Unassessed outcomes.
- Optimistic Achieved claims.
- Weak assurance.
- Contradictions between outcomes.
- Open adverse findings.
- Failed tests.
- Missing supplier or stakeholder evidence.
- Inconsistent targets.
- Missing owners or dates.

15. Write the conclusion

The conclusion should state:

- System and boundary.
- Current Profile position.
- Material strengths.
- Material gaps.
- Evidence and testing limitations.
- Residual risk.
- Treatment priorities.
- Confidence.
- Review date.

- Reassessment triggers.
- Decision owner.

Avoid saying "NIST compliant" or "NIST certified."

16. Confirm and maintain

Confirmation is an accountable human statement about the recorded Profile at that time. It is not permanent. Review it after a material change or by the scheduled date.

Completion checklist

109. Core, Profiles and Playbook

How the NIST AI RMF Core, Current Profile, Target Profile and voluntary Playbook fit together in a Gamut assessment.

The NIST AI RMF is designed for flexible, risk-based use. Its parts serve different purposes and should not be collapsed into a single checklist or maturity score.

The Framework

NIST AI 100-1 provides:

- A framing of AI risk and trustworthiness.
- Intended audiences and AI actors.
- The four-function AI RMF Core.
- The concept of Profiles for tailoring the framework.

The framework is voluntary and outcome-oriented. Organisations apply it according to their role, resources, risk tolerance, use case and operating context.

The Core

The Core contains:

- **4 functions**
- **19 categories**
- **72 subcategories**

The subcategories express risk-management outcomes. They do not constitute an ordered checklist, and NIST does not assign a universal implementation score to them.

Function	Relationship
GOVERN	Cross-cutting conditions, accountability and culture that inform the other functions
MAP	Context and risk identification
MEASURE	Analysis, evaluation, testing and monitoring
MANAGE	Prioritisation, treatment, response, recovery and communication

Profiles

A Profile tailors the Core to a particular application, sector, technology, organisation or system. Profiles help describe what is relevant and how outcomes should be prioritised.

Current Profile

The Current Profile describes what is presently achieved for the selected system and context.

It should answer:

- What practice exists now?

- Is it specific to this system?
- Who owns it?
- Is it documented?
- Is it operating?
- Has effectiveness been tested?
- What evidence or adverse finding affects the conclusion?

Target Profile

The Target Profile describes the desired risk-management position.

It should reflect:

- Intended purpose and deployment context.
- Organisational risk tolerance.
- Legal, contractual and policy requirements.
- Potential impacts on people, organisations, society and the environment.
- Technical and operational dependencies.
- Lifecycle stage.
- Resource and implementation constraints.
- Relevant stakeholder expectations.

Profile gap

The difference between Current and Target Profiles becomes a prioritised improvement plan. A gap may require:

- A new or strengthened practice.
- Better ownership or documentation.
- Additional evidence.
- A test or evaluation.
- Supplier action.
- A deployment restriction.
- Risk treatment or acceptance.
- Monitoring and reassessment.

How Gamut presents Profile outcomes

Gamut uses plain assessment language:

Current outcome	Meaning
Not assessed	No defensible determination has been made.
Not achieved	The outcome is absent or materially ineffective for this system.
Partially achieved	Some elements operate, but material gaps remain.
Achieved	The outcome operates for this system and is sufficiently supported.

The Target Profile uses the substantive target outcomes: Not achieved, Partially achieved or Achieved.

These are **Gamut assessment labels**, not a NIST maturity scale.

The Playbook

The NIST AI RMF Playbook provides suggested actions aligned to Core subcategories. NIST describes it as a voluntary companion resource, not a checklist or mandatory sequence.

Use the Playbook to:

- Explore possible implementation actions.
- Identify documentation and evidence ideas.
- Tailor practices to role, sector and use case.
- Compare alternative ways to achieve an outcome.
- Support workshops and improvement planning.

Do not:

- Treat every suggestion as mandatory.
- Assume one suggestion is sufficient in every context.
- Copy a suggested action without linking it to the selected system.
- Represent a Playbook action as evidence that the outcome is achieved.

The Generative AI Profile

NIST AI 600-1 is a cross-sectoral companion Profile for generative AI. It identifies generative-AI risk considerations and recommended actions that supplement the AI RMF Core.

When a selected system has generative-AI characteristics, Gamut highlights the Generative AI Profile as an additional lens. The Core remains active; the companion Profile does not replace it.

See Generative AI Profile.

Tailoring without losing accountability

Tailoring should be explicit. Record:

- The system and context.
- The outcome being considered.
- The reason for its priority.
- The intended target.
- Any decision to defer work.
- The accountable decision-maker.
- The trigger for revisiting the decision.

Tailoring does not mean:

- Quietly removing inconvenient outcomes.
- Treating low priority as not applicable.
- Reducing evidence because a system is familiar.
- Using another system's assessment.
- Ignoring an outcome because a mapped framework appears stronger.

Frequently confused concepts

Concept	It is	It is not
Core outcome	A NIST risk-management outcome	A prescriptive technical control
Current Profile	Assessed present position	A general organisational aspiration
Target Profile	Intended future position	Proof of implementation
Playbook action	A voluntary implementation suggestion	A mandatory checklist item
Gamut priority	Assessment-planning emphasis	A NIST severity rating
Assurance depth	Strength of support for a claim	The Current outcome itself
Crosswalk	A traceability aid	Automatic equivalence

Official reading

110. Evidence, testing and findings

Practical guidance for supporting NIST AI RMF Profile outcomes with system-specific evidence, tests, findings and monitoring.

NIST AI RMF outcomes are outcome-oriented. A defensible assessment therefore needs more than a policy statement or a Yes answer. It needs evidence that the relevant practice exists and works for the selected system.

Evidence hierarchy

1. Assertion

An owner states that the practice exists.

Useful for discovery, but unverified.

2. Design evidence

Shows what should happen:

- Policy.
- Procedure.
- Architecture.
- Standard.
- Role description.
- Test plan.

3. Implementation evidence

Shows the practice is configured or established:

- Approved system record.
- Training completion.
- Supplier assessment.
- Monitoring configuration.
- Access or oversight workflow.
- Release approval.

4. Operating evidence

Shows what actually happened:

- Logs and alerts.
- Completed reviews.
- Test results.
- Incident records.
- Appeal and override records.
- Monitoring trends.
- Change and decommissioning records.

5. Independent assurance

Adds challenge or validation:

- Independent review.
- Internal audit.
- External assessment.

- Reperformance.
- Red-team or resilience exercise.

Evidence quality questions

Ask whether evidence is:

- **Relevant** - does it address this outcome?
- **Scoped** - does it belong to this system?
- **Current** - does it reflect the assessed configuration and period?
- **Authentic** - is provenance clear?
- **Complete** - does it cover the material system boundary?
- **Approved** - has an authorised reviewer accepted it?
- **Consistent** - does it agree with tests, findings and other records?

Common evidence by function

GOVERN

- AI policy and risk-tolerance statements.
- Inventory and ownership records.
- Role and escalation matrices.
- Training and competence records.
- Stakeholder-engagement procedures.
- Supplier-governance standards.
- Review and decommissioning procedures.

MAP

- System and model cards.
- Intended-use and prohibited-use statements.
- Architecture and data-flow diagrams.
- Impact and risk assessments.
- Stakeholder maps.
- Benefits, costs and alternatives analysis.
- System limitations and assumptions.

MEASURE

- TEVV plans and reports.
- Metric definitions and thresholds.
- Evaluation datasets and representativeness analysis.
- Safety, security, privacy and fairness evaluations.
- Explainability and interpretability assessments.
- Production-monitoring results.
- Drift and incident trends.

MANAGE

- Risk-treatment plans.
- Proceed, restrict, stop or decommission decisions.
- Risk acceptance.

- Incident and recovery records.
- Supplier contingencies.
- Change records.
- Continual-improvement actions.

Evidence requests

An evidence request should state:

- Outcome ID.
- Selected system.
- Requested artefact or operating record.
- Why it is needed.
- Owner.
- Due date.
- Acceptance criteria.
- Review status.

Avoid "provide evidence of compliance." Ask for the exact record needed.

Example:

> For `ME-2.4`, provide the current production-monitoring specification, enabled alert rules, three months of monitoring results, threshold-change history and evidence of response to one material alert for the selected system.

Testing

Testing determines whether an implemented practice is effective.

Test record

Include:

- Objective.
- System and outcome.
- Preconditions.
- Environment and sample.
- Procedure.
- Expected result.
- Pass criteria.
- Safety limits.
- Actual result.
- Exceptions.
- Reviewer and date.

Safe testing principles

- Obtain authorisation.
- Prefer non-production or controlled environments.
- Use synthetic or minimised data where possible.
- Define stop conditions.
- Avoid uncontrolled harmful outputs or actions.
- Protect affected people.

- Preserve evidence.
- Record deviations.
- Retest after remediation.

Example tests

GV-1.6 - inventory

1. Select a sample of deployed AI services.
2. Trace each to the inventory.
3. Verify owner, purpose, version, lifecycle and review dates.
4. Search procurement and technical records for unregistered AI.
5. Pass only if the inventory is materially complete and current.

MP-2.2 - knowledge limits and oversight

1. Review documented limitations.
2. Present representative edge cases.
3. Confirm operators recognise uncertainty.
4. Verify escalation or override.
5. Pass only if limitations are accurate and oversight works in practice.

ME-2.7 - security and resilience

1. Select authorised threat scenarios.
2. Test boundary controls and monitoring.
3. Verify safe failure and recovery.
4. Confirm alerts are attributable and actionable.
5. Record bypasses as findings.

MG-2.4 - disengage or deactivate

1. Run a tabletop or controlled stop exercise.
2. Verify authority, communication and dependencies.
3. Confirm the system can be disabled safely.
4. Confirm fallback and recovery objectives.
5. Pass only if the process works within the documented objective.

Findings

Raise a finding when:

- Current outcome is overstated.
- Evidence is missing or rejected.
- A test fails.
- Exceptions are uncontrolled.
- Ownership is unclear.
- Scope excludes a material component.
- Supplier dependency is unresolved.
- Monitoring does not detect material risk.
- The Target Profile has no credible treatment path.

Finding content

Record:

- Outcome ID and system.
- Condition observed.
- Expected outcome.
- Evidence and test basis.
- Risk and affected parties.
- Root cause where known.
- Immediate containment.
- Recommendation.
- Owner and target date.
- Status and closure evidence.

Adverse evidence

Positive narrative must not override:

- Failed tests.
- Rejected or expired evidence.
- Open material findings.
- Incidents.
- Complaints or appeals.
- Known model or supplier limitations.
- Monitoring that contradicts the claimed outcome.

Reuse and crosswalks

Evidence may support multiple frameworks when it genuinely addresses each objective. Reuse should preserve:

- Original provenance.
- System scope.
- Review date.
- Mapping rationale.
- Framework-specific interpretation.

A crosswalk is not evidence by itself.

Outcome-to-evidence checklist

111. Functions and outcome catalogue

Reference catalogue for the 4 NIST AI RMF functions, 19 categories and 72 system-assessment outcomes presented in Gamut.

This page is a navigation aid for the NIST AI RMF Core as presented in Gamut. The short titles are product-safe summaries. Consult the official AI RMF Core for authoritative wording and context.

Catalogue totals

Function	Categories	Outcomes
GOVERN (GV)	6	19
MAP (MP)	5	18

Function	Categories	Outcomes
MEASURE (ME)	4	22
MANAGE (MG)	4	13
Total	**19**	**72**

GOVERN

GOVERN establishes the organisation-wide conditions that support risk management throughout the AI lifecycle. It is cross-cutting and should continue as context, expectations and risks change.

GV-1 - Governance policies and practices

Focus: transparent policies, controls, risk tolerance, monitoring, inventory and decommissioning.

ID	Assessment focus
GV-1.1	Legal and regulatory requirements
GV-1.2	Trustworthy AI characteristics
GV-1.3	Risk-management activity level
GV-1.4	Risk-management controls
GV-1.5	Monitoring and review cadence
GV-1.6	AI system inventory
GV-1.7	Safe decommissioning

GV-2 - Accountability structures

Focus: roles, competence, communication and executive responsibility.

ID	Assessment focus
GV-2.1	Roles and communication lines
GV-2.2	AI risk-management training
GV-2.3	Executive responsibility

GV-3 - Workforce diversity and human-AI oversight

Focus: diverse decision-making and clear human-AI roles.

ID	Assessment focus
GV-3.1	Diverse AI risk decision-making
GV-3.2	Human-AI roles and oversight

GV-4 - Organisational risk culture

Focus: critical thinking, safety-first behaviour, risk communication, testing and incident learning.

ID	Assessment focus
GV-4.1	Critical thinking and safety-first mindset
GV-4.2	Risk and impact communication
GV-4.3	Testing, incident identification and sharing

GV-5 - Stakeholder engagement

Focus: obtaining, adjudicating and integrating external and relevant AI-actor feedback.

ID	Assessment focus
GV-5.1	External impact feedback
GV-5.2	Adjudicated feedback integration

GV-6 - Third-party and supply-chain risk

Focus: supplier risks, third-party rights and contingency planning.

ID	Assessment focus
GV-6.1	Third-party AI risk policies
GV-6.2	High-risk third-party contingency

MAP

MAP establishes the system's context and identifies benefits, costs, affected parties, limitations, impacts and risks. Weak MAP work undermines later measurement and treatment.

MP-1 - Context established

Focus: intended purpose, deployment context, mission, risk tolerance and socio-technical requirements.

ID	Assessment focus
MP-1.1	Intended purpose and deployment context
MP-1.2	Interdisciplinary context input
MP-1.3	Mission and AI goals
MP-1.4	Business value or use context
MP-1.5	Organisational risk tolerance
MP-1.6	System requirements and socio-technical implications

MP-2 - AI system categorisation

Focus: tasks, methods, knowledge limits, human oversight and scientific integrity.

ID	Assessment focus
MP-2.1	Tasks and methods
MP-2.2	Knowledge limits and human oversight
MP-2.3	Scientific integrity and TEVV considerations

MP-3 - Capabilities, usage, benefits and costs

Focus: benefits, costs, application scope, competence and oversight.

ID	Assessment focus
MP-3.1	Potential benefits
MP-3.2	Potential costs
MP-3.3	Targeted application scope
MP-3.4	Operator and practitioner proficiency
MP-3.5	Human oversight processes

MP-4 - Risk and benefit mapping

Focus: technology, legal, component and third-party risks and internal controls.

ID	Assessment focus
MP-4.1	Technology and legal risk mapping
MP-4.2	Internal risk controls

MP-5 - Impact characterisation

Focus: likelihood and magnitude of impacts and continuing engagement with relevant actors.

ID	Assessment focus
MP-5.1	Likelihood and magnitude of impacts
MP-5.2	Regular engagement on impacts

MEASURE

MEASURE uses quantitative, qualitative or mixed methods to evaluate risks and trustworthy characteristics, monitor operation and validate whether the measurement approach remains useful.

ME-1 - Measurement methods

Focus: risk metrics, control effectiveness and appropriate assessment participation.

ID	Assessment focus
ME-1.1	Risk-measurement approaches
ME-1.2	Metric and control effectiveness
ME-1.3	Independent and stakeholder-supported assessment

ME-2 - Trustworthy AI evaluation

Focus: TEVV, performance, monitoring, safety, security, transparency, explainability, privacy, fairness and environmental impact.

ID	Assessment focus
ME-2.1	TEVV artefact documentation
ME-2.2	Human-subject evaluation requirements
ME-2.3	Deployment-like performance criteria
ME-2.4	Production monitoring
ME-2.5	Valid and reliable demonstration
ME-2.6	Safety evaluation
ME-2.7	Security and resilience evaluation
ME-2.8	Transparency and accountability risks
ME-2.9	Explainability and interpretability
ME-2.10	Privacy risk examination
ME-2.11	Fairness and bias evaluation
ME-2.12	Environmental impact and sustainability
ME-2.13	TEVV effectiveness

ME-3 - Risk tracking

Focus: existing, unknown and emergent risks, including feedback and appeal.

ID	Assessment focus
ME-3.1	Existing and emergent risk tracking
ME-3.2	Tracking where metrics are immature
ME-3.3	End-user and impacted-community feedback

ME-4 - Measurement feedback

Focus: connecting measurement to context and validating whether results and trends remain useful.

ID	Assessment focus
ME-4.1	Context-connected measurement approaches
ME-4.2	Trustworthiness-results validation
ME-4.3	Performance-change tracking

MANAGE

MANAGE turns mapped and measured risk into decisions, treatment, resource allocation, response, recovery, communication and continual improvement.

MG-1 - Risk response

Focus: proceed decisions, prioritisation, treatment and residual-risk communication.

ID	Assessment focus
MG-1.1	Proceed decision
MG-1.2	Risk-treatment prioritisation
MG-1.3	High-priority risk responses
MG-1.4	Residual-risk documentation

MG-2 - Benefit maximisation and impact reduction

Focus: resources, non-AI alternatives, sustained value, unknown risks and deactivation.

ID	Assessment focus
MG-2.1	Risk resources and non-AI alternatives
MG-2.2	Sustaining deployed AI value
MG-2.3	Previously unknown risk response
MG-2.4	Supersede, disengage or deactivate

MG-3 - Third-party risk management

Focus: ongoing monitoring and control of third-party resources and pre-trained models.

ID	Assessment focus
MG-3.1	Third-party monitoring and controls
MG-3.2	Pre-trained model monitoring

MG-4 - Treatment monitoring and incident communication

Focus: post-deployment monitoring, appeal, override, incident response, recovery, change and continual improvement.

ID	Assessment focus
MG-4.1	Post-deployment monitoring plans
MG-4.2	Continual improvement in updates
MG-4.3	Incident and error communication

How to use the catalogue

For each outcome:

1. Read the official Core wording.
2. Confirm the selected system and context.
3. Use the Gamut advisory to plan questions, evidence and testing.
4. Record Current and Target Profile outcomes.
5. Record assurance depth separately.
6. Link evidence, tests and findings.
7. Define monitoring and reassessment.

> **Catalogue titles are navigation aids:** The short titles on this page are not substitutes for the official NIST text. Do not quote them as if they were verbatim NIST requirements.

112. Generative AI Profile

How to use NIST AI 600-1 with the NIST AI RMF Core for generative AI systems in Gamut.

NIST AI 600-1, published in July 2024, is a cross-sectoral companion Profile for generative artificial intelligence. It helps organisations identify generative-AI risks and select actions that fit their goals and priorities.

When to use it

Use the Profile when the system generates or materially transforms:

- Text.
- Images.
- Audio.
- Video.
- Code.
- Synthetic data.
- Multimodal content.

It is also relevant to applications built on foundation models, chatbots, copilots and other generative services.

Relationship to the Core

The Generative AI Profile:

- Supplements the AI RMF Core.
- Does not replace GOVERN, MAP, MEASURE or MANAGE.
- Does not create NIST certification.
- Should be tailored to the selected system.
- Should be revisited as models, data, suppliers and capabilities change.

The 12 GAI risk families

Gamut highlights the risk families identified in NIST AI 600-1:

Risk family	Assessment questions
CBRN information or capabilities	Could the system meaningfully facilitate harmful chemical, biological, radiological or nuclear knowledge or capability?
Confabulation	Can plausible but false output cause decisions, harm or loss of trust?
Dangerous, violent or hateful content	Can the system generate, amplify or enable harmful content?
Data privacy	Can training, prompts, retrieval, memory or outputs expose or infer personal information?
Environmental impacts	Are training, inference and infrastructure impacts understood and managed?
Harmful bias and homogenisation	Can outputs disadvantage groups, erase variation or reinforce harmful patterns?
Human-AI configuration	Are reliance, automation bias, anthropomorphism and human roles appropriately managed?
Information integrity	Can the system create or amplify misleading, manipulated or unverifiable information?
Information security	Can prompts, models, tools, data, outputs or integrations be attacked or misused?
Intellectual property	Are rights, licences, provenance and output-use risks understood?
Obscene, degrading or abusive content	Can the system generate or facilitate abusive or exploitative material?
Value chain and component integration	Are model, data, platform, tool and downstream responsibilities and dependencies controlled?

Assessing a GAI risk family

For each relevant family:

1. Define the credible scenario.
2. Identify affected people, assets and decisions.
3. Identify model, data, prompt, retrieval, tool and supplier contributors.
4. Review preventive and detective controls.
5. Select evaluation methods.
6. Define pass criteria and limitations.
7. Record incidents, near misses and external evidence.
8. Determine treatment and monitoring.
9. Map the work back to relevant Core outcomes.

Common generative-AI evidence

- Model and system cards.
- Supplier documentation and change notices.
- Data provenance and rights records.
- Prompt and retrieval architecture.
- Red-team and misuse testing.
- Hallucination or factuality evaluation.
- Safety-filter and policy testing.
- Privacy testing.
- Bias and representation evaluation.
- Security testing.
- Human-oversight procedures.
- Content labelling and provenance mechanisms.
- Incident, complaint and takedown records.
- Monitoring by use case, user group and model version.

Important evaluation dimensions

Context

Evaluate the actual use case. The same model may have very different risk when used for creative drafting, medical support, recruitment or autonomous tool use.

Capability

Consider what the system can do with:

- Long context.
- Retrieval.
- Tools.
- Code execution.
- Memory.
- Multimodal input.
- Fine-tuning.
- External data.

Access

Consider who can use it, which information it can reach and whether outputs can trigger action.

Scale

Consider volume, speed, user reach, automation and the ability to replicate harmful output.

Change

Model updates, system prompts, retrieval sources, safety settings and suppliers can change risk without changing the product name.

Confabulation example

For a customer-support assistant:

- MAP identifies high-impact topics, escalation rules and knowledge limits.
- MEASURE evaluates factuality using representative enquiries and current source material.
- MANAGE requires abstention, source presentation, human escalation and incident response.
- GOVERN assigns ownership and review cadence.

An acceptable average accuracy does not excuse a dangerous error class. Evaluate the severity and distribution of errors, not only the mean.

Supplier responsibility

Using a third-party model does not transfer all responsibility. Record:

- Supplier responsibilities.
- Application-provider responsibilities.
- Deploying-organisation responsibilities.
- Customer or user responsibilities.
- Evidence dependencies.
- Escalation and contingency arrangements.

Monitoring triggers

Reassess after:

- New foundation model or model version.
- Fine-tuning or prompt change.
- New retrieval source.
- New tool or action permission.
- New user population.
- New high-impact use.
- Safety-control change.
- Material incident or misuse.
- New external research on capability or risk.

Critical-infrastructure note

NIST announced a critical-infrastructure Profile concept note in April 2026. A concept note should not be described as a final Profile. Check the official AI RMF page for current status.

113. Outcomes, assurance and conclusions

Explain Gamut's NIST AI RMF Current outcome labels, Target Profile, assurance depth, confirmation and defensible conclusion language.

Gamut separates **what outcome is achieved** from **how strongly that conclusion is supported**. This prevents a documented policy from being mistaken for operating effectiveness.

Current Profile outcomes

Not assessed

No defensible determination has been made.

Use when:

- The outcome has not been reviewed.
- Required context is missing.
- Ownership is unresolved.
- Evidence is unavailable.

Do not use it as a neutral or low-risk answer.

Not achieved

The outcome is absent or materially ineffective for the selected system.

Examples:

- No responsible owner exists.
- Required practice is not implemented.
- A control is routinely bypassed.
- A test shows the practice fails its objective.

Partially achieved

Some elements operate, but material gaps remain.

Examples:

- Policy and ownership exist, but execution is inconsistent.
- The practice covers development but not production.
- Evidence covers only part of the system boundary.
- Monitoring exists but lacks response thresholds.
- A supplier process is incomplete.

Achieved

The outcome operates for the selected system and is sufficiently supported.

An Achieved label should be consistent with:

- Clear ownership.
- Appropriate design.
- Current implementation.
- System-specific operating evidence.
- Effective testing or evaluation where relevant.
- No unresolved adverse finding that contradicts the claim.

Assurance depth

Depth	Label	Meaning
0	**Unverified**	Assertion only; no system-specific evidence has been reviewed.
1	**Documented**	Design documentation and accountable ownership have been reviewed.
2	**Implemented**	System-specific operating evidence demonstrates implementation.
3	**Assured**	Operating evidence, effective testing and adverse-finding review support the conclusion.

Assurance depth does not replace the outcome.

Examples:

Current outcome	Depth	Interpretation
Partially achieved	2	The assessor has evidence that the partial implementation operates as described.
Achieved	1	The outcome is asserted from design documentation but is not yet strongly assured.
Not achieved	3	Strong evidence and testing confirm a real gap.
Achieved	3	The strongest supported positive conclusion.

Target Profile outcome

The Target Profile records the intended future state:

- Not achieved.
- Partially achieved.
- Achieved.

Most material risk-management outcomes will target Achieved, but the target should still be contextual and accountable. A lower target requires a clear reason, risk decision and review trigger.

Gamut priority labels

Baseline, Priority and Enhanced describe assessment emphasis. They do not alter Current or Target outcomes and are not NIST terms.

What confirmation means

Confirmation means an authorised human has reviewed the system-specific Profile and conclusion.

A confirmable assessment should:

- Consider all 72 Core outcomes.
- Set a Current and Target outcome.
- Explain material gaps.
- Support Achieved claims with strong assurance.
- Reflect failed tests and open findings.
- Identify owner, confidence, residual risk, review date and triggers.

Confirmation does not mean:

- NIST certification.
- Legal compliance.
- Risk elimination.
- Permanent approval.
- External audit opinion.

Conclusion fields

Assessment owner

The accountable person responsible for the conclusion and follow-up.

Confidence

How reliable and complete the assessment basis is:

- Low.
- Medium.
- High.

Confidence does not change the Current Profile. It explains uncertainty.

Residual risk

The remaining risk after current practices, limitations, evidence and treatment are considered.

Gamut uses:

- Low.
- Moderate.
- Elevated.
- High.
- Critical.

Residual risk is not the same as an outcome completion percentage.

Review due

The date by which the Profile should be reviewed even if no trigger occurs.

Reassessment triggers

Material events that require earlier review, such as:

- Model or supplier change.
- New data.
- Increased autonomy.
- New geography or user group.
- Incident or complaint.
- Performance drift.
- Change to purpose or human oversight.

Assessor conclusion

A concise narrative connecting:

- Scope.
- Current and Target Profiles.
- Material evidence.
- Important gaps.
- Residual risk.
- Treatment.
- Constraints.
- Monitoring.
- Decision.

Safe conclusion patterns

Strong position

> The selected system has a substantially achieved Current Profile across the assessed NIST AI RMF Core, with the stated positive outcomes supported by current operating evidence and testing. Remaining gaps, residual risk and reassessment triggers are recorded below. This is the organisation's risk-management conclusion, not NIST certification.

Partial position

> The Current Profile is partially achieved. Governance foundations are documented, but material gaps remain in system-specific measurement, supplier assurance and post-deployment monitoring. The Target Profile and treatment plan identify the required improvements before broader use.

Insufficient position

> A reliable Current Profile cannot yet be concluded because system context, operating evidence and test results are incomplete. The assessment remains open and no positive alignment claim should be made until the identified evidence requests are resolved.

Language to avoid

114. Reference and glossary

Quick reference for NIST AI RMF functions, Profiles, Gamut outcome and assurance labels, evidence terms and safe explanatory language.

Quick reference

Item	Value
Framework	NIST Artificial Intelligence Risk Management Framework
Published baseline	AI RMF 1.0, NIST AI 100-1
Published	26 January 2023
Nature	Voluntary
Functions	GOVERN, MAP, MEASURE, MANAGE
Categories	19
Core outcomes	72
Companion implementation resource	NIST AI RMF Playbook
Generative AI companion	NIST AI 600-1
Gamut assessment scope	One selected AI system
Assessment structure	Current Profile and Target Profile

AI RMF Core

The set of functions, categories and subcategories that provides outcomes for managing AI risk.

Function

A high-level group of AI risk-management activity:

- GOVERN.
- MAP.
- MEASURE.
- MANAGE.

Category

A group of related outcomes within a function.

Subcategory

A more specific Core outcome. Gamut uses stable short IDs such as `GV-1.1`, `MP-2.2`, `ME-2.7` and `MG-4.1` for navigation.

Profile

A tailoring of the AI RMF Core to a use case, sector, technology, organisation or system.

Current Profile

The assessed present position for the selected system.

Target Profile

The intended future risk-management position.

Profile gap

The difference between Current and Target Profiles. It should drive treatment, evidence, testing, ownership and monitoring.

Playbook

NIST's voluntary companion resource containing suggested actions related to Core outcomes. It is not a mandatory checklist.

Trustworthiness characteristics

- Valid and reliable.
- Safe.
- Secure and resilient.
- Accountable and transparent.
- Explainable and interpretable.
- Privacy-enhanced.
- Fair, with harmful bias managed.

TEVV

Testing, evaluation, verification and validation.

AI actor

A person or organisation performing a role in the AI lifecycle, such as developer, deployer, operator, evaluator, affected individual or supplier.

Socio-technical

The combined interaction of technology, people, organisations, processes and context.

Risk tolerance

The level and type of risk an organisation is prepared to accept in pursuit of objectives.

Current outcome labels

Not assessed

No defensible determination has been made.

Not achieved

The outcome is absent or materially ineffective.

Partially achieved

Some elements operate, but material gaps remain.

Achieved

The outcome operates for the selected system and is sufficiently supported.

These are Gamut assessment labels, not NIST maturity levels.

Assurance depth

Unverified / 0

Assertion only.

Documented / 1

Design documentation and ownership reviewed.

Implemented / 2

System-specific operating evidence supports implementation.

Assured / 3

Operating evidence, effective testing and adverse-finding review support the conclusion.

Priority labels

Baseline

Foundational assessment depth retained in the Profile.

Priority

System facts make the outcome particularly relevant.

Enhanced

The context calls for deeper system-specific evidence, testing or review.

These are Gamut planning labels, not NIST severity ratings.

Accepted evidence

Evidence that is relevant, scoped, current, authentic, sufficiently complete and reviewed.

Evidence request

A request for a specific artefact or operating record needed to support an outcome.

Control test

A bounded procedure with objective, expected result, pass criteria, safety limits and actual result.

Finding

A documented gap, failed test, unsupported assumption, exception or adverse condition requiring action or risk decision.

Adverse evidence

Information that contradicts or weakens a positive conclusion, such as a failed test, open finding, incident or rejected evidence.

Residual risk

Risk remaining after current practices, evidence, limitations and treatment are considered.

Reassessment trigger

A change or event requiring review before the normal review date.

Generative AI Profile

NIST AI 600-1, a cross-sectoral companion Profile for generative AI.

The 12 GAI risk families

1. CBRN information or capabilities.
2. Confabulation.
3. Dangerous, violent or hateful content.
4. Data privacy.
5. Environmental impacts.
6. Harmful bias and homogenisation.
7. Human-AI configuration.
8. Information integrity.
9. Information security.
10. Intellectual property.
11. Obscene, degrading or abusive content.
12. Value-chain and component integration.

Crosswalk

A traceability reference between related framework concepts. It is not automatic equivalence.

Confirmation

An accountable human sign-off on the recorded Profile at a point in time. It is not NIST certification.

Frequently asked questions

Is the NIST AI RMF mandatory?

The framework itself is voluntary. Other law, regulation, contract or policy may independently require risk-management activity.

Does NIST certify AI systems against the AI RMF?

The Gamut assessment must not be represented as NIST certification or endorsement.

Is the AI RMF a checklist?

No. The Core is outcome-oriented, and the Playbook is a voluntary set of suggestions.

Does every outcome need the same effort?

No. Tailor effort to context and risk, while explicitly considering the Core and documenting the reason for priority.

Is Achieved the same as Assured?

No. Achieved describes the outcome. Assured describes the strength of support.

Can an outcome be Not achieved at Assured depth?

Yes. Strong evidence and testing can confirm a real gap.

Can a documented policy justify Achieved?

Not by itself. System-specific implementation and effectiveness matter.

Does a mapped GTSAF or ISO control prove a NIST outcome?

No. It may provide relevant evidence or implementation support, but the NIST outcome still needs direct assessment.

Does AI assistance complete the Profile?

No. It provides advisory analysis for human review.

Does an API key grant access?

No. Access remains subject to the user's authorised workspace, role, plan and feature availability.

What happens when the system changes?

Review the Profile, evidence, tests, findings, priorities and conclusion. Material change may require a new assessment basis.

Is AI RMF 1.0 still current?

It is the published baseline described here. NIST states that a revision is in progress. Check the official AI RMF page for current status.

Safe one-sentence explanation

> Gamut applies the voluntary NIST AI RMF 1.0 Core to a named AI system through Current and Target Profiles, separates outcome from assurance strength, links evidence, testing and findings, adds the NIST Generative AI Profile when relevant, and keeps final risk decisions with accountable humans.

Official links

115. Reporting and governance

How to interpret NIST AI RMF system and portfolio reporting, write accountable conclusions and govern ongoing reassessment.

NIST AI RMF reporting should communicate the selected system's Profile, evidence position, gaps and decisions without implying NIST certification or universal trustworthiness.

System report

A system report should identify:

- System and assessed boundary.
- Framework publication used.
- Active companion Profiles.
- Current and Target outcomes.
- Outcome coverage.
- Assurance depth.
- Evidence and testing position.
- Findings.
- Material strengths and gaps.
- Residual risk.
- Treatment priorities.
- Owner, review date and reassessment triggers.

Function summaries

Gamut may summarise the proportion of outcomes marked Achieved in:

- GOVERN.
- MAP.
- MEASURE.
- MANAGE.

This percentage is a Gamut reporting summary. It is:

- Not a NIST maturity level.
- Not a compliance percentage.
- Not evidence quality.

- Not a certification score.

  Read it with assurance depth, evidence, tests and findings.

Portfolio reporting

A portfolio view can show:

- Systems assessed.
- Outcome coverage.
- Common gaps.
- Recurring evidence needs.
- High-priority or Enhanced areas.
- Systems requiring review.
- Generative-AI exposure.
- Upcoming review dates.

Portfolio reporting does not replace system-level Profiles. Averaging systems can conceal a severe gap in one consequential system.

Cross-framework reporting

NIST outcomes may be cross-referenced with:

- GTSAF.
- EU AI Act.
- MAESTRO.
- Agentic Trust Framework.
- ISO/IEC 42001.
- ISO/IEC 42005.

Mappings support:

- Evidence discovery.
- Duplicate-work reduction.
- Integrated remediation.
- Common ownership.
- Multi-framework reporting.

Mappings do not prove equivalence. Validate each framework's own objective, scope and evidence.

Governance decisions

The Profile should support decisions such as:

- Proceed.
- Proceed with conditions.
- Restrict use.
- Require remediation.
- Increase monitoring.
- Obtain independent review.
- Accept defined residual risk.
- Pause.
- Supersede.

- Disengage or deactivate.

The decision should name:

- Authority.
- Conditions.
- Evidence basis.
- Residual risk.
- Review date.
- Trigger for escalation or reversal.

Reporting adverse information

Do not hide:

- Unassessed outcomes.
- Weak assurance.
- Failed tests.
- Rejected or stale evidence.
- Open findings.
- Incidents and complaints.
- Supplier limitations.
- Uncertainty.
- Deferred Target Profile work.

Safe external language

Prefer:

> The organisation has assessed the named system using the NIST AI RMF 1.0 Core and documented system-specific Current and Target Profiles. The report identifies the outcomes reviewed, supporting evidence, assurance limitations, findings and required treatment.

Avoid:

> The system is NIST compliant.

Management review agenda

1. Has the system or context changed?
2. Are material outcomes still achieved?
3. Is assurance current?
4. Did tests or monitoring reveal adverse evidence?
5. Are findings overdue?
6. Has residual risk changed?
7. Are Target Profile actions funded and owned?
8. Are supplier dependencies acceptable?
9. Is generative-AI risk still appropriately assessed?
10. Should deployment conditions change?

Reassessment governance

Review:

- On the scheduled date.
- After a material trigger.
- Before significant expansion.
- After a serious incident.
- After a major model or supplier change.
- When authoritative framework publications change.

Records to retain

- Scope and Profile basis.
- Assessment outcomes and rationale.
- Evidence and test records.
- Findings and remediation.
- Risk decisions.
- Approvals.
- Monitoring and incidents.
- Review history.
- Superseded conclusions.

Retention should follow applicable law, contract and organisational policy.

116. System scope and prioritisation

How a NIST AI RMF assessment is bound to one AI system and how context changes assessment priority without removing Core outcomes.

NIST AI risk management is contextual. A useful assessment must be connected to the system that is being designed, developed, deployed, used or evaluated.

Why system scope matters

Two systems in the same organisation can have very different:

- Purposes and users.
- Affected people.
- Data and privacy risks.
- Models and suppliers.
- Autonomy and human oversight.
- Deployment environments.
- Performance and safety requirements.
- Threat exposure.
- Legal and societal impacts.
- Monitoring and recovery needs.

An enterprise policy may support both systems, but it does not establish that the same outcome is implemented or effective for both.

The linked System Record and Intake establish the current authoritative routing basis. NIST AI RMF remains available as a universal risk-management baseline; those facts influence priority, assurance and reassessment rather than removing Core outcomes.

The scope statement

Before assessing outcomes, record:

- System name and reference.
- Intended purpose and prohibited or out-of-scope uses.
- Current lifecycle stage.
- Deployment environment and geography.
- Users and operators.
- Affected individuals, groups and communities.
- Data sources and classifications.
- Model, service and supplier dependencies.
- Human-AI roles and decision authority.
- Autonomy, tools and reachable systems.
- Relevant laws, contracts and internal policies.
- Known incidents, limitations and assumptions.

If the boundary is unclear, do not compensate with optimistic outcome labels. Resolve or document the uncertainty.

Core availability

Gamut makes the NIST AI RMF Core available for every registered AI system. Context affects **priority and assurance depth**, not whether the Core disappears.

This supports a complete Profile while allowing proportionate assessment effort.

Priority labels

Gamut may present these assessment-planning labels:

Label	Meaning
Baseline	Retain the outcome in the Profile and assess it at a proportionate foundational depth.
Priority	System facts make the outcome particularly relevant and deserving of focused evidence.
Enhanced	The risk context calls for deeper, more system-specific evidence, testing or review.

These labels:

- Help sequence work.
- Explain why some outcomes need deeper assurance.
- Do not change the official NIST outcome.
- Do not prove implementation.
- Do not grant access to another framework or product feature.
- Are not NIST risk ratings.

Typical priority drivers

Personal or sensitive data

Often increases attention to:

- Legal and regulatory understanding.
- Privacy risk.
- Fairness and harmful bias.

- Stakeholder feedback and redress.
- Residual-risk communication.

Consequential decisions

Often increases attention to:

- Executive accountability.
- Human-AI roles.
- Knowledge limits.
- Human oversight.
- Validity, reliability, safety and fairness.
- Proceed, stop and deactivation decisions.

Public-facing use

Often increases attention to:

- Transparency.
- Feedback and complaints.
- Explainability.
- Incident and error communication.
- Monitoring of actual use and misuse.

Third-party or pre-trained components

Often increases attention to:

- Supplier risk.
- Intellectual-property and rights considerations.
- Dependency inventory.
- Contingency arrangements.
- Monitoring of third-party resources.

Agentic or tool-using behavior

Often increases attention to:

- Production monitoring.
- Security and resilience.
- Safe failure.
- Emergent-risk tracking.
- Recovery, disengagement and deactivation.

Retrieval or external information access

Often increases attention to:

- Data and source provenance.
- Information integrity.
- Security evaluation.
- Production monitoring.
- Unknown and emergent risks.

Human-subject evaluation

Directly increases attention to applicable protection, representation and evaluation requirements.

Large-scale model training or compute

Increases attention to environmental impact, sustainability and resource use.

Critical-infrastructure context

Increases attention to safety, resilience, continuity, stop decisions and recovery. NIST released a critical-infrastructure Profile concept note in April 2026; treat a concept note as informative, not as a final published Profile.

Priority is not applicability

Avoid these mistakes:

- "Baseline" does not mean optional.
- "Priority" does not mean non-compliant.
- "Enhanced" does not mean the system is unsafe.
- An outcome with no special trigger may still be important.
- An outcome cannot be marked Achieved merely because it was low priority.

Changing systems

When the selected system changes:

- The active Current and Target Profiles change.
- Outcome rationale changes.
- Linked evidence, tests and findings should change.
- AI-assisted analysis should be regenerated.
- A conclusion for one system must not be presented as the conclusion for another.

Reassessment triggers

Review the scope and priorities after:

- Material purpose or user change.
- New country or regulated sector.
- New model, model version or supplier.
- New data source or sensitive-data use.
- Increased autonomy or new tools.
- New external retrieval.
- Significant performance drift.
- Incident, near miss or complaint.
- New affected population.
- Change to human oversight.
- New threat intelligence.
- Major control or infrastructure change.
- Decommissioning or replacement decision.

Assessor checklist

117. Worked example

End-to-end NIST AI RMF example for a generative customer-support assistant, covering scope, Profiles, evidence, testing, findings and reassessment.

This example demonstrates method and terminology. It is not a template conclusion for every customer-support system.

System

Customer Support Copilot

- Drafts responses for support agents.
- Retrieves approved knowledge-base articles.
- Uses a third-party large language model.
- Can summarise customer account history.
- Cannot send a response without agent approval.
- Processes personal data.
- Operates in production.

Scope

The assessment includes:

- Application.
- Foundation-model service.
- Retrieval layer.
- Knowledge base.
- Prompt and response controls.
- Human approval workflow.
- Monitoring and incident process.

It excludes automated account changes because the copilot has no action-capable tool for them.

Active Profile lenses

- NIST AI RMF 1.0 Core.
- NIST AI 600-1 Generative AI Profile.

Priority drivers include:

- Personal data.
- Public/customer-facing output.
- Third-party model dependency.
- Retrieval.
- Generative output.

GOVERN example: GV-6.1

Outcome focus: third-party AI risk policies.

Evidence:

- Supplier due-diligence record.
- Contract and data-processing terms.
- Model/service documentation.
- Change-notification clause.

- Incident-escalation route.

Gap:

- No documented process for evaluating a major foundation-model version change before production use.

Current Profile:

> Partially achieved

Assurance:

> Implemented

Target:

> Achieved

Treatment:

- Require advance supplier notification where available.
- Evaluate material model changes before release.
- Maintain fallback and rollback arrangements.

MAP example: MP-2.2

Outcome focus: knowledge limits and human oversight.

Evidence:

- System card describing limitations.
- Agent training.
- Mandatory approval workflow.
- Escalation procedure.

Test:

1. Present unsupported product and policy questions.
2. Confirm the copilot indicates uncertainty.
3. Confirm source articles are visible.
4. Confirm the agent can reject the draft.
5. Confirm high-risk topics route to a specialist.

Result:

- Unsupported answers sometimes appear confident.
- Human rejection works.
- Escalation does not consistently trigger.

Current Profile:

> Partially achieved

Finding:

> Confidence presentation and high-risk escalation do not reliably reflect knowledge limits.

MEASURE example: ME-2.10

Outcome focus: privacy risk examination.

Evidence:

- Data-protection impact assessment.
- Data-flow diagram.

- Prompt logging rules.
- Retention configuration.
- Access review.

Test:

1. Use synthetic account histories containing unnecessary sensitive details.
2. Verify the retrieval layer minimises context.
3. Check whether output repeats irrelevant personal data.
4. Verify logs do not retain disallowed content.
5. Check deletion and access controls.

Initial result:

> Failed - the retrieval layer passes complete account notes when only a small subset is needed.

Current Profile:

> Not achieved

Treatment:

- Field-level retrieval controls.
- Context minimisation.
- Sensitive-data filtering.
- Retest before wider rollout.

MEASURE example: ME-2.5

Outcome focus: valid and reliable demonstration.

Evaluation set:

- Common enquiries.
- Rare product issues.
- Policy questions.
- Ambiguous requests.
- Adversarial or misleading prompts.
- Different customer language patterns.

Metrics:

- Factual correctness.
- Source support.
- Appropriate abstention.
- Escalation accuracy.
- Material-error rate.
- Performance by enquiry type.

Result:

Average performance is strong, but policy questions have a materially higher error rate.

Current Profile:

> Partially achieved

The assessor does not mark Achieved merely because the average score is high.

MANAGE example: MG-4.1

Outcome focus: post-deployment monitoring.

Monitoring covers:

- Unsupported-answer rate.
- Agent rejection rate.
- Escalation rate.
- Privacy-filter events.
- Supplier model version.
- Customer complaints.
- Incident severity.

The plan also includes:

- User and agent feedback.
- Override.
- Deactivation.
- Incident response.
- Recovery.
- Change management.

Test:

A tabletop simulates a supplier update that increases unsupported policy answers.

Initial result:

- Detection works.
- Deactivation authority is unclear.
- Rollback time exceeds the intended objective.

Current Profile:

> Partially achieved

Treatment:

- Name stop authority.
- Pre-approve rollback criteria.
- Maintain a tested fallback model or non-AI workflow.

Generative-AI risks

The assessment gives specific attention to:

- Confabulation.
- Data privacy.
- Human-AI configuration.
- Information integrity.
- Information security.
- Intellectual property.
- Value-chain and component integration.

Other NIST AI 600-1 risk families are considered and documented according to their relevance.

Current Profile summary

Strengths

- System is inventoried and owned.
- Human approval exists.
- Supplier due diligence exists.
- Evaluation and monitoring have begun.
- Incident and rollback concepts are documented.

Material gaps

- Privacy minimisation failure.
- Weak high-risk escalation.
- Policy-question reliability.
- Supplier-change assurance.
- Unclear deactivation authority.
- Slow rollback.

Target Profile

The Target Profile sets Achieved for the material outcomes above, supported by:

- Passed privacy retest.
- Improved abstention and escalation.
- Risk-segmented performance thresholds.
- Supplier-change review.
- Named stop authority.
- Tested rollback and non-AI fallback.

Conclusion

Customer Support Copilot has a partially achieved Current Profile. Foundational governance, ownership, human approval, supplier review and monitoring are present, but material gaps remain in privacy minimisation, knowledge-limit handling, policy-response reliability and recovery. Wider deployment remains conditional on remediation and successful retesting. A supplier model change, new action capability, new sensitive-data source, material incident or significant performance drift requires reassessment.

What the example demonstrates

Part 4: Agentic CISO and runtime governance

This part compiles the current public operating guidance for the workflows in scope.

118. Agent Launchpad

Bring an agent into governance through a controlled sequence from registration and assessment to simulation and bounded operation.

Agent Launchpad guides an agent from proposed use to governed readiness. It coordinates the required records but does not override an incomplete gate.

Recommended sequence

1. Define the agent boundary, purpose, prohibited purposes and owners.
2. Register the agent and runtime identity.
3. Complete system intake, ACRS, ATF and relevant MAESTRO assessment.
4. Establish data movement, tools, connections and least-privilege permissions.
5. Draft and independently approve exact Runtime Access Policies.
6. Test simulations, approval, denial, timeout, retry and containment paths.
7. Resolve critical findings and accept residual risk through authorised governance.
8. Enable only the intended environment and monitor the first operating period closely.

Readiness does not equal authority

Launchpad progress indicates that governance work has been completed or remains outstanding. Live execution still requires Gateway to validate authenticated identity, connection, permission, policy, approval and request context at runtime.

Reassessment

Return to Launchpad after changes to purpose, owner, model, tools, data, supplier, environment, autonomy, delegation or risk. Suspend authority first when the change invalidates the trusted boundary.

Next

119. Agentic CISO

Agentic CISO is where Gamut governs agentic AI, the agent register, ATF assessment, tool and data governance, approvals, red-team evidence and runtime trajectory, the system of record that Gateway enforces and Claw executes against.

Agentic CISO is the governance home for agentic AI inside Gamut AI. It is the **system of record**: it owns the truth about which agents exist, what they are allowed to do, who is accountable for them, and what they have actually done. Gateway enforces that truth at runtime, and Claw (or a BYO runtime) executes the work. Nothing an agent does at runtime is valid unless it traces back to a record held here.

The separation of duties

```
Agentic CISO = governance system of record (this page)
Gamut Gateway = runtime policy decision and enforcement
Gamut Claw = secure execution layer
```

Agentic CISO never executes agent work and never holds provider credentials. Its job is to define and hold the governance state that Gateway reads on every single agent action. This is what keeps Gamut a trust plane rather than just another agent orchestrator.

What Agentic CISO holds

Governing an agent across its lifecycle takes more than an inventory entry. Agentic CISO brings together a connected set of records, each one a control surface that Gateway can enforce against:

Record	What it governs
Agent register	The inventory of agents, their owners, purpose, ATF level, risk tier and lifecycle status.
Org nodes & identity profiles	Where an agent sits in the organisation, the escalation path for high-impact actions, and the identity it acts under.
Runtime Access Policies (Access Matrix)	Exact, versioned authority for an agent, governed connection, action, resource, data class, environment, purpose and time boundary.
Tool permissions	The business authorisation binding an agent to the specific tools it may use.
Data flows	What classified data an agent may process, and how it must be handled.
Approval gates	The human approval required before sensitive or mutating actions run.
ATF assessments	The trust and autonomy assessment for each agent.

Record	What it governs
Incident playbooks	The pre-agreed response for prompt injection, rogue-agent and unauthorised-action events.
Red-team tests	Adversarial test scenarios and their results, evidence that the agent was challenged.
Gateway simulations	Recorded "what would Gateway decide" runs against a policy snapshot.
Memory & checkpoints	Governed agent memory and resumable state, kept inside the trust boundary.
Trajectory events	The stream of what agents actually did at runtime, fed back from Gateway.

Every one of these is workspace-scoped and access-controlled. Reading, changing, approving and simulating agentic governance are separate permissions, further constrained by the workspace plan. See users & roles and plans & entitlements.

The agent register

The agent register is the precondition for everything else. **No unregistered agent may take any autonomous action**. Gateway blocks it outright with a critical severity. A register entry carries, at minimum:

- A **human owner** and a **security owner**, both named. Gateway fails the action if either is missing.
- An **ATF current level** (L1 Intern through L4 Principal), which sets the agent's autonomy boundary.
- A **risk tier** (low through critical) and capability flags such as can spend money and can access customer data, which raise the bar on what controls Gateway demands.
- A **lifecycle status**. A `suspended` agent is blocked from all actions until it is reactivated through formal change control.

Whether an agent runs on Claw or an external runtime, it is registered here first. Registration is what makes a runtime identity issuable at all.

The two layers of tool authorisation

Agentic CISO and Gateway divide tool authorisation cleanly, and an agent may use a tool **only when both layers agree**:

- **Tool permissions in Agentic CISO** are the business authorisation layer: a deliberate decision that this agent may use this tool, with an approved capability scope, an audit-logging requirement, and an approval requirement for critical-risk tools.
- **Connector registration in Gateway** is the technical capability layer: the governed adapter that actually performs the call, holding the credential and the endpoint policy.

If a tool is permitted here but no connector exists, the action cannot run. If a connector exists but the tool is not permitted here, Gateway blocks it. Capability and authorisation are deliberately kept on different sides of the boundary.

Setting up an agent

Bringing an agent under governance is a short, ordered process. Each step is workspace-scoped and audited, and the Workflow Studio wizard will deep-link you to whichever of these screens still needs attention.

1. **Register the agent** in the Agent Register: give it a name, a **named human owner** and **security owner**, its purpose and **risk tier**, and its capability flags (can it spend money, take external action, reach customer data). Gateway refuses any action by an unregistered agent, or one missing an owner.
2. **Set the ATF level** (L1 Intern to L4 Principal), the autonomy ceiling Gateway enforces. Start low and promote through the ATF gates.
3. **Grant Tool Permissions** for each tool the agent needs: bind the agent to the tool adapter, set the allowed capability scope and data classes, **enable audit logging**, and require approval for critical-risk tools. A tool permission that is inactive, not runtime-enabled, or missing audit logging blocks the agent at preflight and at Gateway.
4. **Author Runtime Access Policies** for each material authority the agent needs. Bind the agent to the governed adapter and target, exact actions, resource and data boundaries, environment, purpose, dates, logging, monitoring and approval requirements.
5. **Validate and submit each policy**. Review structural validity, breadth, overlap and potential conflicts before freezing the version for independent approval.
6. **Independently approve runtime authority** through a workspace-scoped Runtime Policy Approver. Draft, pending, rejected, suspended, revoked and expired policies cannot permit runtime actions.

7. **Map the agent into the Org Chart** so high-impact actions have an escalation path back to the CEO, orchestrator, security owner or other accountable node. This proves accountability only: org chart placement does not grant inherited permissions.
8. **Configure the matching connector** so the tool can actually execute (the technical capability layer). Both layers must agree.
9. **Define approval gates** for any external, financial or code-modification action that needs human sign-off.
10. **Record data flows** for any classified data the agent processes.
11. **Simulate** the agent's intended action (see below) and resolve any control gaps before going live.

See Runtime Access Policies for the complete field model, lifecycle, approval workflow and troubleshooting guidance.

Runtime policy lifecycle

Runtime Access Policies move through controlled states:

```
Draft -> Pending approval -> Active
↑ ↓
└──── Rejected

Active -> Suspended or Revoked
Active -> Expired or Superseded
```

Submitted and authoritative versions are immutable. A change to active authority is proposed as a new draft and must pass validation and independent approval again. In a locked workspace, policies remain visible for audit and inspection while authoring and lifecycle changes remain disabled.

The **Validate & impact** view helps assess policy completeness, broad authority, overlap and potential allow/deny conflicts. Validation confirms that the policy can be reviewed; it does not replace human risk judgement.

How ATF level bounds autonomy

The ATF level on the register entry is not a label, it is an enforced ceiling. Gateway applies a different action boundary at each level:

Level	Boundary enforced at runtime
L1 Intern	Read, observe and report only. Any write or external action is blocked.
L2 Junior	Routine actions allowed; every external or financial action requires an approval gate.
L3 Senior	Operates within guardrails; financial and high-risk external actions require approval.
L4 Principal	Strategic autonomy; critical or top-secret data access still requires documented approval.

Raising an agent's autonomy is therefore a governed decision made here, with immediate runtime consequences. See ATF for the full level model and promotion gates.

Runtime evidence and accountability

Because every agent action flows through Gateway, Agentic CISO receives a continuous stream of **trajectory events**: what was requested, which controls passed or failed, what Gateway decided, and what happened. Combined with the journal that Claw keeps for each task, this turns agent activity from an opaque black box into the same reviewable, auditable evidence and findings model the rest of Gamut uses. Open risks or findings linked to an agent are surfaced back into the Gateway decision, so an agent with unresolved critical findings cannot quietly keep operating.

Gateway decisions also carry bounded policy evidence back into Gamut: the matched policy and version, access decision, org escalation path, matched approval gate, controls passed, controls failed and decision path. The Gateway Event Centre shows these events without exposing raw payloads or secrets.

Simulating before you authorise

Before an agent goes live, you can run a **Gateway simulation** from Agentic CISO: choose an agent (or a transient template agent), a tool, an action type, a data classification, a target system and an environment, and Gateway returns the full decision it would make, the controls passed, the controls failed, the missing evidence, the decision path, the runtime-policy match and the org escalation path, against a frozen policy snapshot with a 90-day validity and explicit retest triggers. This lets you close control gaps before any real action is ever attempted.

Break-glass: suspending an agent or halting the runtime

Governance includes the ability to stop. Gamut offers graduated controls, all audited:

- **Suspend one agent.** Set the agent's lifecycle status to `suspended` in the register, or use **Suspend Agent Runtime** in the admin console. Gateway then blocks every action by that agent, fail-closed, until it is reactivated through change control.
- **Stop one running step.** Use the **Stop** button in Workflow Studio, which cooperatively cancels that step's Claw task at the next safe checkpoint.
- **Halt all agentic execution (kill switch).** An administrator can flip the **Gateway kill switch** in the admin console (Gateway Status, Disable Gateway Runtime). Gateway then blocks every agent action, and Gamut itself fail-closes every dispatch and skips scheduled runs, so nothing executes until it is cleared. A separate Claw kill switch, set on the Claw service, halts the execution worker entirely. See Gateway.

Next

120. Agent identity, ownership and organisation

Establish accountable agent identities, human ownership and escalation relationships before granting runtime authority.

Agentic governance starts with a stable identity. An agent name or job title is not an identity and does not grant authority. The Agent Register, Identity & Ownership and Organisation Chart establish who or what the agent is, who is accountable and where escalation goes.

Agent Register

Create one record for each independently governed agent boundary. Record purpose, owner, operating status, runtime, model, environment, data access, tools, autonomy, criticality and review dates. Separate agents when they have materially different authority, owners, runtime identities or failure consequences.

An active register entry is necessary but not sufficient for execution. Tool Permissions, a ready connection, an active Runtime Access Policy and any required approval must also pass.

Identity & Ownership

Bind the governance record to the runtime identity used in authenticated requests. Record the business owner, technical owner and accountable executive or service owner as appropriate. Owners remain responsible for purpose, risk, access review, incidents and retirement; the agent cannot own its own risk acceptance.

Rotate or revoke runtime identity material through the approved service workflow. Never place raw credentials, service secrets or private keys in the Agent Register.

Organisation Chart

The chart explains reporting, delegation and escalation. It can help determine who must be consulted or approve a high-impact action, but position in the chart does not create runtime permission. Gateway evaluates exact policy and request context.

Lifecycle controls

1. Register the agent and owners.
2. Complete risk tiering and ATF assessment.
3. Configure least-privilege tools and governed connections.
4. Validate simulations and approval paths.
5. Activate only the authority required for the intended environment.
6. Review after change, incident, expiry or owner transfer.
7. Suspend identity and runtime authority before retirement.

Review checklist

- [ ] Unique agent boundary and runtime identity.

- [ ] Named, current human owners.
- [ ] Purpose and prohibited purposes defined.
- [ ] Environment and data boundaries recorded.
- [ ] Escalation path is usable but not treated as permission.
- [ ] Dormant, duplicate and retired identities are disabled.

Next

121. Runtime Access Policies

Define, validate, approve and maintain exact, versioned authority for AI-agent actions, with least privilege, independent review and execution-time enforcement through Gamut Gateway.

A **Runtime Access Policy** defines the exact authority an AI agent may exercise through Gamut Gateway. It answers a narrow question:

> May this agent perform this action, through this governed connection, on this resource, with this class of data, in this environment, for this purpose, at this time?

The answer is never inferred from an agent's title, organisational position or general access to Gamut. Missing or ambiguous authority is denied.

Runtime Access Policies are managed from **Agentic CISO -> Access Matrix**. The navigation label reflects the relationship between agents and resources. The records themselves are exact, versioned policies with a controlled lifecycle.

Where a Runtime Access Policy fits

A policy is one part of a defence-in-depth decision. It does not make an action permissible on its own.

Control	What it proves
Agent register	The agent has an accountable identity, owners and an active lifecycle state.
Gateway tool registration	The requested adapter is a governed capability known to Gateway.
Governed connection	The exact tenant and workspace connection is configured and ready.
Tool Permission	This agent is allowed to request the selected capability.
Runtime Access Policy	The requested action, resource, data, environment and purpose are within an active policy.
Approval gate	A qualifying sensitive action has the required human decision.
Runtime controls	Logging, monitoring, current policy validity and request-bound authority remain satisfied when the action executes.

Every applicable control must pass. A broad Tool Permission cannot compensate for a missing Runtime Access Policy, and an active policy cannot compensate for a suspended agent or unavailable connection.

Policy fields

Identity and effect

- **Policy name**: a human-readable description of the authority being governed.
- **Agent**: the registered agent to which the policy applies.
- **Effect**: `Allow` or `Deny`. Deny policies take precedence when an allow and deny boundary intersect.
- **Risk rating**: the assessed risk of the governed authority.

Capability and target

- **Gateway adapter**: selected from the agent's active Tool Permissions.
- **Governed target / connection**: the exact ready connection for that adapter. Where Gamut provides a governed internal service, that service can appear as the target instead.

- **Resource boundary**: the specific object, collection or path the action may affect. A bounded trailing prefix can cover a defined collection. An unrestricted wildcard is treated as privileged authority and requires stronger justification and review.

The governed-target list is not free text. It is derived from the selected adapter and the connections that are available to the current tenant and workspace.

Action and information boundaries

- **Exact actions**: the operations the policy covers, such as read, retrieve, create, update, execute, send a message, export data or invoke a model.
- **Allowed data classifications**: the highest and specific information classes the action may handle.
- **Environments**: development, test, staging or production.
- **Workflow boundary**: an optional exact workflow to which the authority is restricted.
- **Purpose boundary**: an optional business-purpose restriction.

Choose only what the use case needs. If an agent requires unrelated actions or targets, create separate policies so each authority can be reviewed, suspended and replaced independently.

Assurance and time boundaries

- **Human approval**: whether execution also requires an approval gate.
- **Gateway logging**: mandatory for governed runtime authority.
- **Continuous monitoring**: whether ongoing runtime monitoring is required.
- **Approver role**: the accountable role expected where approval applies.
- **Effective date**: when the authority may begin.
- **Expiry date**: when the authority stops being valid.
- **Review due date**: when the business and security basis must be reconsidered.
- **Business and security justification**: why the authority is necessary, proportionate and time-bounded.

Dates have different purposes. Expiry controls runtime validity. Review due prompts governance attention. Reaching a review date does not silently extend an expired policy.

Policy lifecycle

Status	Meaning	Gateway treatment
Draft	The author is defining and validating the policy.	Not authoritative and not used to permit actions.
Pending approval	The submitted version is awaiting independent review.	Not active and not used to permit actions.
Active	The approved version is within its effective and expiry dates.	Eligible for exact runtime matching.
Rejected	The reviewer returned the proposal with a reason.	Not active. The author may revise and resubmit it.
Suspended	Previously active authority has been paused.	Immediately unavailable for new actions.
Revoked	Authority has been withdrawn.	Permanently unavailable.
Expired	The policy passed its expiry date.	Unavailable without a newly approved version.
Superseded	A controlled replacement became authoritative.	The replacement version is used; the prior version remains historical.

Submitted and authoritative versions are immutable. To change active authority, clone the policy, edit the new draft, validate it and send that new version through approval. This preserves a clear history of what was authorised at any point in time.

Authoring workflow

1. Establish the prerequisites

Before creating a policy:

1 Register the agent and assign its accountable owners.

2. Confirm the agent's lifecycle and ATF level.
3. Register or enable the required Gateway adapter.
4. Configure and test the governed connection where one is required.
5. Grant the agent the matching Tool Permission.
6. Define an approval gate for actions that require human approval.

If the adapter or target does not appear in the builder, resolve the corresponding Tool Permission or connection first. Do not replace a missing governed target with a descriptive name.

2. Define the smallest useful boundary

Select one agent, one governed capability and a precise target. Add only the required actions, resource scope, information classifications and environments. Record the purpose and the reason the authority is proportionate.

3. Validate and review impact

Use **Validate & impact** before submission. Validation checks structural completeness and the relationship between the proposed policy and its governed context. Impact analysis helps identify:

- missing prerequisites;
- broad or privileged boundaries;
- overlapping active or pending policies;
- potentially contradictory allow and deny coverage;
- other active policies that could be affected.

Structural validity means the policy is complete enough to review. It does not mean the business risk is acceptable or that the policy is approved.

4. Save and submit

Save the proposal as a draft while it is being developed. When the content and impact are ready, submit it for independent review. Submission freezes that version and moves it to **Pending approval**.

5. Independently review

The reviewer should confirm:

- the agent, tool and target are correct;
- the action and resource boundaries match the stated need;
- the data classifications and environments are no broader than necessary;
- approval, logging and monitoring requirements match the risk;
- effective, expiry and review dates are appropriate;
- the justification explains necessity and proportionality;
- overlapping or conflicting policies have been resolved;
- the reviewer is independent of the creator and submitter.

The reviewer may approve and publish the policy or reject it with a reason.

6. Operate and reassess

An active policy remains subject to Gateway's complete runtime decision chain. Reassess it when the agent, connection, tool, resource, purpose, data, environment, risk, approval requirement or organisational ownership changes.

Independent approval

Runtime authority can affect live systems, so approval uses a dedicated **Runtime Policy Approver** role. The role is assigned to the exact workspace containing the policies that person may review.

An approver can:

- inspect policy details and relevant supporting context;

- filter the approval inbox;
- approve a pending policy;
- reject a pending policy with a recorded reason.

An approver cannot use that role to draft, edit, delete, submit, suspend or revoke policies, administer users, operate Gateway or change agent governance records. Approval and rejection both require the approver's own password and current MFA code through a short-lived step-up verification. The creator or submitter cannot approve the same policy.

See Users & roles for assignment and access details.

Execution-time validity

Approval is necessary, but it is not treated as permanently valid. When Gateway is ready to execute an allowed action, it checks that the authority still matches the request and that the governance state has not changed.

The execution is refused if, for example:

- the policy was suspended, revoked, expired or replaced;
- the policy content or version no longer matches the authorised decision;
- the agent, tool, connection, action, resource, data class, environment, workflow or purpose no longer matches;
- a required approval no longer applies to the exact request;
- the request-bound authority is missing, expired, replayed or inconsistent;
- required logging or monitoring cannot be satisfied.

This closes the gap between a decision being made and an action being executed. See Gamut Gateway for the complete runtime decision model.

View-only workspaces

When a workspace is locked in **View only** mode, existing policies remain available for inspection. Drafting and lifecycle changes are disabled. This supports audit and review without allowing the governance record to be altered.

Troubleshooting

The governed target list is empty

Confirm that the selected agent has an active Tool Permission for the adapter and that the required tenant and workspace connection is configured and ready.

A submitted policy is not visible to an approver

Confirm that:

1. the person has the dedicated Runtime Policy Approver account role;
2. that account has the Runtime Policy Approver workspace role for the exact workspace;
3. the policy is in that workspace;
4. the workspace assignment has not expired;
5. the policy status is **Pending approval**.

MFA is required to make the decision, not to list the assigned policy.

Approval asks for identity verification

The approver must enrol MFA, then complete the password and authenticator step-up. After successful verification, return to the policy and confirm the decision.

A previously allowed action is blocked

Review the current agent, connection, Tool Permission, policy, approval gate and runtime status. Gateway intentionally refuses to rely on an earlier decision when the current context no longer matches.

Worked example

A security triage agent needs to create a finding in Gamut from a verified SIEM alert.

The policy could specify:

- **Agent**: Security Alert Triage Agent
- **Adapter**: Gamut finding creation
- **Target**: the governed Gamut service
- **Resource**: the designated findings collection
- **Action**: create
- **Data**: internal
- **Environment**: production
- **Purpose**: convert verified security alerts into governed findings
- **Logging and monitoring**: required
- **Expiry and review**: time-bounded to the operational approval

The policy does not allow the agent to delete findings, modify users, change permissions or act on another workspace. Any additional authority requires its own explicit governance decision.

Next

122. Gamut Gateway

Gamut Gateway is the policy decision and enforcement engine for agentic AI. Every agent action passes through it, is checked against a sequence of governance controls, and is allowed, approval-gated, degraded or blocked, with a signed decision and a complete audit record.

Gamut Gateway is the policy decision and enforcement engine of the agentic stack. It is the **ATF runtime**: every agent action passes through Gateway, which evaluates it against the full governance context, decides, performs the call only if permitted, and records the decision either way.

What Gateway does

Gateway sits between agents and the models, tools and data they want to use. Nothing reaches a tool directly. For each requested action it assembles the governance context, agent register, Runtime Access Policies, org chart, tool permissions, approval gates, data flows, incident playbooks, red-team tests, linked risks and findings, and runs a sequence of controls before allowing anything.

The decision model

Gateway resolves every request to one of four runtime decisions:

Decision	Meaning
allow	All controls passed. The action proceeds, and is logged.
require_approval	A high-impact action with control gaps. A named human must approve first.
degrade	The action proceeds in a reduced, safer form.
block	The action is refused.

For high-risk or external actions that pass, the decision becomes **allow and log** with mandatory detailed audit logging. Where critical control failures coincide with an open critical finding, the decision escalates to **block and open incident**.

The controls Gateway evaluates

For each action, Gateway runs an ordered series of checks. Any of the first two short-circuit to an immediate block:

1. **Agent registration.** An unregistered agent is blocked at critical severity. No exceptions.
2. **Lifecycle status.** A `suspended` agent is blocked until formally reactivated.
3. **Human owner** is assigned.

4. **Security owner** is assigned.
5. **ATF maturity level** is assigned, and the action sits within that level's boundary (see below).
6. **Tool permission.** The tool is registered for this agent in Agentic CISO, with audit logging enabled, and human approval required if it is a critical-risk tool.
7. **Approval gate.** External, financial or code-modification actions are covered by an active approval gate that matches the action type, tool, target system, data class and environment.
8. **Runtime Access Policy.** Material actions must have a matching active, least-privilege policy for the agent, adapter, governed connection, action, resource, data class, environment and any workflow or purpose boundary. Missing or non-matching coverage fails closed. Draft, pending, rejected, suspended, revoked and expired policies never permit execution.
9. **Org escalation path.** High-impact actions must map back to an org-chart reporting path so escalation and accountability are clear. The org chart does not grant permission and does not create inherited access.
10. **Data movement.** Customer or regulated data (PII, confidential, restricted) has a documented data-flow record.
11. **Incident playbooks.** High-risk agents have playbooks for prompt injection, unauthorised external action and rogue-agent scenarios.
12. **Linked risks and findings.** Open risks or findings tied to the agent are surfaced; open critical findings can escalate the decision.
13. **Red-team validation.** High-risk or external actions have at least one passed red-team test on record.

Each check contributes to a **decision path** and, on failure, a **rule trigger**, so every decision is fully explainable: which controls passed, which failed, what evidence is missing, and exactly what would change the decision.

Gateway also records a bounded **policy trace**. The trace shows the runtime-policy match, the org escalation path, the matched approval gate and the decision path without storing raw payloads or secrets. The Gateway Event Centre surfaces this evidence for operators.

Runtime Access Policy and org chart enforcement

The **Access Matrix** screen manages Runtime Access Policies. It is an enforcement control, not just documentation. For a material action, Gateway checks for an active policy covering the requested agent, adapter, governed connection, action, resource, data, environment, workflow and purpose. If no policy covers the exact request, Gateway blocks it and records the gap. A matching deny policy takes precedence over an allow policy.

The org chart has a different purpose. It proves reporting, escalation and accountability. Gateway uses it to build an escalation path for high-impact decisions, but it never uses the org chart to grant access, inherit permissions or override Runtime Access Policies, tool permissions or approval gates.

ATF-level action boundaries

Gateway enforces the agent's ATF level as a hard ceiling on autonomy:

- **L1 Intern** may only read, observe and report. Writes and external actions are blocked.
- **L2 Junior** may take routine actions; every external or financial action requires an approval gate.
- **L3 Senior** operates within guardrails; financial and high-risk external actions require approval.
- **L4 Principal** has strategic autonomy; critical or top-secret data access still requires documented approval.

Why enforcement lives in one place

Concentrating enforcement in Gateway is what makes the zero-trust model work:

- **Credentials stay on Gateway.** Model provider keys and connector credentials live only on the Gateway side, never with the agent. A compromised agent cannot leak keys it never held.
- **Policy is consistent.** Every action is judged by the same engine, so governance does not depend on each agent behaving well.
- **Evidence is complete.** Because every action flows through one point, the runtime evidence fed back to Agentic CISO is comprehensive.

Request-bound, short-lived authorisation

Gateway does not just return a verdict. It issues **signed, time-bounded and request-bound authorisation**. The authority is tied to the tenant, workspace, agent, request, governed connection and the exact active policy version. Only an allowed and fully satisfied request can receive executable authority.

Immediately before execution, Gateway confirms that the policy and surrounding governance state are still current. The action is refused if the policy has changed, expired, been suspended, revoked or replaced, or if the request no longer matches the authorised identity, action, resource, connection or context. A required approval is bound to the same request and cannot be reused for unrelated work.

Execution authority is short-lived and replay-resistant. Repeated delivery of the same request does not create duplicate side effects, and concurrent execution attempts cannot both consume the same authority. This closes the time-of-check to time-of-use gap between deciding and acting.

Connectors and approval gates

Tools are exposed to agents as governed **connectors** registered in Gateway, model gateways, retrieval, HTTP and webhook adapters, ticketing, notification, storage, CRM and more, each with its own action type, risk tier, payload limits and response handling. See the connector catalog. Sensitive or mutating actions can require explicit human approval; those **approval gates** are defined as governance in Agentic CISO and enforced here.

Fail-closed by default

If Gateway cannot reach a dependency, cannot verify a signature, or encounters an error mid decision, it fails closed: the action does not proceed. Safety never depends on a call succeeding.

The kill switch (break-glass)

When you need to stop everything, an administrator can activate the **Gateway runtime kill switch** from the admin console (**Gateway Status, Disable Gateway Runtime**). It requires gateway-configuration permission and a step-up re-authentication, and it is audited. While it is active:

- **Gateway blocks every agent action** with a kill-switch decision, regardless of agent or policy.
- **Gamut itself fail-closes** every dispatch path, so a new Claw task, a workflow dispatch and a scheduled run are all refused before they reach the runtime. Scheduled runs are **skipped, not failed**, so they resume automatically once the switch is cleared.

This is defence in depth: the halt does not depend on any single service. A separate **Claw kill switch**, set on the Claw execution service, stops the worker entirely. Reactivating the Gateway runtime is itself a privileged, audited action. To halt a single agent or a single step instead, use agent suspension or the Workflow Studio **Stop** button (see Agentic CISO).

Next

123. Tool permissions and governed connections

Define the maximum tool authority an agent may request and bind runtime policies to exact Gateway-managed targets.

Tool Permissions define the outer boundary of what an agent may request. **Gateway Tooling** defines the governed adapters and connections capable of carrying out those requests. Neither one alone authorises execution.

Tool Permission

Grant the smallest action set, resource boundary, data classification and environment required for the agent's purpose. Separate read, search, create, update, delete, execute, external communication, data export, financial action, permission change and code modification. Broad wildcards require exceptional justification and stronger approval.

Review permissions after purpose, owner, supplier, model, environment or risk changes. Remove unused actions rather than relying on the agent not to call them.

Gateway Tooling

A governed target or connection comes from the Gateway tooling catalogue available to the tenant and agent. It identifies a configured adapter and exact target under Gateway custody. A free-text name cannot create a connection or bypass readiness.

Before use, confirm:

- the connector is approved and healthy;
- credentials are stored by the governed connection, not the agent;
- the target and environment are correct;
- egress, resource and data boundaries are explicit;
- monitoring and logging requirements can be met;
- the connection owner and review date are current.

Zero-trust intersection

Runtime authority is the intersection of authenticated agent identity, active connection, Tool Permission, exact Runtime Access Policy, required approval and live Gateway checks. A permissive record at one layer cannot compensate for a missing or narrower record at another.

Change and retirement

Changing a connection, credential, target, permission or environment can invalidate policy assumptions. Revalidate affected policies and simulations. Revoke or disable retired connections before removing descriptive records so runtime authority fails closed.

Next

124. Gamut Claw

Gamut Claw is the secure agent execution layer. It plans and reasons but acts only through Gateway-controlled paths, holds no credentials, runs governed task types under leases, redacts output and keeps a hash-chained journal of every step.

Gamut Claw is the secure execution layer of the agentic stack. It is where governed agent work actually runs. The defining rule is simple: **Claw may think, but it may only act through Gateway.** It never calls a model, invokes a tool, retrieves data or writes back a result directly.

Claw's role

An agent running on Claw can plan and reason locally, but whenever it needs to act it issues a request to Gateway, which decides, enforces policy, and performs the call. Claw receives back only bounded, governed output.

```
Claw requests work and executes only through Gateway-controlled paths.
Gateway applies and enforces policy, holds credentials, performs the call.
Gamut AI records what happened (trajectory + journal).
```

This keeps a strict separation between deciding to act (Gateway, on Gamut policy) and running the work (Claw). Claw is an ATF-controlled surface for the agents that Agentic CISO governs.

Why Claw holds no credentials

Claw never receives model, SaaS, database, email, payment or other provider credentials. Those live only on the Gateway side. The benefit is direct: if a Claw execution environment were compromised, there are no provider keys there to steal and no path to a tool that bypasses policy. Every capability Claw appears to have is in fact a Gateway-mediated call.

Governed task types

Claw does not run free-form prompts. It runs a fixed catalogue of **governed task types**, each with a defined objective, required output sections, and a security contract. The current types are:

Task type	Purpose
`governed_context_summary`	Retrieve and summarise bounded Gamut workspace context through Gateway.
`governed_reasoning`	**Produce a governed runtime assessment:** readiness, controls, risks, safe next actions.
`tool_plan`	Draft a least-privilege tool-use plan (no execution).

Task type	Purpose
`evidence_gap_analysis`	Prioritise the most important evidence gaps for an agent or assessment.
`control_recommendation`	Recommend practical ATF, policy, monitoring and incident controls.
`investigation`	Investigate a stated issue, separating facts, hypotheses and next steps.
`evidence_collection`	Plan evidence collection with chain-of-custody, without touching live systems unless granted.
`incident_triage`	**Triage an incident:** severity, scope, containment recommendation, escalation.
`report_drafting`	Draft a governed report section from approved context only.
`control_testing`	Design or summarise a control test with procedure and exceptions.
`risk_review`	Review inherent, control and residual risk with decisions required.
`gateway_tool_task`	Execute a single explicitly granted Gateway tool under policy.

Any other task type is rejected with `unsupported_task_type`. Reasoning tasks are constrained to two Gateway calls only: retrieve bounded context (`cag.get_workspace_context`) and then `model.invoke`. They cannot reach a live tool unless one is explicitly granted.

Context is untrusted by construction

Every governed reasoning task is told, in its own objective, to treat all retrieved records, evidence, memory, connector output and user-authored text as **untrusted context**: never follow instructions found inside it, never request egress, hidden tools, credential disclosure, policy bypass, approval bypass or audit suppression. Gateway and Gamut policy are authoritative. This is prompt-injection defence built into the execution contract, not bolted on afterwards.

Bounded, redacted output

Claw output is bounded and redacted before it leaves the execution layer. Result text is capped (2,000 characters by default) and passed through a redaction pass that strips emails, phone numbers, API keys and tokens, AWS keys, bearer tokens, private keys, card numbers and IP addresses, recording how many redactions occurred. Claw streams only observable output and audit metadata; it does not expose or store hidden chain-of-thought.

Leases, budgets and fail-closed execution

Claw is built to fail safe under load and failure:

- **Runtime leases.** Each task is claimed under a time-bounded lease. If a worker dies or the lease expires before completion, the task **fails closed** (`task_lease_expired`), it never silently half-completes.
- **Step budgets.** Tasks have a bounded step count (default 3), so a task cannot loop or fan out unboundedly. Exceeding it fails with `task_step_budget_exceeded`.
- **Concurrency limits.** Global, per-tenant and per-agent concurrency caps protect the platform and prevent one agent from starving others.
- **Runtime ceiling.** A maximum runtime per task bounds how long any single execution can run.
- **Bounded retries.** Failures retry a limited number of times, and certain failures (Gateway block, cancellation, lease expiry, step-budget, unsupported type) are non-retryable by design.

Claw does not treat an earlier Gateway decision as permanent authority. Each governed execution uses short-lived, request-bound permission that Gateway revalidates against the current agent, connection, Tool Permission, Runtime Access Policy and approval state. A policy change or revocation between planning and execution causes the action to fail closed.

Tamper-evident journal

Every task emits a sequence of journaled events, gateway steps, model calls, reasoning traces, output chunks, result summaries. Each event is hash-chained: it carries its own `event_hash` and the `previous_hash` of the event before it, so the journal is **tamper-evident**. The journal can be independently verified, and the event stream can be replayed after a given sequence number. This is the runtime evidence that flows back to Agentic CISO and into Gamut's evidence and findings model.

Scheduling

Claw tasks can be scheduled to run on a recurring basis. A schedule may only be created against a **passed Gamut preflight attestation**, and that attestation must still be valid when the task fires. If the attestation is missing or expired, the scheduled run is refused (`schedule_preflight_attestation_required` / `_expired`). Schedules can be paused and resumed, but they can never escape the same governance checks a live task faces.

Claw versus bring-your-own runtime

Claw is Gamut's own execution layer, but it is not the only option. If you already run agents on other frameworks, the BYO agent runtime lets them execute under the same model, **think anywhere, act through Gateway**, without adopting Claw. The governance and enforcement model is identical; only the execution surface differs.

Next

125. Bring-your-own runtime

Run external agent frameworks, LangGraph, CrewAI, Hermes, OpenClaw and custom code, while keeping Gamut as the trust and enforcement plane. Think anywhere, act through Gateway, via a signed SDK that fails closed.

The **BYO agent runtime** lets you run external agent frameworks while keeping Gamut as the trust and enforcement plane. You are not required to adopt Gamut Claw to benefit from governed agentic AI. Your agent keeps its own orchestration logic; Gamut governs what it is allowed to do.

The operating rule

> **Think anywhere. Act through Gateway.**

External agents may plan and reason locally, in whatever framework you prefer. But context, model calls, tools, connectors, approvals, run events and governed writeback must go through Gamut Gateway. Planning is unconstrained; action is always governed.

The SDK and its three governed actions

Gamut ships a BYO agent SDK (Node and Python) that an external runtime embeds. It exposes exactly three governed actions, and nothing else reaches a provider:

Action	What it does
context	Retrieve a bounded, tenant-scoped Gamut workspace context pack through Gateway.
model	Run a governed model inference over an approved context pack, with provider keys held by Gateway.
invokeTool	Invoke a named, permitted connector through Gateway, optionally carrying an approval warrant.

Each call is signed and tenant-scoped, and carries the agent's registered identity. The runtime receives only a single Gamut runtime secret. It is given **no** model-provider, SaaS, database, payment, email, SIEM, SOAR, shell or browser credentials, and the SDK actively rejects any attempt to pass direct provider configuration or secret material in a payload.

Configuration

An external runtime is wired with just its Gamut coordinates and one secret:

```
GAMUT_BASE_URL=https://your-gamut-host.example
GAMUT_TENANT_SLUG=main
GAMUT_WORKSPACE_ID=12
GAMUT_AGENT_ID=14
GAMUT_BYO_AGENT_SECRET=...
```

The `GAMUT_AGENT_ID` must correspond to an agent already registered in Agentic CISO. Registration is the precondition for the secret to authorise anything.

Fails closed, always

The SDK fails closed, the action does not proceed, on any of:

- a non-2xx response from Gamut,

- Gateway returning `invoked: false` (policy denial),
- a request timeout,
- a stale or revoked credential,
- an identity mismatch,
- any policy denial.

Safety never depends on a call succeeding. A failed governance check is a stop, not a warning.

Supported frameworks

The SDK ships bridges that keep popular frameworks inside the trust boundary, plus a minimal base client for custom Node or Python agents:

- **LangGraph** and **CrewAI** wrapper tools, so graph and crew steps act through Gateway.
- **Hermes Agent** chat, responses and toolset wrappers, backed by Gateway.
- **OpenClaw** run, stream, cancel, model and tool helpers, backed by Gateway.
- **Custom agents** via the base Node and Python clients.

These bridges must not receive model-provider keys, direct tool credentials or alternate provider base URLs; the SDK rejects direct provider configuration and continues to require signed Gamut runtime credentials for every governed action.

The security model

The BYO runtime preserves the same zero-trust posture as Claw:

- **No credentials to the agent.** Provider credentials remain on the Gateway side.
- **Registered identity.** Each external agent is registered in the Agentic CISO agent register, with an owner, scope and ATF level.
- **Governance parity.** A BYO agent passes the same Gateway controls, ATF boundary, Runtime Access Policies, org escalation evidence, tool permissions, approval gates and data flows as any other governed agent before any action.
- **Enforced at runtime.** Gateway enforces tool permissions, action types, data classes, rate limits, exact policy coverage and approval gates on every request, then revalidates the short-lived authority before execution.
- **Fully logged.** Gamut and Gateway record every action, giving the same runtime evidence as any other governed agent, including bounded policy trace evidence for Gateway Event Centre review.

Tool brokers

External tool brokers (such as MCP-based brokers) are supported only as governed Gateway tool connections, with declared profiles and explicit per-tool scopes. External agents never receive a broker credential or a direct broker URL; they request the tool through Gateway like any other connector. See the connector catalog.

Delegation

For frameworks that spawn sub-agents or delegate work, Gamut can issue short-lived, governed child identities so delegated work stays inside the same trust and enforcement model rather than escaping it. A delegated sub-agent is no less governed than its parent.

Next

126. Data movement and approval gates

Govern where agent data can move and when a human decision is required before an action can execute.

The **Data Movement** view documents permitted flows between an agent, governed connection, resource and destination. **Approval Gates** identify actions that require a bound human decision. Together they make authority understandable before Gateway evaluates a live request.

Data Movement

For each flow, document source, destination, purpose, data classification, affected people, geography, retention and permitted action. Include indirect movement through model providers, plugins, queues, logs and generated attachments.

Do not infer permission from technical connectivity. A reachable destination is not an approved destination. Restrict sensitive and regulated data to the minimum fields and period needed, and test redaction and egress controls.

Approval Gates

Require approval when the action is high-impact, irreversible, external, financially material, privilege-changing, broad, production-facing or otherwise outside routine bounded authority. Specify the approver role, decision context, expiry and conditions.

An approval must refer to the same agent, action, resource, target, environment and request context that will execute. Generic standing consent should not be used where transaction-level review is required.

Separation of duties

The requester or policy submitter cannot satisfy an independent approval requirement for the same authority. Runtime Policy Approvers can inspect assigned policy context and approve or reject; they cannot draft or edit policy through that role. MFA and step-up verification apply to sensitive decisions.

Review checklist

- [ ] Every material ingress and egress path is represented.
- [ ] Data classification and geographic restrictions are current.
- [ ] External recipients and subprocessors are identified.
- [ ] High-impact actions have an appropriate approval gate.
- [ ] Approval is request-bound and time-bounded.
- [ ] Denial, timeout, withdrawal and dependency failure stop execution.

Next

127. Connector catalog

How Gamut exposes broad tool access to agents through governed Gateway connectors, models, retrieval, HTTP, ticketing, notification, storage, CRM and more.

Connectors are how Gamut exposes broad tool access to agents **without weakening the zero-trust architecture**. Agents never call tools directly. Tools are registered, governed and invoked through Gamut Gateway after policy, permission, tenant, assessment and identity checks.

How a connector is governed

Each connector is a governed adapter with a defined contract, rather than a raw API call. A connector carries:

- A stable **tool name** (for example `siem.search_alerts`).
- An **action type**, the policy category, such as retrieve, report, ticket, notify or model.
- A default **risk tier** for runtime classification.
- A **credential reference** to a managed Gateway-side secret, never a raw secret in a workflow.
- An **endpoint policy**, allowlisted destinations and safety rules.
- **Payload and response policies**: accepted input shape and size, plus redaction and retention rules for outputs.
- An **audit policy**, the fields recorded for every decision and invocation.

Because credentials live on the Gateway side and destinations are allowlisted, agents get capability without ever holding keys or reaching arbitrary endpoints.

Built-in connector families

Gamut ships with connector families spanning common enterprise needs. Availability and enablement depend on your plan & entitlements and workspace configuration.

Family	Typical use
Model gateway	Model reasoning or generation (provider keys live only on Gateway).
Context (CAG)	Retrieve governed Gamut workspace context, tenant-scoped.
Retrieval (RAG)	Search approved knowledge stores with redaction policy.
MCP brokers	Broker approved MCP tools with explicit per-tool scopes.
Configurable HTTP	Wire customer REST APIs without code changes.
Webhook	Send bounded JSON events to approved destinations.
SIEM / SOAR	Security investigation context and incident workflows.
Ticketing & work	Create governed work items and agent tasks.
Document	Produce governed workpapers and summaries.
Database	Read approved operational data under strict query policy.
Gamut writeback	Materialise approved findings and evidence requests into Gamut.
Notification	Notify approved recipients by email or chat.
CRM	Read or update customer records under PII controls.
Productivity	Mail, calendar, drive and directory APIs, path-scoped.
Content & wiki	Retrieve or update approved knowledge stores.
Object storage	Read or write scoped object storage.
Research & market data	News, RSS and market research from allowlisted sources.
Finance	Finance or payment operations (critical risk, approval-gated).
Public channels	Publish approved content externally (approval-gated).

Two layers must agree

An agent can use a connector only when **both** authorisation layers align:

- **Tool permissions** in Agentic CISO, the business decision that this agent may use this tool.
- **Connector registration and policy** in Gateway, the technical capability, plus a policy that allows the requested action, context and payload.

Add the required approval gates and a valid scoped runtime token, and only then does the action proceed. This is the mechanism that lets Gamut offer broad reach safely.

Configuring a connector

Connectors are configured in the **connector catalogue** panel of the Workflow Studio launchpad. Configuring one is a governed, audited action that requires an editable session and the gateway-enforce permission. For each connection you provide:

1. The **adapter** to use (for example a model gateway, an MCP tool broker, or a configurable HTTP adapter) and, where relevant, the **provider**.
2. A **connection name**.
3. The **endpoint** (an HTTPS destination, which must fall within the adapter's allowlist policy).
4. The **credential** (an API token or OAuth secret). The credential is **encrypted in the browser to a Gateway public key** and handed to Gateway custody; Gamut never stores it in plaintext and an agent never holds it.
5. The explicit **scopes**, **allowed actions** and **allowed data classes** the connection may use.

Once saved, the credential is validated and the connector becomes available to permit to agents. A connection that has not passed validation, or whose credential is not in Gateway custody, surfaces as a runtime-readiness blocker in the Workflow Studio wizard. Tools that do not need an external credential (such as governed Gamut context retrieval) require no connector configuration.

Validated connections also populate the **Governed target / connection** selector when a Runtime Access Policy is authored for a compatible agent and adapter. Selecting from this catalogue binds the policy to an exact tenant-scoped connection; an

arbitrary name cannot create a new runtime target or bypass connector readiness.

Next

128. Visual Workflow Studio

Design, govern, run and schedule agentic workflows in Gamut. The Plan Runtime Wizard walks you from an empty workspace to a dispatched, Gateway-enforced runtime and tells you exactly what to fix at each gate.

Visual Workflow Studio is where you design, run and schedule governed agentic workflows. You compose the steps an agent takes, and every action step remains governed by Gateway. It turns the zero-trust model into something you can build, review, run, watch and stop, before and while it executes.

The Studio opens as its own page (the **Workflow Studio** link in the app top bar). It needs an editable, Advanced-tier session with Agentic CISO access, and the platform must be running in its database mode.

The building blocks

A workflow is assembled from a small set of governed objects, each held in Agentic CISO and each carrying its own policy:

Object	What it is
Workflow plan	**The overall design:** a name, objective, root agent, and the ordered steps.
Workflow step	A single step with a task type. Action steps reference a governed connector by tool name.
Task type	What a step does, for example governed reasoning, tool plan, evidence collection, investigation, or a single Gateway tool call.
Toolset / connector	A governed connector adapter with allowed actions and data classes, Gateway-enforced, with approval required for high-risk actions.
Checkpoint	A saved, resumable state with a rollback policy, so long-running work can pause and resume inside the trust boundary.

The Plan Runtime Wizard

The fastest way to a working runtime is the **Plan Runtime Wizard**, a guided panel at the top of the Studio. It steps you through eight stages and, at each gate, tells you **exactly what to fix**, with a button that takes you straight to where the fix is made.

1. **Prerequisites.** Confirms the runtime is ready: Claw and Gateway configured and reachable, tool audit logging enabled, and a connector in proper credential custody. Each blocker shows a **Fix it** button (for example Open Tool Permissions) and a warning does not block you.
2. **Starting point.** Choose **Bootstrap the governed demo** (provisions a working agent, its tool permissions and a plan that already passes preflight, the fastest path to a first success), start from a **template**, or start **blank**.
3. **Define plan.** Set the plan name, objective, risk level, **root agent** (the orchestration identity), maximum parallel steps and timeout.
4. **Build steps.** Add steps from the task-type palette, set each step's agent, tool (for a Gateway tool call), dependencies and whether it needs approval. A **wave preview** shows the execution order and flags any dependency cycles.
5. **Preflight and fix.** Saving the plan runs the authoritative **server preflight**, which checks every step against exactly what Gateway will enforce at runtime. Anything missing is listed as a categorised checklist, each item with a **clickable fix** that opens the right Agentic CISO screen (with the specific agent pre-selected for per-step issues).
6. **Approve.** The governance sign-off, recorded in the audit trail. It attests the plan passed preflight under zero-trust enforcement.
7. **Dispatch and run.** Sends ready steps to the runtime. See Running a plan.
8. **Done.** A summary of the run, with the option to create another.

Every gate is enforced by the server, not just the wizard: you cannot reach a confirmed, dispatched runtime without passing the real preflight and approval.

Setting up and running a plan

Prerequisites

Before a plan can run, an operator and an administrator make sure the runtime is in place. The wizard's first step verifies all of this for you:

- The **Claw** execution service and the **Gateway** enforcement service are configured and reachable (an operations task, see Claw and Gateway).
- At least one **connector** is configured, with its credential held in Gateway custody (configure these in the Studio's connector catalogue panel).
- Every registered **agent** that a step uses has a named human owner, an assigned ATF level, and the required **Tool Permissions** active with **audit logging enabled**.
- Material action steps have matching **Runtime Access Policy** coverage and the agent is mapped into the **Org Chart** where high-impact escalation is required.
- For any step that needs human sign-off, an **approval gate** exists.

If something is missing, the wizard's checklist links you straight to the screen that fixes it.

Build the plan

1. Open **Workflow Studio** and select your workspace.
2. Use the wizard's **Bootstrap demo** for a guided first run, or pick a template, or build from scratch.
3. In **Define plan**, choose the **root agent** and set parallelism and timeout.
4. In **Build steps**, add each step: choose its **task type**, assign a **registered agent**, set a **tool** for Gateway tool calls, mark **dependencies** (which steps it runs after), and tick **requires approval** for sensitive steps.

Preflight and fix

Save the plan to run preflight. The readiness scorecard shows a score and which steps are ready. For anything not ready, the checklist tells you the precise problem and gives a **Fix it** button:

Blocker	Where the fix-it button takes you
Tool permission not active / not runtime-enabled / audit logging off	Tool Permissions (Agentic CISO), with that agent focused
Runtime Access Policy coverage missing	Access Matrix (Agentic CISO), with that agent focused
Org escalation path missing	Org Chart (Agentic CISO), with that agent focused
Agent or root agent not registered	Agent Register
Approval gate missing	Approval Gates
ATF level	ATF assessment
Connector missing	the Studio's connector catalogue
Draft issues (step keys, dependencies)	the wizard's Build step

Fix, return, and re-check. The plan is ready when preflight passes.

Approve and dispatch

Approve the plan (the governance sign-off), then dispatch. Dispatch sends each **ready step** to the runtime under a signed runtime identity, and Gateway re-validates every action at dispatch, so a step can still be blocked or held for approval at runtime.

Running a plan

Steps run in **waves** by dependency: steps with no unmet dependencies run first, then the next wave once they complete.

- **Auto-run (default on):** the Studio automatically dispatches each next wave as the previous one completes, so the whole plan runs to completion on its own. Turn it off to step through manually, one wave at a time, with **Sync + continue**.
- **Approval-gated steps still wait for a human** on every run, and every Gateway decision still holds, so Auto-run never bypasses governance.
- A **live stream** shows each step's Gateway decisions, model calls and bounded output, backed by a hash-chained Claw journal.

Stopping a running step

A running step is backed by a Claw task. Each active step has a **Stop** button that cooperatively cancels it: Claw stops at the next safe checkpoint, marks the task cancelled and will not half-complete. The step is terminal, but you can re-dispatch the plan afterwards. Stopping is governed and audited.

For a broader halt, an administrator can use the **Gateway kill switch** (admin console, see Gateway), or suspend an individual agent in Agentic CISO.

Scheduling a plan

An **approved** plan can run on a recurring schedule. On the selected plan's **Schedule** panel:

1. Choose a **cadence** (hourly, daily, weekly, monthly or quarterly).
2. Click **Enable schedule**. Enabling is a sign-off and needs an approver-level permission plus an approved plan.
3. **Pause**, **Resume** or **Disable** the schedule at any time. The panel shows the next run and the last run with its outcome.

A schedule is never a standing bypass. On every scheduled run, Gamut:

- re-checks the schedule owner's live permission and the tenant entitlement (and pauses the schedule if either was revoked),
- re-runs the full **preflight**,
- **skips** the run if a previous run is still in flight (no overlap),
- **skips** while the kill switch is active (runs resume automatically once it is cleared),
- and dispatches through the same Gateway/Claw-enforced path, where per-step approvals still hold.

Every scheduled run is audited with its outcome.

Deleting a plan

Each saved plan in the sidebar has a delete control. Deleting removes the plan and its steps (with a confirmation) and is audited. Any Claw tasks already running are governed independently and are unaffected.

Governed by construction

A workflow built in Studio does not bypass the agentic stack. Each action step references a governed connector by tool name, and that call faces the same checks as any other: the tool must be permitted to the agent, Gateway must allow the action within the agent's ATF level boundary, a Runtime Access Policy must cover the action, the org chart must support high-impact escalation, and any approval gates must be satisfied. The design surface changes; the enforcement model does not.

Runtime backends

A workflow targets a registered **runtime backend**, either Gamut Claw or, for external frameworks, a BYO runtime. Backends are governed: they default to a **deny** network posture, **no raw secret mounts**, and **gateway-only egress**. In every case the rule holds: think anywhere, act through Gateway.

Next

129. Runtime control and live approvals

Monitor agents, inspect pending decisions and suspend or stop authority without weakening the fail-closed Gateway boundary.

Agent Runtime Control is the operational view of agent status, pending decisions and governed execution. It does not replace Gateway enforcement. Use it to observe, approve where authorised, and contain operation.

Before enabling operation

- Agent identity and owners are current.
- Risk tier and ATF readiness support the intended autonomy.
- Required connectors are healthy and least-privileged.
- Tool Permissions and Runtime Access Policies are exact and active.
- Approval, logging, monitoring and expiry requirements are configured.

- Simulation and failure-path testing are complete.

Live approvals

Inspect the complete action context: agent, purpose, requested action, target, resource, data class, environment, risk, policy version, expiry and expected effect. Approve only when the request matches the authorised business purpose and remains necessary at decision time.

Approval is not execution. Immediately before action, Gateway revalidates policy, approval, connection, identity and request context. Changed, stale, replayed, expired or already-consumed authority is rejected.

Suspension and containment

Suspend an agent or revoke policy when identity, purpose, owner, connection, evidence or control assumptions are no longer trustworthy. Use the narrowest control that safely contains the issue, but do not delay a broader stop when impact is uncertain.

Record reason, decision owner, affected work, communications and recovery conditions. Test that pending and retried requests cannot continue under revoked authority.

Monitoring

Review denials, unusual request volume, repeated approval prompts, policy mismatches, connector failures, timeouts and evidence gaps. A high denial rate may show malicious behaviour, poor agent planning or an incorrectly designed policy; investigate rather than automatically widening access.

Next

130. Agent risk and ATF Control Tower

Combine agent risk tiering, ATF readiness, tests, incidents and runtime evidence without treating one score as authority to operate.

Agent Risk Tiering prioritises governance based on capability, access, autonomy and potential impact. **ATF Control Tower** brings together per-agent maturity, control evidence and operational signals. Neither view grants runtime access.

Risk tiering

Assess the actual deployment, not the agent's intended personality or job title. Consider tool and data reach, independence, delegation, persistence, external communication, reversibility, privilege, scale and affected people. Record evidence and uncertainty for each judgement.

Use the resulting tier to set approval, testing, monitoring, review and escalation depth. Do not lower the tier merely because controls are planned; evaluate residual exposure after tested controls separately.

ATF Control Tower

Use the Control Tower to identify agents below target maturity, stale evidence, failed promotion gates, overdue reviews, incidents and runtime-control weaknesses. Drill into the agent and requirement before assigning remediation.

ATF readiness is per agent. Portfolio averages can hide one highly privileged unready agent. Promotion requires the evidence and elapsed operating experience specified by the assessment, plus human confirmation.

Decision use

Combine risk tier, ATF result, MAESTRO threats, runtime tests, incidents, open findings and policy scope. Record the human decision and reassessment trigger. A green dashboard is not permission to operate without the exact runtime authority chain.

Next

131. Incident response and red teaming

Prepare, test, contain and learn from agentic failures while preserving evidence and runtime control.

Agentic incident response must address both the harmful outcome and the authority path that allowed or attempted it. Red teaming tests those paths before an attacker, model failure or unsafe workflow does so in production.

Incident workflow

1. **Triage:** identify agent, action, target, environment, affected systems and current exposure.
2. **Contain:** suspend the agent, policy, connection or runtime path necessary to stop further harm.
3. **Preserve:** retain requests, decisions, approvals, runtime evidence, connector responses and relevant business records.
4. **Assess:** determine data, safety, financial, legal, customer and operational impact.
5. **Eradicate and recover:** correct policy, code, credentials, data or process; retest failure paths.
6. **Review:** record causes, control failures, lessons, residual risk and reassessment triggers.

Do not delete the evidence needed to understand the event. Coordinate privacy, legal, security, safety and regulatory notification decisions through authorised people.

Red Team

Define written scope, authorisation, environment, stop conditions and evidence handling before a test. Cover prompt injection, indirect instruction, identity spoofing, permission escalation, policy mismatch, approval manipulation, replay, data exfiltration, unsafe delegation, tool misuse, connector failure and containment.

Test both deny and allow paths. A system that blocks everything is not operationally adequate, and a successful expected action does not prove malicious variants are contained.

Convert confirmed weaknesses into findings, risks, remediation and retests. Record limitations and untested paths rather than describing the exercise as comprehensive when scope was bounded.

Next

132. Agentic reporting and evidence

Produce defensible Evidence Packs, Board Reports, architecture views, red-team reports, Control Tower reports and framework mappings.

Agentic reports translate live governance and runtime records for different decision makers. Select the output whose purpose matches the audience; do not use one report as a substitute for all assurance work.

Output	Primary use
Evidence Pack	Trace controls, policies, approvals, tests and runtime evidence.
Agentic Board Report	Summarise exposure, incidents, readiness and decisions for leadership.
Security Architecture Report	Explain identities, trust boundaries, connections, data movement and enforcement.
Red Team Report	Record authorised scope, scenarios, results, limitations and remediation.
Control Tower Report	Summarise per-agent ATF posture, gates, exceptions and trends.
Framework Mapping	Trace agentic controls to supporting framework requirements without claiming equivalence.

Generation checklist

1. Confirm workspace, systems, agents, period and intended audience.
2. Check that the underlying records and evidence are current.
3. Resolve unexplained scope, policy and incident contradictions.
4. Include open limitations, accepted risks and overdue work.

5 Generate and record the run metadata.

6 Review content and redact unnecessary sensitive operational detail before distribution.

7 Regenerate after material changes rather than editing an old export into a new conclusion.

Interpretation

Counts and maturity summaries support prioritisation but do not demonstrate that a specific action was authorised. Use request-bound Gateway evidence for runtime decisions. Framework mappings show support and reuse opportunities; they do not create compliance inheritance.

Next

Part 5: Administration, access and assurance

This part compiles the current public operating guidance for the workflows in scope.

133. Workspaces & tenancy

Each organisation uses Gamut in an isolated tenant, with assessments as workspaces inside it, keeping AI systems, evidence and users separated by design.

Gamut is multi-tenant. Each organisation operates in its own **tenant**, an isolated environment whose AI systems, assessments, evidence and users are kept separate from every other organisation. Inside a tenant, each **assessment** acts as a workspace that work and membership attach to.

Isolation by construction

Tenant isolation is a deliberate security property, enforced at the data layer rather than by application filtering alone: each tenant's records live in their own normalised database schema. Your governance records, evidence and users are never visible to another organisation, and access is always scoped to the tenant you belong to. See Security & data handling.

What lives in a tenant

- The AI system registry and discovery results.
- Assessments, evidence and findings.
- The agent register and agentic governance.
- Users and roles.
- Tenant configuration, including SSO and entitlements.

Assessment-level membership

Within a tenant, membership and roles can be scoped to an individual assessment. A user carries a tenant role for what they can do organisation-wide, and an additive workspace role, Lead Assessor, Contributor, Reviewer or Viewer, on the specific assessments they are added to. This is how you give someone broad read access but write access only where they actually work.

The **Runtime Policy Approver** role is also workspace-scoped. It reveals submitted Runtime Access Policies only for assigned workspaces and does not grant general editing authority.

View-only workspaces

Locking a workspace in **View only** mode preserves its governance record for inspection while blocking changes. Existing Runtime Access Policies can still be opened and reviewed, including their boundaries and history. Policy drafting, editing and lifecycle actions remain unavailable until the workspace is deliberately unlocked by an authorised user.

Belonging to more than one tenant

Some people work across more than one organisation or environment, for example advisers and auditors. Where that applies, a user can belong to multiple tenants and switch between them, always acting within the one currently selected and never carrying data across the boundary.

Tenant status

A tenant has a status that controls access. An active tenant operates normally; a suspended tenant is locked out entirely until reactivated. Provisioning, suspension and reactivation are audited account lifecycle actions.

Next

134. Users & roles

Gamut governs access with a fixed catalogue of roles across three scopes, enforced server-side as a permission set, combined with plan entitlements, and additive at the assessment level.

Access in Gamut is governed by **roles**. A role maps to a precise set of namespaced permissions (for example `system.update`, `finding.export`, `agent.approve`), and every state-changing action checks for the permission it requires. Roles are not free-form; they are a defined catalogue, enforced the same way in the interface and behind every API call.

Three scopes of role

Roles exist at three scopes, and a user's effective access is the combination.

Tenant roles

What a user can do across the organisation:

Role	Access
Administrator	Full tenant access, user management and the admin console.
Runtime Policy Approver	Approval-only access to inspect and decide submitted Runtime Access Policies in explicitly assigned workspaces.
Advanced	Advanced-tier frameworks, AI analysis, Agentic CISO, simulation and exports; Enterprise runtime and licensed ISO access remain separately gated.
Standard	Core frameworks, bounded AI and workpaper/report capabilities. No control testing, AI Consultant or agentic stack.
Subscriber	Login only. No product access.

Workspace (assessment) roles

Additive permissions scoped to a single assessment, layered on top of the tenant role:

Role	Access within the assessment
Lead Assessor	Full read/write, can share and delete, can approve agents and export.
Runtime Policy Approver	Read supporting agentic context and approve or reject submitted Runtime Access Policies only.
Contributor	Create and update records. Cannot delete or export.
Reviewer	Read access plus notes and comments. Cannot delete.
Viewer	Read-only.

Permission and entitlement: both must pass

A role granting a permission is necessary but not always sufficient. Sensitive product capabilities are gated by **both** an RBAC permission **and** a plan entitlement. A tenant administrator is deliberately not exempt: administration authority never unlocks a capability the tenant has not purchased. So "can this user do X" is answered by two questions: does their role grant the permission, and does the plan include the feature.

Object ownership

Independently of role, the **owner** of an object gets read and update on that object, for systems, risks, findings, evidence and agents. Ownership never escalates to delete or to audit-sensitive actions; it is a deliberately narrow grant so owners can maintain their own records without broad access.

Least privilege

Assign the least privilege needed for someone to do their job. It is easier to grant more access later than to recover from over-sharing, and the additive workspace roles mean you can give someone broad read access at the tenant level and write access only on the assessments they work on.

Runtime Policy Approver

Runtime policy decisions can change the authority exercised by AI agents. Gamut therefore provides a dedicated **Runtime Policy Approver** rather than requiring reviewers to hold an administrator role.

When creating an approver account, the administrator selects the exact workspace the person may review. Additional workspace assignments must be granted explicitly. A policy is visible only when it belongs to an active, assigned workspace.

The role can:

- inspect submitted Runtime Access Policies and relevant supporting context;
- use the pending-approval inbox;
- approve and publish a pending policy;
- reject a pending policy with a recorded reason.

The role cannot:

- draft, edit, delete or submit a policy;
- suspend or revoke active authority;
- create or change agents, tools, connections, approval gates or data flows;
- manage users, roles, workspaces or tenant settings;
- operate Gateway or bypass plan entitlements.

The restriction is enforced behind the interface as well as in it. Account-security actions such as changing the approver's own password and enrolling MFA remain available.

Both approval and rejection require the approver's password and current MFA code through a short-lived step-up verification. The person who created or submitted the policy cannot approve that version. Every assignment and decision is audited.

See Runtime Access Policies for the review checklist and policy lifecycle.

Managing users

Administrators manage users from **Administration -> Users**:

- **Invite** colleagues by email and assign a role. See Invite your team.
- **Adjust** roles as responsibilities change.
- **Assign approvers** to the exact workspace containing the policies they may review.
- **Suspend** a user to immediately revoke access. An inactive user fails every permission check, and suspension ends their active sessions.

Sessions

Administrators can review active sessions and revoke them individually, useful for offboarding or responding to a concern. Sessions also expire on their own over time.

Privileged controls and break-glass

The **admin console** holds the most sensitive controls, behind both a permission gate and a **step-up re-authentication**. They include the runtime break-glass switches:

- **Disable Gateway Runtime** (the Gateway kill switch): blocks every agent action across the tenant, fail-closed, until reactivated. See Gateway.

- **Suspend Agent Runtime**: halts a single agent's actions until it is reactivated.
- Gateway approval review, connector and integration configuration.

Every one of these is audited. They are deliberately kept out of the normal app and behind step-up so a break-glass action is always a deliberate, attributable decision.

Accountability

Granting or revoking a role is itself an audited action, and every state-changing action a user takes is written to the audit log, so access and activity remain accountable and reviewable.

Single sign-on

For organisations with an identity provider, SSO governs authentication (proving who someone is) while roles continue to govern authorisation (what they can do).

Next

135. Account security and MFA

Protect human accounts with MFA, step-up verification, secure sessions and a controlled recovery process.

Every person should use their own account. Shared accounts weaken attribution, separation of duties and incident response.

Enrol MFA

Use the **MFA** control in the sidebar to begin enrolment. Scan the displayed authenticator setup with an approved application, enter the current code and store any recovery material according to your organisation's security process. Do not send setup secrets or recovery codes through chat or email.

MFA enrolment protects future sign-ins and supports **step-up verification** for sensitive actions. Runtime Policy Approvers must enrol MFA before they can approve or reject policies. MFA does not hide their assigned pending policies; it prevents the decision until strong verification is complete.

Step-up actions

Actions such as security changes, write-token creation, trial activation and sensitive approvals may require the current password and/or authenticator code. Reauthentication is deliberately short-lived and applies to the exact operation being performed.

Account lifecycle

- Invite users to the correct tenant and role.
- Verify identity through the approved organisational process.
- Require MFA for privileged and approval roles.
- Review sessions, tokens and role assignments after job changes.
- Suspend access promptly when no longer required.
- Use administrator-supported recovery rather than creating a replacement shared account.

If compromise is suspected

Suspend the account, revoke sessions and API tokens, reset credentials, review the Audit Trail and investigate affected records. Restore access only after identity and device security are confirmed.

Next

136. Single sign-on (SSO)

Connect your identity provider to Gamut with OpenID Connect so people sign in with corporate credentials and access follows your existing identity process.

Gamut supports **single sign-on (SSO)** via **OpenID Connect (OIDC)**, so people sign in with their corporate credentials and your existing identity process governs access.

Why use SSO

- **One set of credentials.** People sign in with the account they already use.
- **Centralised control.** Access follows your joiner / mover / leaver process in your identity provider.
- **Stronger security.** You apply your own MFA and conditional-access policies at the identity provider.

How it works

SSO is configured per workspace by an administrator, using the standard OpenID Connect authorization code flow:

1. In your identity provider, register Gamut as an application and obtain the connection details (discovery URL, client ID and client secret).
2. In Gamut, open the **Administration -> SSO** panel and enter those details.
3. Save the configuration. Once configured, the sign-in page offers a **Sign in with SSO** option.
4. People authenticate at your identity provider and are returned to Gamut, where their account is provisioned from the verified identity.

Connection secrets are stored encrypted. Gamut verifies identity tokens cryptographically on each sign-in.

Configuring SSO

SSO is configured per tenant by an administrator:

1. In your identity provider, register Gamut as an **OpenID Connect application** and note the issuer (discovery) URL, client ID and client secret. Set the redirect (callback) URL to the one Gamut provides for your tenant.
2. In Gamut's tenant administration, enter the **issuer URL, client ID and client secret**, and choose how provider identities map to Gamut users (for example which email domains are allowed and the default role for a newly provisioned user).
3. **Test** a sign-in with a provider account, then enable enforcement so users authenticate through the provider.

Credentials are held server-side, never exposed to the browser, and the sign-in is CSRF- and nonce-protected. Roles and entitlements still apply on top of SSO, see roles with SSO.

Supported providers

Any standards-compliant OpenID Connect identity provider can be used, for example the major enterprise identity platforms. If your provider supports the OpenID Connect authorization code flow with discovery, it will work with Gamut.

Roles with SSO

SSO governs authentication, proving who someone is. Their authorisation in Gamut is still controlled by roles within the workspace.

Next

137. Plans & entitlements

Current Gamut plan pricing, framework access, quotas, billing behavior and server-enforced entitlements, capped by both plan tier and user role.

A workspace's **plan** determines which frameworks, modules, quotas and billing options it can use. **Entitlements** are enforced server-side, not only hidden in the interface. A capability is available only when the workspace plan allows it and the signed-in user's role allows it.

Current plan tiers

Free

- **Price:** $0.
- **Scale:** 1 paid seat, up to 3 AI systems and 3 intake records, 1 active assessment at a time, and up to 3 assessment creations per rolling 12 months.
- **Framework access:** Choose 1 at signup: GTSAF, NIST AI RMF, EU AI Act or NAGF.
- **Included:** Dashboard, Learning Hub, AI System Records, AI Use Case Intake, assessment, risk register, model cards, CSV export and a limited platform-funded AI-analysis allowance.
- **Not included:** AI Consultant, reports, evidence tracker, findings register, policy generation, audit-trail module, Discovery, Agentic CISO, Gateway and Claw.

Standard

- **Price:** $499/month, or $4,990/year.
- **Scale:** 5 paid seats, up to 25 AI systems and 50 assessments per year.
- **Framework access:** GTSAF, EU AI Act, NIST AI RMF and NAGF.
- **Included:** Free capabilities plus policies, AI analysis, bounded policy generation, reports, report export, evidence tracker, findings register, workpaper packs and unlimited free read-only auditor seats.
- **Not included:** Control testing, AI chat, Agentic CISO, Gateway simulation, Gateway, Claw and Discovery.

Advanced

- **Price:** $1,499/month, or $14,990/year.
- **Scale:** 25 paid seats, up to 200 AI systems and 200 assessments per year.
- **Framework access:** GTSAF, EU AI Act, NIST AI RMF, NAGF, MAESTRO, ACRS and ATF.
- **Included:** Standard capabilities plus control testing, AI chat, policy generation, Agentic CISO and Gateway simulation.
- **Not included:** Production Gateway enforcement, Claw runtime execution and Discovery, which remain Enterprise-only.

Enterprise

- **Price:** Starting at $40,000/year, contact us.
- **Scale:** 100 paid seats, effectively unlimited AI systems and effectively unlimited assessments.
- **Framework access:** GTSAF, EU AI Act, NIST AI RMF, NAGF, MAESTRO, ACRS and ATF, plus ISO/IEC 42001 and ISO/IEC 42005 when licence-confirmed.
- **Included:** Full product access, including Discovery, Agentic CISO, Gateway, Claw, production runtime enforcement, OpenID Connect single sign-on and contracted data-residency options.
- **Independent runtime review:** the dedicated Runtime Policy Approver role is available for workspace-scoped policy decisions without granting administrator authority.

  ISO/IEC 42001 and ISO/IEC 42005 are not automatic plan entitlements. They are enabled only after a valid licence confirmation is recorded.

  Standard annual billing saves $998 compared with monthly billing. Advanced annual billing saves $2,998 compared with monthly billing.

Free signup framework choice

New public signups start on **Free**. During signup the user chooses one starting framework from GTSAF, NIST AI RMF, EU AI Act or NAGF. The Free workspace is locked to that selected framework.

GTSAF is the recommended default when the user does not have an immediate framework-specific need: it has the deepest control coverage and maps outward to the other frameworks. If the user needs a specific regulatory or regional view first, they can choose NIST AI RMF, EU AI Act or NAGF instead.

The ACRS score used by intake can still be calculated to route risk. The standalone ACRS assessment module is not included on Free unless the plan is upgraded or a valid boost is active.

Advanced Trial Boost

Free and Standard workspaces can activate a one-time **14-day Advanced Trial Boost** from the billing page. The boost:

- temporarily grants Advanced-tier features;
- includes 50 AI Consultant interactions;
- preserves all data created during the boost after it expires;
- can be used once per workspace;
- requires administrator step-up MFA before activation.

After the boost expires, the workspace returns to its prior plan and the normal entitlements apply.

Quotas

Plan	AI analysis quota	Policy generation quota	Audit retention
Free	20 daily, 20 monthly	0 daily, 0 monthly	90 days
Standard	20 daily, 200 monthly	2 daily, 10 monthly	365 days
Advanced	50 daily, 500 monthly	100 daily, 1,000 monthly	1,095 days
Enterprise	999 daily, 9,999 monthly	999 daily, 9,999 monthly	2,555 days

Policy generation has a separate entitlement and quota. If the `policy_generation` entitlement is off, the policy quota is zero even if general AI analysis is available.

Agentic access

The agentic stack is split deliberately:

- **Advanced** can use Agentic CISO, ATF assessment and Gateway simulations.
- **Enterprise** is required for production Gateway enforcement, Claw dispatch, Claw schedules, BYO runtime credentials, Gateway Event Centre runtime actions and Discovery.

This lets teams design and assess agentic governance before allowing live runtime execution. Runtime-sensitive actions stay behind the Enterprise `gateway` entitlement and are checked again server-side at the access point.

Runtime Access Policy approval is also Enterprise-gated. Assigning an approver role does not create a plan entitlement and cannot make production Gateway enforcement available on another plan.

Two caps, whichever is lower

Effective access is the **intersection** of two limits:

- the workspace **plan tier**;
- the user's **role product ceiling**.

A Standard-role user is capped at Standard capabilities even on an Enterprise workspace. An Advanced-role user can use Advanced capabilities but cannot reach Enterprise-only Gateway or Claw runtime execution. External auditor users are read-only across the assurance surface and do not inherit runtime privileges from the tenant plan.

How gating is enforced

Every gated capability binds a feature entitlement to one or more RBAC permissions. Requests are checked server-side, including API calls, asset access, report and workpaper scopes, framework writes, Gateway actions and runtime dispatch. The UI is not the security boundary.

The access model uses role-based access plus entitlement dual-gating, row-level tenant isolation, fail-closed checks, CSRF protection, rate limiting, step-up MFA for sensitive changes and full audit records for state-changing actions.

Administering your plan

Administrators can review billing and the current plan from **Billing**. Standard and Advanced checkout use secure payment flow. Enterprise is handled by contract and invoice.

> **Note:** The live workspace's returned quota and entitlement payload remains authoritative for that tenant and user.

Next

138. Audit log

Gamut writes an audit entry before every state-changing action returns, with a defined set of high-sensitivity actions that mandate an audit record, giving each workspace a reviewable history.

Gamut records state-changing actions in an **audit log**. It gives every workspace an accountable record of what happened, who did it and when, which is essential for a platform whose whole purpose is defensible governance.

Written before the action returns

Auditing is not an afterthought in Gamut. Every state-changing API handler writes its audit entry before it returns, so a successful change and its audit record are inseparable. The log captures the actions that change governance state, creating or updating AI systems, running assessments, managing evidence and findings, changing users and roles, agentic governance decisions, rather than routine read-only activity.

Actions that mandate an audit record

A defined set of high-sensitivity permissions **require** an audit entry. These are the actions where after-the-fact accountability matters most:

- **Access control**: granting or revoking roles.
- **Risk decisions**: formal risk acceptance, and risk-register export.
- **Findings**: deleting or exporting findings.
- **Exports**: report, workpaper, evidence and audit-log exports.
- **Gateway**: altering enforcement decisions, deleting simulations and changing runtime policy lifecycle state.
- **Policy**: generating, publishing or deleting policy documents.
- **Agents**: deleting an agent, and ATF promotion or approval-gate sign-off.
- **Account and billing**: tenant lifecycle, billing changes and SSO configuration.

Marking these as audit-required in the permission model means the obligation travels with the permission, it cannot be forgotten when a new endpoint is added.

Why it matters

The audit log is part of what makes Gamut governance trustworthy:

- **Accountability.** Every entry is attributable to a user and a time.
- **Reviewability.** Internal audit and reviewers can see how the governance record came to be what it is.
- **Integrity.** A consistent record of changes supports investigation and assurance.

It complements the evidence and findings model: evidence proves the state of controls, while the audit log proves the history of actions. Exporting the audit log is itself an audited action.

Runtime evidence for agents

For agentic AI, the equivalent record is the runtime evidence captured as every agent action passes through Gateway and flows back to Agentic CISO, alongside the tamper-evident, hash-chained journal that Claw keeps per task. Together, the audit log and runtime evidence give a complete account of both human and agent activity.

Runtime policy decision history

Each Runtime Access Policy also exposes its authoritative history. The combined policy history and audit log record:

- draft creation and controlled updates;
- submission for independent review;
- approval and publication;

- rejection and its reason;
- suspension, revocation, expiry or replacement;
- the identity responsible for each decision;
- successful and failed approver-verification events.

Policy content is frozen while pending or authoritative, so the reviewed version cannot be silently edited after the decision.

Next

139. Billing, plans and Trial Boost

Review effective entitlements, manage billing and activate a time-limited Advanced Trial Boost without changing permanent plan rights.

The Billing area shows the workspace plan, quotas and available upgrade or trial actions. The server-returned entitlement profile for the current tenant and user is authoritative.

Before changing a plan

Review required frameworks, system and assessment volume, seats, AI usage, reporting, Agentic CISO, Discovery and production runtime needs. A higher plan expands the tenant ceiling; the user's role still limits effective access.

Advanced Trial Boost

Eligible Free and Standard workspaces can activate one 14-day Advanced Trial Boost. Activation is an administrator action protected by step-up MFA. The boost temporarily grants Advanced capabilities and a bounded AI Consultant allowance; it does not grant Enterprise-only production Gateway, Claw or Discovery.

When the boost expires, records remain but unavailable modules become read-only or inaccessible according to the restored plan. Export or review important outputs before expiry and do not design an operating dependency on a temporary entitlement.

Billing controls

- Use the approved checkout or contractual process.
- Confirm workspace, plan, term and price before payment.
- Restrict billing administration to authorised users.
- Review entitlement state after checkout or contract changes.
- Contact support when payment and effective entitlement do not reconcile.

Next

140. API keys and model providers

Configure personal model-provider keys securely and distinguish them from Gamut API tokens and agent runtime credentials.

My API Keys stores user-authorised model-provider configuration for supported AI features. These keys are different from Gamut bearer API tokens and from Gateway-managed runtime credentials.

Credential	Purpose	Custody
Model-provider key	Run permitted AI Assist, Consultant or generation features.	Stored and used server-side for the authorised user context.
Gamut bearer token	Integrate with supported Gamut API resources.	Created by the user and stored by the external integration.
Gateway connection credential	Invoke an approved external tool or provider at runtime.	Held by the governed Gateway connection, never the agent.

Configure safely

1. Obtain a dedicated organisational provider key where possible.
2. Restrict provider-side project, spend and capability.
3. Add it through **My API Keys**; do not paste it into assessment notes or chat.
4. Run a low-risk test and confirm the selected provider.
5. Monitor usage and rotate according to policy.
6. Remove the key promptly when no longer required or suspected exposed.

The browser does not receive the stored provider secret during ordinary AI use. Entitlement, permission and usage quotas still apply even when the user supplies the provider key.

Privacy decisions

Configuring a key does not itself authorise sending every workspace record. Use scores-only privacy mode for framework AI Assist when detailed context is unnecessary. The AI Consultant uses the permitted workspace context shown on its screen and should not be used where that context is not appropriate for the configured provider.

Next

141. Auditor and independent review access

Invite read-only reviewers and dedicated Runtime Policy Approvers without granting administrator authority.

Gamut provides separate access patterns for assurance review and runtime-policy decisions. Neither requires an administrator role.

Auditor access

Auditors receive a read-only assurance profile constrained by tenant, workspace, plan and role. They can inspect the permitted systems, assessments, risks, policies, evidence, findings, tests, workpapers and audit records, but cannot mutate governance records or export where the role does not permit it.

Invite the reviewer to the exact workspace and period required. Review the invitation status and remove access when the engagement ends. Free-plan review invitations may be time and count limited; paid-plan behaviour follows the current entitlement.

Runtime Policy Approver

The dedicated approver role is for independent decisions on submitted Runtime Access Policies in assigned Enterprise workspaces. The approver can inspect policy context and approve or reject. The role cannot draft, edit, delete, submit, suspend or revoke policy; administer users; operate Gateway; or mutate unrelated product records.

MFA must be enrolled before an approval or rejection. The creator or submitter cannot approve the same policy. The plan must include production Gateway and policy approval; assigning the role does not create an entitlement.

Review checklist

- Correct tenant and workspace assignment.
- Minimum role for the review purpose.
- Individual account and verified identity.
- MFA for sensitive approval.
- No creator-approver conflict.
- Access expiry and removal recorded.

Next

Part 6: Worked scenarios

This part compiles the current public operating guidance for the workflows in scope.

142. Scenario guides

Vertical-agnostic quick-start guides for the most common AI-governance jobs, governing a GenAI chatbot, EU AI Act readiness, shadow-AI discovery, vendor due diligence, board assurance, agentic workflows and ISO 42001 certification.

Where the industry playbooks start from "who you are", **scenario guides** start from "what you need to do". They are short, outcome-focused quick starts for the governance jobs that come up again and again, whatever your industry.

Each guide names the goal, the outcome you will produce, and a numbered path through the real Gamut modules to get there.

Use every scenario defensibly

Before starting, name the system, owner, decision maker, assessment period and intended outcome. During the work, preserve the sources used, unknowns, reviewer decisions and limitations. Finish only when the following are clear:

- the system and use-case boundary;
- applicable routes and unresolved screening;
- evidence and test coverage;
- open findings, risks and remediation;
- the human conclusion and authority for it;
- the events that will trigger reassessment.

Scenario outputs are point-in-time governance records. They do not replace legal advice, certification decisions or an organisation's own risk acceptance.

The guides

Scenario	The job
Govern a GenAI chatbot	Bring a customer- or staff-facing assistant under governance.
EU AI Act readiness	Classify and evidence a system against the EU AI Act.
Shadow-AI discovery sprint	Find unregistered AI and bring it under governance.
Vendor AI due diligence	Assess a third-party AI tool before and during use.
Board assurance pack	Produce a board-ready view of AI governance posture.
Govern an agentic workflow	Put an action-taking agent under runtime control.
ISO 42001 certification readiness	Prepare the AI management system for certification.

How scenarios and playbooks fit together

The two are complementary. Pick the industry playbook that matches your sector for the framework routing and drivers that apply to you, then use a scenario guide for the specific job in front of you. Both run the same governance lifecycle; they are just two ways into it.

Next

143. Govern a GenAI chatbot

A quick-start guide to bringing a customer- or staff-facing GenAI assistant under Gamut governance, from registration and risk tiering to transparency evidence and runtime control.

A GenAI assistant, whether it answers customers or helps staff, is one of the most common AI systems an organisation needs to govern. This guide takes one from unmanaged to governed.

When to use this

You have deployed, or are about to deploy, a chatbot or copilot built on a language model: customer support, internal knowledge, complaint handling, document Q&A or similar.

What you will produce

A registered, risk-tiered chatbot with documented transparency and oversight controls, supporting evidence, and, if it can take action, enforced runtime governance.

Steps

1. **Register it.** Add the assistant in AI System Records with a named owner and a model card capturing the model, provider and intended use.
2. **Run intake.** In AI Use Case Intake, capture purpose, users, data exposure and oversight, and flag public-facing use, personal data and any retrieval (RAG) sources. Confirm the risk tier.
3. **Route it.** Most customer-facing assistants route to GTSAF for depth and the EU AI Act for transparency obligations.
4. **Evidence the key controls.** Through the Evidence Tracker, capture: disclosure that users are interacting with AI, human-oversight and escalation paths, content and safety controls, and data-handling for any retrieved sources.
5. **Decide if it acts.** If the assistant only answers, you are largely done. If it can take action (issue refunds, change records, call tools), govern it as an agentic workflow: register it in Agentic CISO and enforce action through Gateway.
6. **Track and report.** Raise findings for any gaps, work them on the Remediation Roadmap, and report posture via reporting.

Evidence and review checklist

- Intended and prohibited uses, model/provider version and RAG sources.
- User disclosure, safe escalation, complaint and human-review paths.
- Prompt, output, abuse, privacy and security testing for representative users.
- Data retention, provider use, access controls and cross-border processing.
- Accuracy limitations, monitoring signals and incident response.
- Reassessment after model, prompt, retrieval, audience or action-capability change.

The final conclusion should state whether the assistant only advises or can act, which users and data were assessed, unresolved limitations and who approved continued use.

Modules and frameworks involved

AI System Records, intake & risk tiering, GTSAF, EU AI Act, evidence & findings, and the agentic stack if the assistant takes action.

Next

144. EU AI Act readiness

A quick-start guide to classifying an AI system against the EU AI Act in Gamut, routing it to the article-anchored categories, and evidencing the obligations that apply.

The EU AI Act is risk-tiered, so readiness starts with honest classification and ends with evidenced obligations. This guide takes a system from "we think the Act applies" to a defensible readiness position.

When to use this

You need to demonstrate readiness for the EU AI Act for a specific system, or to triage which of your systems the Act touches and how heavily.

What you will produce

A system classified into the correct EU AI Act risk category, with the applicable obligations identified and evidenced, and a readiness report.

Steps

1. **Register and ground the system.** Capture it in AI System Records and run intake, the intake signals (automated decisions, public-facing, personal and special-category data, sector) drive classification.
2. **Route against the Act.** Use the EU AI Act framework to establish scope, Article 5 status, high-risk classification, Article 50 behaviours, GPAI position and other independently applicable routes. These are not one mutually exclusive risk-class selector.
3. **Identify every role.** Provider, deployer, product manufacturer, importer, distributor, authorised-representative and GPAI roles can create different duties. Determine roles from the facts rather than the contract label.
4. **Assess the atomic obligations.** Work every routed item, recording the legal answer separately from its assurance depth and non-applicability governance.
5. **Evidence them.** Through the Evidence Tracker and Testing Centre, capture the evidence each obligation needs: risk management, data governance, transparency, human oversight, accuracy and robustness, and record-keeping.
6. **Add depth where it matters.** For high-risk systems, pair the Act with GTSAF for control depth and ISO/IEC 42005 for an impact assessment.
7. **Track and report.** Work gaps on the Remediation Roadmap and produce a readiness workpaper pack.

Evidence and conclusion checklist

- Legal snapshot, territorial nexus, system boundary and relevant dates.
- Every economic-operator role and independently applicable route.
- Article 5 screen, high-risk route, transparency route and GPAI route considered separately.
- Evidence linked to the correct obligation, system version and assessment period.
- Current law separated from voluntary or future readiness.
- Named legal/compliance reviewer, limitations and reassessment triggers.

Record **compliance conclusion**, **evidence assurance** and **readiness work** separately. A high completion percentage does not by itself establish legal compliance.

Modules and frameworks involved

intake & risk tiering, EU AI Act, GTSAF, ISO/IEC 42005, evidence & findings and reporting.

Next

145. Shadow-AI discovery sprint

A quick-start guide to finding unregistered AI across your organisation with Gamut Discovery and bringing it under governance through review, registration and risk tiering.

The hardest part of AI governance is knowing what exists. This guide runs a focused sprint to surface shadow AI, GenAI tools, vendor AI and emerging agentic workflows in use but never registered, and bring it under governance.

When to use this

You suspect (or know) that AI is being used across the organisation faster than governance can track, and you need an honest inventory before you can govern anything.

What you will produce

A reviewed set of discovered AI, with the real systems promoted into the registry, tiered and routed, and a clear picture of coverage.

Steps

1. **Set up sources.** In Discovery, configure discovery sources, connector-based collectors or manual imports, with an owner and cadence.
2. **Run discovery.** Each run produces **candidates** (suspected AI in use) and **artifacts** (the underlying usage signals), normalised against rules into canonical apps and vendors.
3. **Triage candidates.** Review each candidate's confidence, signal and guessed owner, and decide: promote, dismiss or investigate. Artifacts carry a reconciliation status so each signal is tied to a known asset or flagged.
4. **Promote the real ones.** Promote confirmed candidates into AI System Records, where they enter intake and risk tiering like any other system.
5. **Tier and route.** Run intake on the newly registered systems and route the higher-risk ones to GTSAF and the EU AI Act.
6. **Report coverage.** Use reporting to show leadership what was found, what was registered, and where coverage still has gaps.

Sprint controls and completion criteria

- Define included business areas, sources, accounts, regions and time period.
- Use least-privilege collection and an approved lawful basis.
- Record rule versions, scan health and known detection limitations.
- Give every candidate and alert an attributable disposition.
- Reconcile duplicates before promotion.
- Route promoted systems through normal intake rather than auto-approving them.
- Report unresolved artifacts and uncovered areas.

A zero-candidate run is not proof that no shadow AI exists. The conclusion must explain source coverage, collector health and detector limitations.

Modules and frameworks involved

Registry & Discovery, intake & risk tiering, GTSAF, EU AI Act and reporting.

> **Note:** Discovery is an enterprise capability. Availability depends on your plan & entitlements.

Next

146. Vendor AI due diligence

A quick-start guide to assessing a third-party or vendor AI tool in Gamut before and during use, structured assessment, evidence requests to the vendor, and ongoing oversight.

Most AI an organisation uses is bought, not built. This guide assesses a third-party or vendor AI tool, before adoption and on an ongoing basis, so procurement and risk decisions are structured and defensible.

When to use this

You are evaluating a vendor AI product, or you already use one and need to bring it under proper governance and periodic review.

What you will produce

A registered vendor system with a documented assessment, vendor-supplied evidence captured against controls, and a clear accept/condition/reject position with ongoing review dates.

Steps

1. **Register the vendor system.** Add it in AI System Records, recording the vendor, deployment type and a model card for the vendor model. Mark it as vendor-provided.
2. **Run intake and tier it.** Capture the use case, data exposure and oversight in intake, and confirm the risk tier that sets how deep the diligence should go.
3. **Route to the right frameworks.** Higher-risk vendor systems route to GTSAF and the EU AI Act; for the vendor's own management system, ISO/IEC 42001 evidence is a strong signal.
4. **Request evidence from the vendor.** Use evidence requests to ask the vendor (or the internal owner) for specific artefacts against specific controls: validation, security, data handling, model documentation and assurance certifications.
5. **Assess and decide.** Score the controls with rationale, raise findings for gaps, and reach an accept, accept-with-conditions or reject decision with the gaps tracked on the Remediation Roadmap.
6. **Set ongoing review.** Record review dates so the vendor is reassessed as the product and your reliance on it change, and report the portfolio via reporting.

Minimum diligence evidence

- Contracted purpose, roles, service boundary and subcontractors.
- Model and material component provenance and change notification.
- Data use, retention, location, deletion and training commitments.
- Security, incident, resilience and business-continuity evidence.
- Performance, bias, safety, accessibility and human-oversight evidence.
- Audit rights, termination, data return and exit plan.
- Independent assurance interpreted within its actual scope and period.

Document unavailable evidence as a limitation or condition. A supplier questionnaire response is a claim until corroborated by appropriate evidence, testing or contractual assurance.

Modules and frameworks involved

AI System Records, intake & risk tiering, evidence & findings, GTSAF, EU AI Act, ISO/IEC 42001 and reporting.

Next

147. Board assurance pack

A quick-start guide to producing a board-ready view of AI governance posture in Gamut, inventory, risk, evidence quality, findings and remediation, traceable to the underlying records.

Boards increasingly ask a direct question: are we governing our AI, and can we prove it? This guide produces a concise, defensible board pack that answers it from live records rather than a hand-assembled slide deck.

When to use this

You need to brief a board, risk committee or executive on AI governance posture, and you want a view that is concise for leadership yet traceable back to evidence.

What you will produce

A board pack covering AI inventory, risk picture, evidence quality, open findings, remediation progress and readiness priorities, every figure traceable to its underlying records.

Steps

1. **Confirm the inventory is current.** Check AI System Records is complete, running a discovery sprint first if coverage is in doubt. A board view is only as honest as the inventory beneath it.
2. **Confirm tiers and assessments.** Make sure systems have been through intake and risk tiering and that material systems carry an assessment.
3. **Surface findings and remediation.** Review open findings and the Remediation Roadmap, including overdue items, so the board sees real status, not a clean fiction.
4. **Generate the pack.** In reporting, produce a **Board pack** or **Board summary**: it composes inventory, risk, estate, evidence quality, findings, decisions and roadmap into a leadership view, with agentic posture included where relevant.
5. **Keep it defensible.** Because the pack is generated from connected records and carries run-level traceability metadata, any figure can be followed back to its evidence if challenged.
6. **Optionally accelerate the narrative.** Use the AI Consultant's Board Summary mode to draft executive narrative grounded in the same records.

Board quality checklist

- Coverage denominator, unassessed estate and inventory limitations are visible.
- High and critical exposure, overdue findings and accepted risks are not averaged away.
- Trend changes distinguish scope, completion and genuine control improvement.
- Agentic runtime posture and material incidents are included where relevant.
- Each requested board decision has an owner and deadline.
- Pack ID, cutoff date, source workspace and reviewer are recorded.

Review AI-drafted narrative against the source records. Regenerate the pack after material changes; do not edit an old export into a new reporting period.

Modules and frameworks involved

reporting, AI System Records, intake & risk tiering, evidence & findings, Remediation Roadmap and the AI Consultant.

Next

148. Govern an agentic workflow

A quick-start guide to putting an action-taking AI agent under runtime control in Gamut, register it, set its ATF level, enforce policy through Gateway and execute through Claw.

When AI stops answering and starts acting, calling tools, moving data, making changes, design-time assessment is no longer enough. This guide puts an agentic workflow under enforced runtime governance.

When to use this

You have, or are building, an agent that takes action: it calls tools and APIs, retrieves and writes data, or runs multi-step workflows, whether on Gamut Claw or your own framework.

What you will produce

A registered, ATF-assessed agent designed to act only through Gateway-enforced authority, with provider credentials outside the agent and request-level runtime evidence.

Steps

1. **Register the agent.** Add it to the Agentic CISO agent register with a human owner and a security owner. An unregistered agent is blocked from acting at all.
2. **Set its autonomy.** Assign an ATF level (Intern through Principal). Gateway enforces a different action boundary at each level, from read-only to strategic autonomy.
3. **Score its capability risk.** Run an ACRS assessment to set how tightly to govern, and model the threats its capabilities introduce with MAESTRO.
4. **Authorise its tools.** Grant tool permissions in Agentic CISO and confirm the matching governed connectors exist. An agent can use a tool only when both layers agree.
5. **Define Runtime Access Policies.** Bind each required authority to the exact agent, governed connection, action, resource, data classification, environment, purpose and time boundary. Validate the policy, review its impact and submit the immutable version for independent approval.
6. **Configure approval gates.** Require human approval for sensitive or mutating actions (external calls, financial actions, code changes). Gateway enforces them on every request.
7. **Independently approve runtime authority.** A workspace-scoped Runtime Policy Approver reviews the submitted policy and either publishes it or rejects it with a reason. Approval requires the reviewer's own MFA-protected step-up.
8. **Choose the runtime.** Run it on Gamut Claw or, for an external framework, the BYO runtime. Either way the rule holds: think anywhere, act through Gateway. Credentials live on Gateway, never with the agent.
9. **Simulate, then watch.** Run a Gateway simulation to see what Gateway would decide and close gaps before going live, then rely on the hash-chained runtime evidence and the audit log for ongoing oversight.

Go-live evidence and gates

- Authenticated agent identity, owners and purpose boundary.
- ATF target, ACRS risk and MAESTRO threat treatment.
- Exact Tool Permissions, governed connections and data flows.
- Independently approved Runtime Access Policies with expiry and review.
- Allow, deny, approval, timeout, replay, failure and containment tests.
- Monitoring ownership, incident playbook and kill/suspend procedure.
- Residual risk decision and material change triggers.

Readiness is not standing authority. Gateway revalidates the exact request immediately before execution, and missing or stale context must stop the action.

Modules and frameworks involved

Agentic CISO, Runtime Access Policies, Gateway, Claw, BYO runtime, ATF, ACRS and MAESTRO.

Next

149. ISO 42001 certification readiness

A quick-start guide to preparing your AI management system for ISO/IEC 42001 certification in Gamut, route to the standard, evidence the clauses and Annex A, and produce an audit-ready pack.

ISO/IEC 42001 is the AI management-system standard organisations and their buyers increasingly ask for. This guide prepares your management system for certification by evidencing the standard inside Gamut.

When to use this

You are pursuing ISO/IEC 42001 certification, or a buyer or board has asked you to demonstrate an AI management system aligned to it.

What you will produce

An assessment against ISO/IEC 42001 with each clause and Annex A control evidenced, gaps tracked to closure, and an audit-ready workpaper pack for the certification body.

Steps

1. **Set the scope.** Confirm which AI systems and processes are in scope in AI System Records; the management system governs how you govern them.
2. **Route to the standard.** Create an assessment against ISO/IEC 42001, which is structured as clauses 4 to 10 plus Annex A controls.
3. **Assess each section.** Work through the management-system clauses (context, leadership, planning, support, operation, performance evaluation, improvement) and the Annex A controls, recording maturity and rationale.
4. **Evidence the controls.** Through the Evidence Tracker and Testing Centre, capture the artefacts each clause needs, policies, roles, risk and impact assessments, operational controls, monitoring and review records. The policy generator can draft the policies and procedures the standard expects.
5. **Close the gaps.** Raise findings for shortfalls and work them on the Remediation Roadmap ahead of the audit.
6. **Produce the audit pack.** Generate a workpaper pack that traces each control conclusion to its evidence, the package a certification body can test against.

Readiness evidence and audit questions

- Approved AIMS scope, interested parties and exclusions.
- AI policy, objectives, roles, competence and communication.
- Risk and impact assessment methods applied to the in-scope portfolio.
- Statement of Applicability with inclusion and exclusion rationale.
- Operational procedures and records showing they are followed.
- Monitoring, internal audit, management review, corrective action and improvement.
- Evidence period, sampling limitations and open nonconformities.

Use a licensed copy of ISO/IEC 42001 and confirm certification scope with the selected certification body. Gamut supports readiness and traceability; it does not issue the certificate or guarantee an audit result.

Modules and frameworks involved

ISO/IEC 42001, AI System Records, evidence & findings, policy generation, Remediation Roadmap and reporting.

Next

Part 7: Integration and API guidance

This part compiles the current public operating guidance for the workflows in scope.

150. API overview

Integrate Gamut with your stack using its HTTP API, authenticate with bearer tokens and work with governance data programmatically.

Gamut exposes an HTTP API so you can integrate governance into the rest of your stack, automating inventory, assessments, evidence and reporting rather than doing everything by hand.

> **Note:** API capabilities depend on your plan & entitlements and the role of the token's owner. A token can only do what its owner could do in the interface.

Base URL

The API is served from your Gamut workspace under a versioned prefix:

```
https://run.gamutassure.com/api/compass/v1/
```

All endpoints sit beneath this prefix. The version segment (`v1`) lets the API evolve without breaking existing integrations.

Authentication

Programmatic access uses **bearer tokens**, named and revocable API tokens tied to a user, a workspace, an access level and an expiry. Pass the token in the `Authorization` header:

```
Authorization: Bearer <your-token>
```

See Authentication for creating, using and revoking tokens.

What you can do

The API exposes supported governance resource families described in the resource guide. Not every interactive or administrative operation is a supported public API. Subject to the documented resource, permissions and entitlements, integrations can:

- Keep your AI inventory in sync with other systems.
- Drive or retrieve assessment data.
- Manage evidence and findings programmatically.
- Pull data for external reporting.
- Use workspace-scoped API access within the token's read-only or read/write boundary.

Two authentication modes

The API has two distinct authentication paths for two distinct callers:

- **Bearer tokens** for user-scoped automation against the governance objects above. A token can do only what its owner could do in its bound workspace, within its access level and before its expiry. Token administration remains an interactive account operation and is not available through a bearer token. See Authentication.
- **Signed service requests** for external agent runtimes, under `/api/compass/v1/external-agent/`. These endpoints (context, model, tool invoke, child-agent request, heartbeat, result) use authenticated, anti-replay service requests rather than user bearer tokens, and every action is brokered through Gateway. This is the path the BYO agent runtime SDK uses; agents act through it but never receive provider credentials.

Conventions

The API follows consistent conventions for requests, responses and errors. See Conventions & errors.

Next

151. API authentication

Authenticate to the Gamut API with named, revocable bearer tokens. Create, list, use and revoke tokens, and keep them secure.

The Gamut API authenticates with **bearer tokens**, named, revocable and expiring credentials tied to a user and an accessible workspace. They let automation and integrations act on the API without a browser session while retaining the same tenant, role and entitlement boundaries.

Creating a token

Create a named token from your account, select the exact workspace and choose a read-only or read/write scope. A token inherits the role and entitlements of the user who created it: it can do what that user can do within the selected workspace, and no more.

Write-enabled tokens require the user to re-enter their current password. Where MFA is enrolled, the current authenticator code is also required. This prevents an unattended browser session from silently creating durable write authority.

> **Shown once:** A token's secret value is shown **only once, at creation**. Copy it immediately and store it in a secrets manager. If you lose it, revoke the token and create a new one.

Using a token

Send the token as a bearer credential on each request:

```
GET /api/compass/v1/... HTTP/1.1
Host: run.gamutassure.com
Authorization: Bearer <your-token>
```

Bearer-authenticated requests are intended for server-to-server automation. Do not embed tokens in browsers, mobile apps or any client where users could extract them.

Managing tokens

You can:

- **List** your tokens to see their name, workspace, scope, expiry and recent use.
- **Revoke** a token immediately when it is no longer needed or may be exposed.

  Revoking a token takes effect at once, any integration using it will stop being authorised.

Keeping tokens secure

- **Store in a secrets manager**, never in source control.
- **Scope by user.** Create tokens under a user whose role matches what the integration needs, apply least privilege.
- **Scope by workspace.** Use a separate token for each workspace boundary.
- **Prefer read-only.** Grant write scope only when the integration must change governance data.
- **Use a short lifetime.** Set an expiry appropriate to the integration and rotate before it.
- **Rotate** tokens periodically, and immediately if exposure is suspected.
- **One token per integration**, so you can revoke one without disrupting others.

Sessions vs. tokens

Interactive use of Gamut in the browser uses a signed-in session; programmatic use uses bearer tokens. For human sign-in, including single sign-on, see Users & roles.

Bearer tokens cannot manage account credentials, create other tokens or change their own security boundary. Account suspension, password reset and other qualifying account-security events revoke existing programmatic authority so retained credentials cannot outlive the account decision.

Next

152. Conventions & errors

Request and response conventions for the Gamut API, JSON, versioning, rate limits, status codes and error handling.

The Gamut API follows consistent conventions so integrations are predictable. This page covers the request and response shape, status codes, rate limiting and error handling.

Requests and responses

- **JSON.** Requests and responses use JSON. Send `Content-Type: application/json` on requests with a body.
- **HTTPS only.** All requests are over HTTPS. Plain HTTP is not accepted.
- **Versioned path.** Endpoints live under `/api/compass/v1/`. The version segment lets the API evolve without breaking existing integrations.
- **Authentication.** Send a bearer token on every request, see Authentication.

Status codes

The API uses standard HTTP status codes:

Code	Meaning
`200` / `201`	Success.
`400`	The request was malformed or failed validation.
`401`	Missing or invalid authentication.
`403`	Authenticated, but not permitted, check the token owner's role and entitlements.
`404`	The resource does not exist, or is not visible in your workspace.
`429`	Rate limit exceeded, slow down and retry later.
`5xx`	An unexpected server-side error.

Errors

Error responses return a JSON envelope with a machine-readable code, a human-readable message and the HTTP status:

```
{
"code": "quota_exceeded",
"message": "Daily limit reached (20/20). Resets at midnight UTC.",
"status": 429
}
```

Handle errors by checking `code` (stable, safe to branch on) rather than parsing the message (human-facing, may change). Treat `4xx` as something to fix in your request, and `5xx` as transient unless it persists.

Rate limiting and quotas

AI-assisted endpoints are bound by your plan's usage quotas, enforced server-side. Two limits apply independently:

- a **daily** limit that resets at midnight UTC, and
- a **monthly** limit that resets on the 1st.

Exceeding either returns `429` with code `quota_exceeded` and a message stating the current usage and reset point. Policy generation carries its own separate daily and monthly quotas. Back off and retry after the stated reset rather than retrying immediately.

Listing and limits

List endpoints accept a `limit` query parameter to bound the number of records returned, and relevant filters (for example `workspace_id`, `status`) as query parameters. Request only what you need.

Idempotency, retries and concurrency

Do not assume every write is safe to repeat. Retry network and `5xx` failures only with bounded backoff and only when the operation is known to be idempotent or the current contract provides a request/idempotency mechanism. Do not retry `400`, `401`, `403` or validation failures without changing the underlying request or authority.

For updates, re-read the current record and use returned version or update metadata where the resource supports it. Route conflicting governance edits to review rather than silently applying last-write-wins. External-agent operations have their own signed, request-bound anti-replay contract; never reuse stale signed material.

Validation and data minimisation

Send only documented fields and the records needed for the integration purpose. Do not place credentials, access tokens or unrestricted evidence payloads in notes or logs. Treat server-side validation as a second boundary, not a replacement for client validation.

CSRF

CSRF protection applies to **cookie-based browser sessions**, which must send the per-session token as an `x-csrf-token` header (or form field) on state-changing requests. **Bearer-token** API calls are not cookie-authenticated and do not need a CSRF token; authenticate with the `Authorization` header instead.

Permissions and scope

Every call runs within the token owner's workspace and is subject to their role. You will only ever see and affect data in that workspace, isolation applies to the API exactly as it does in the interface.

Next

153. Supported API resources

Plan integrations around supported Gamut resource families while respecting workspace, role, entitlement and record-lifecycle boundaries.

The versioned API exposes supported resource families for authorised integrations. Availability is determined by the current product version, token scope, user role and tenant entitlement. Do not depend on private browser endpoints or administrator routes discovered through inspection.

Resource families

Customer integrations commonly work with:

- workspace context available to the token owner;
- AI systems and intake records;
- assessment records and framework scores;
- risks, evidence, tests and findings;
- model cards, remediation and reporting artifacts;
- user-owned tokens and account-security operations where explicitly supported;
- entitled Discovery and agentic governance resources;
- exports permitted by the user's role.

Use the product documentation for the lifecycle and meaning of each record. The API does not make a field optional merely because an HTTP request can be accepted; governance completeness still depends on the workflow and human decisions.

Integration design

1. Create a dedicated integration user with the minimum role.
2. Bind a short-lived token to one workspace and prefer read-only.
3. Read the current context and entitlement before attempting a feature.
4. Use stable record identifiers, not display names, for reconciliation.

5 Validate input and handle partial or incomplete governance states.

6 Preserve returned identifiers and timestamps for traceability.

7 Re-read the record after a write rather than assuming local state is authoritative.

Unsupported dependency warning

UI routes, internal administration functions and undocumented response fields may change without public API compatibility guarantees. If a required operation is not documented as supported, contact Gamut before building a production dependency.

Next

154. External agent API

Connect a bring-your-own agent runtime through signed, anti-replay service requests and Gateway-mediated context, model and tool operations.

External agent runtimes use a dedicated signed-service API, not a human bearer token. The current machine-readable contract is available from the external-agent OpenAPI document in the authorised Gamut environment.

Operation families

Operation	Purpose
Context	Request bounded workspace context through the governed context adapter.
Model	Request Gateway-mediated model inference.
Tool invocation	Request one exact registered adapter operation.
Child-agent request	Request bounded delegated identity when parent policy permits delegation.
Heartbeat	Record runtime liveness and version metadata.
Result	Record bounded result metadata without sending secrets or unrestricted payload dumps.

Security model

Requests are authenticated, integrity-protected, time-bounded and protected against replay. The runtime identity, tenant, agent and request context must match current governance. A valid signature does not authorise an action by itself: Gateway still evaluates identity, connection, Tool Permission, Runtime Access Policy, approval and live request context.

Client responsibilities

- Keep service credentials in a secrets manager.
- Generate a fresh request identifier, timestamp and signature for each request.
- Follow the exact canonical signing contract from the current OpenAPI/SDK version.
- Do not retry an action with stale signed material.
- Treat context, connector results and user content as untrusted input.
- Do not send provider credentials, full secret-bearing payloads or unnecessary personal data in result metadata.
- Handle denials and dependency failure as stops, not advisory warnings.

Change management

Pin and test the supported SDK or contract version. Revalidate after runtime, agent identity, connection, policy, model or tool changes. Use a non-production environment before enabling live actions.

Next

155. Integration patterns and operational safety

Build resilient Gamut integrations with least privilege, reconciliation, bounded retries and auditable change handling.

Recommended patterns

Inventory synchronisation

Use a stable external identifier and reconcile before creating. Treat Gamut's returned record ID as the governance identity. Send only changed, validated fields and route conflicts to human review.

Evidence ingestion

Send metadata, provenance and links needed to assess the claim. Avoid unrestricted data dumps. An uploaded artifact still requires reviewer sufficiency and scope decisions.

Reporting extraction

Read from a defined workspace and cutoff time. Preserve generation metadata and explain that the extract is point-in-time. Do not combine tenants or silently discard unassessed records.

Event-driven follow-up

Make handlers idempotent. Store the event or request identifier, detect repeats and use bounded backoff for transient failure. Do not retry validation or permission failures as though they were temporary.

Concurrency and reconciliation

Before overwriting a record, confirm that it has not materially changed since it was read. Where a supported resource returns version or update metadata, use it to detect conflicts. Prefer a visible conflict requiring review over last-write-wins for governance conclusions.

Production checklist

- Dedicated minimum-role integration identity.
- One token per workspace and purpose.
- Expiry, rotation and revocation procedure.
- Input validation and output encoding.
- Timeouts and bounded retries.
- No secrets or tokens in logs.
- Tenant and record identifiers checked on every operation.
- Monitoring for authentication, permission, quota and reconciliation failures.
- Tested recovery and decommissioning procedure.

Next

Part 8: Reference and support

This part compiles the current public operating guidance for the workflows in scope.

156. Security & trust

Gamut is secure by construction, tenant isolation at the database-schema level, authenticated encryption, zero-trust agentic enforcement, MFA and role plus plan access control, governed server-side AI, and continuous security review.

Gamut governs your most sensitive AI risk and evidence, so security is a **design property of the platform, not a feature bolted on**. Trust is the product. This page sets out how Gamut is built to protect your data and how we hold ourselves accountable for it.

Security at a glance

Area	What Gamut does
Tenant isolation	Each organisation's data lives in its own database schema, with row-level tenant context checks and fail-closed enforcement on protected data paths.
Encryption	Sensitive secrets are protected with authenticated **AES-256-GCM** encryption at rest; all traffic is encrypted in transit over HTTPS with HSTS in production.
Access	Multi-factor authentication, step-up MFA for sensitive changes, single sign-on (OpenID Connect), role-based access and entitlement dual-gating.
Agentic AI	**Zero-trust:** agents hold no credentials, exact Runtime Access Policies require controlled approval, and every action is revalidated and enforced through Gateway, fail-closed.
AI handling	Model provider keys never reach the browser; all AI calls are governed server-side.
Accountability	Audited state changes and tamper-evident runtime evidence for governed agent actions.
Application hardening	Rate limiting, CSRF protection, content security policy and other production browser-hardening headers.
Assurance	Continuous security review, including dependency, static and authorised dynamic testing, with results assessed as part of release readiness.

Tenant isolation

Each organisation operates in its own tenant, and isolation is enforced at the data layer: each tenant's records live in their own database schema, not merely filtered by an application query. One organisation's AI systems, assessments, evidence and users are kept strictly separate from every other, and access is always scoped to the tenant a user belongs to. Protected data paths also bind tenant context and fail closed when that context cannot be verified.

Encryption

Sensitive secrets and credentials are protected with **AES-256-GCM**, an authenticated cipher, so stored secrets cannot be silently read or altered. The platform will not run in production without its encryption key, so secrets are never stored in plaintext. Passwords are stored as salted, one-way **scrypt** hashes and are never recoverable. All traffic is encrypted in transit over HTTPS, with HSTS enforced.

Authentication and access

People sign in with a password or via single sign-on using OpenID Connect, so organisations can apply their own multi-factor and conditional-access policies. Gamut also supports its own multi-factor authentication, with **step-up verification** required before security-sensitive changes.

Access is governed by role-based access control and plan entitlements: a sensitive capability requires **both** the role permission and the plan entitlement, enforced server-side on every action. Administrators are deliberately not exempt from entitlement checks. Suspending a user or tenant revokes access immediately. Browser sessions carry CSRF protection, programmatic access uses named, revocable API tokens, and responses carry modern browser-hardening headers including a strict content-security policy. Production also applies rate limiting and HSTS.

API tokens are bound to an accessible workspace, limited to read-only or read/write scope and time-bounded. Creating write authority requires reauthentication, and account-security events invalidate existing sessions and programmatic credentials where continued access would no longer be appropriate.

Zero-trust agentic AI

Gamut governs AI that takes action, not just AI that answers questions, and it does so on a zero-trust basis. **Agents never hold credentials and never call tools directly.** Every agent action passes through Gamut Gateway, where policy is enforced and provider keys live, and the execution layer (Claw) redacts sensitive values before any output leaves it. Enforcement is **fail-closed**: if a decision cannot be made or verified safely, the action does not proceed. See the agentic stack overview.

Runtime Access Policies bind authority to an exact agent, governed connection, action, resource, data class, environment and time boundary. Policy creation and policy approval are separated. Immediately before execution, Gateway confirms that the policy, approval and request context remain current and refuses stale, changed, expired or replayed authority.

Governed, server-side AI

All AI analysis is proxied **server-side**. Model provider keys are never exposed to the browser, and prompts and responses are handled by Gamut rather than sent directly from a user's device to a model provider. This keeps model usage governed and credentials protected.

What is sent depends on the feature:

- **Framework AI Assist, full context:** selected system or agent scope, the relevant assessment item, structured scores/statuses and the authorised notes, evidence metadata and routing context required for that analysis.
- **Framework AI Assist, scores-only privacy mode:** direct identifiers and free-text context are excluded; the model receives the minimum structured assessment and route signals required for the bounded task.
- **AI Consultant:** permitted workspace records and free text shown by the on-screen context manifest. Scores-only privacy mode does not narrow Consultant context.
- **Generated narratives:** use the context described on the generating screen; review the scope before submitting.

AI output remains advisory. It cannot approve scope, applicability, evidence, risk acceptance, policy or the human assessment conclusion. Avoid entering secrets and unnecessary personal data, and apply the configured model provider's contractual and organisational handling requirements.

Accountability by default

State-changing actions are recorded in an audit log, and governed agent actions generate **tamper-evident runtime evidence**. Together these provide a reviewable record of relevant human and agent activity for the configured scope and retention period.

Our security practice

Security is part of how we build, not an afterthought:

- **Secure by construction.** The controls above are architectural, isolation, least privilege, fail-closed enforcement and encrypted secrets are properties of the design.
- **Continuous security review.** We conduct regular security review, including static analysis and authorised dynamic testing, as part of our release process.
- **Least surface area.** Gamut runs on a deliberately minimal dependency footprint to reduce supply-chain and attack surface.
- **Dependency hygiene.** Dependencies are reviewed as part of release readiness. A point-in-time scan is evidence for that release, not a permanent guarantee; newly disclosed issues are assessed and treated according to risk.

Independent validation

We are committed to validating our security posture independently, not just asserting it. We are **commissioning independent third-party penetration testing** and **pursuing recognised certifications** (such as SOC 2 and ISO/IEC 42001) as we scale. Organisations evaluating Gamut can request our current security overview through their account contact.

Reporting a vulnerability

We welcome reports from security researchers and customers. If you believe you have found a security issue, please see our vulnerability disclosure policy for how to report it safely and what to expect from us.

Next

157. FAQ

Common questions about Gamut AI, what it is, who it is for, the frameworks it supports, agentic governance, security and getting started.

What is Gamut AI?

Gamut is an AI governance lifecycle platform. It helps organisations discover the AI they use, classify its risk, assess it against multiple frameworks, and produce audit-ready evidence. See What is Gamut AI.

Who is Gamut for?

Organisations running AI governance, across risk, compliance, security and procurement, and the auditors who review them. It is built for organisations that need to demonstrate structured governance, including banks, financial institutions and regulated enterprises. See Who Gamut is for.

Which frameworks does Gamut support?

GTSAF, the EU AI Act, NIST AI RMF, ISO/IEC 42001 and 42005, NAGF, ACRS, the Agentic Trust Framework (ATF) and MAESTRO. A single AI system can be assessed against one or several. See Frameworks overview.

How is a system's risk tier decided?

Intake captures structured risk signals, and a deterministic governance weighting profile, six weighted dimensions with configurable thresholds and floor rules, turns them into a risk tier. The same inputs always produce the same tier, and the reasoning is traceable. See Intake & risk tiering.

Does Gamut give legal advice or certify compliance?

No. Gamut's framework references help you organise governance work and are not endorsements by framework owners. The product does not replace licensed standards, legal advice, certification audits or regulator guidance. See Frameworks overview.

What is the difference between GTSAF and the regulatory frameworks?

GTSAF is Gamut's native assurance baseline, 358 controls across 17 domains, for depth. The regulatory frameworks (EU AI Act, NIST AI RMF, ISO) map Gamut's workflow to external regimes. Evidence crosswalks between them. See GTSAF.

Which framework should I choose when I create a Free workspace?

Choose GTSAF unless you have an immediate need to start with NIST AI RMF, EU AI Act or NAGF. GTSAF has the deepest coverage and maps outward to other frameworks. Free workspaces are locked to the one framework selected at signup, with additional frameworks available on paid plans. See Plans & entitlements.

How does Gamut govern AI agents?

Through the agentic stack: Agentic CISO governs agents, Gateway enforces policy on every action, and Claw executes work only through Gateway. Agents never hold credentials or call tools directly. See the agentic stack overview.

Can I use my own agent framework?

Yes. The BYO agent runtime lets you run external frameworks (LangGraph, CrewAI, Hermes, OpenClaw and custom code) while Gamut remains the trust and enforcement plane, think anywhere, act through Gateway.

What happens if an agent is not registered?

It is blocked. Gateway refuses any action from an agent that is not in the Agentic CISO register, at critical severity. An agent's autonomy is also capped by its ATF level (Intern through Principal), so it can only act within that boundary.

How are agents scored for risk?

By ACRS, which scores an agent's capability across dependency, action, access and harm to set how tightly to govern it, and by an ATF assessment for trust and autonomy readiness.

What happens when routing facts are incomplete?

Unknown material facts produce **screening required** or a routing-incomplete state; they do not silently exclude frameworks or controls. Complete the named system and intake facts, resolve conflicts, review ACRS and then confirm the route. See Routing & applicability.

Why can I see a system in a framework selector but not select it?

The system exists, but a prerequisite such as saved intake, confirmed routing, ACRS review, framework licence or module-specific scope may be incomplete. Follow the displayed reason rather than entering evidence under the all-controls catalogue. See System selection in assessments.

What does privacy mode send to AI?

For framework AI Assist, privacy mode sends bounded structured scores, statuses and route signals without direct identifiers or free-text records. It does not change the assessment or route. The AI Consultant is different and uses the permitted workspace context shown on its screen. See Security & trust.

How can someone approve a runtime policy without becoming an administrator?

Assign the dedicated Runtime Policy Approver role to the exact workspace. It provides approval-only access to submitted policies and supporting context. MFA is required for a decision, and the creator or submitter cannot approve that version. See Auditor and independent review access.

Is my data isolated from other organisations?

Yes. Each organisation operates in an isolated tenant with its own database schema, not just query filtering. See Security & data handling.

Is my data encrypted?

Sensitive secrets and credentials are encrypted at the application layer with AES-256-GCM, and the server refuses to start in production without its encryption key. Passwords are scrypt-hashed. See Security & data handling.

What actions are recorded?

State-changing actions are written to the audit log before they return, and a defined set of high-sensitivity actions (role grants, exports, policy and agent decisions, billing, and more) mandate an audit record. Agent actions also generate hash-chained runtime evidence.

How is AI analysis handled?

All AI analysis is proxied server-side, so model provider keys are never exposed to the browser. The AI Consultant and other AI features are gated by role and entitlement and metered by daily and monthly quotas. Framework AI Assist can use scores-only privacy mode. The AI Consultant is different: it uses the permitted workspace context shown in its context manifest, so do not use it when that wider context should not be sent to the configured model provider. See Security & data handling.

What plan tiers are there?

Free, Standard, Advanced and Enterprise are the customer-facing plans. Free is $0 and includes one selected framework. Standard is $499/month, or $4,990/year, and includes GTSAF, EU AI Act, NIST AI RMF, NAGF, reports, evidence tracker, findings register, bounded policy generation, workpaper packs and unlimited free read-only auditor seats. Advanced is $1,499/month, or $14,990/year, and adds MAESTRO, ACRS, ATF, Agentic CISO, Gateway simulation, AI chat and control testing, with higher AI and policy-generation allowances. Enterprise starts at $40,000/year and adds production Gateway enforcement, Claw, Discovery, OpenID Connect single sign-on, contracted data-residency options and bespoke scale. A capability needs both the plan entitlement and the user's role permission. See Plans & entitlements.

Is there a trial?

Free and Standard workspaces can activate a one-time 14-day Advanced Trial Boost from billing. It temporarily grants Advanced-tier features, includes 50 AI Consultant interactions, requires administrator step-up MFA and preserves data after the boost expires. See Plans & entitlements.

Does Gamut support single sign-on?

Yes, via OpenID Connect. See Single sign-on.

Is there an API?

Yes, an HTTP API authenticated with bearer tokens. See the API overview.

How do I get started?

Follow the Quickstart: sign in, register an AI system, run intake, and complete your first assessment.

How do I get help?

See Support.

158. Support

How to get help with Gamut AI, access the platform, talk to the team, and find the right documentation.

This page points you to the right place for help with Gamut AI.

Using the platform

Gamut runs at **run.gamutassure.com**. If you do not yet have access, ask your organisation's administrator to invite you, or start a trial from the sign-in page.

Find it in the docs

Most questions are answered here:

- New to Gamut? Start with What is Gamut AI and the Quickstart.
- Working with a framework? See Frameworks overview.
- Governing agents? See the agentic stack overview.
- Integrating? See the API overview.
- Common questions are in the FAQ.

Use the search box at the top of any page to jump straight to a topic.

Before contacting support

Record the information needed to investigate without exposing secrets:

- tenant and workspace name;
- affected module and system record;
- approximate time and time zone;
- action attempted and visible error message;
- whether the problem affects one user or several;
- browser and deployment environment;
- a redacted screenshot or request correlation detail where available.

Do not send passwords, MFA setup secrets, API tokens, model-provider keys, Gateway credentials, private keys or unredacted sensitive evidence by email.

Common first checks

- Confirm the correct tenant, workspace and system.
- Check save status and reload the current record.
- Verify the user's role and workspace plan entitlement.

- Complete MFA where the action requires step-up verification.
- For assessment selection, review intake, routing and ACRS prerequisites.
- For AI features, verify provider configuration and displayed quota.
- For runtime actions, inspect identity, connection, permission, policy, approval and Gateway status.

  Avoid repeatedly retrying a state-changing action if the result is uncertain. Check the record and Audit Trail first to prevent duplicate work.

Talk to the Gamut team

For questions the docs do not cover, plans and enterprise capabilities, security and assurance detail, or reporting a concern, contact the Gamut team:

- **Website:** gamutassure.com
- **Email:** hello@titai.org

> **Note:** Confirm any service commitments with your Gamut account representative; the address above is a general starting point.

Security concerns

If you believe you have found a security issue, please report it responsibly via our vulnerability disclosure policy. For our overall security posture, see Security & trust.