Meridant

AI Threat Readiness Assessment

Assessment Report

Albatross Financial Services Limited

Financial Services

Date:	18 May 2026
SKU:	AI Threat Readiness v1.0
Country:	New Zealand
Survey window:	15 April 2026 – 1 May 2026

Letter of Assessment Summary

18 May 2026

Albatross Financial Services Limited Financial Services Wellington, New Zealand

Letter of Assessment Summary

Based on the responses collected through the Meridant AI Threat Readiness Assessment v1.0, Albatross Financial Services Limited has self-attested to the maturity of its AI threat protection controls as summarised in this report. The assessment was completed on 1 May 2026 and scored against the AI Threat Readiness BARS rubric over six NIST Cybersecurity Framework v2.0 functions (Govern, Identify, Protect, Detect, Respond, Recover).

Summary

The organisation achieved an overall AI Threat Readiness score of **3.27** / **5.00** across 24 capabilities.

NIST CSF Function	Average score	Target	Threshold achieved
Govern	3.25 / 5.00	3.50	No
Identify	3.25 / 5.00	3.75	No
Protect	3.60 / 5.00	3.80	No
Detect	3.00 / 5.00	3.25	No
Respond	3.50 / 5.00	3.50	Yes
Recover	3.00 / 5.00	3.33	No

Methodology

Each of the 24 capabilities was scored on a Behaviourally Anchored Rating Scale (BARS) from L1 (Not Defined) to L5 (Optimised). Respondents selected the anchor that most closely matched observed behaviour. Capability scores are the arithmetic mean of all scored responses; N/A responses are excluded from the count. Function scores are the mean of capabilities mapped to that function via `sku_domain.nist_function_code`. The overall score is a weighted mean across all capabilities. Cross-references to NIST AI RMF, NIST SP 800-53, NIST SP 800-30, NIST SP 800-218 SSDF, CSA MAESTRO, OWASP LLM Top 10, and MITRE ATLAS are provided in the appendix.

Limitations

- Results are self-reported by the assessed organisation. No external verification was performed.
- Scores reflect a point-in-time snapshot as at the survey completion date (1 May 2026).
- Recommended re-assessment cadence: annually, or upon material change to the AI estate.

Scope of Assessment

The assessment covered Albatross Financial Services' technology estate supporting retail banking, business banking, and treasury operations across the Wellington (HQ), Sydney, and London offices. In scope: corporate IT infrastructure, customer-facing digital channels (web banking, mobile app), the customer virtual assistant (production), the AI-assisted credit decisioning pilot (pre-production), internal developer AI tooling (GitHub Copilot Enterprise), and the claims summarisation pipeline. The Pelican Trustee Services subsidiary is governed under a

separate certification programme and is excluded. Third-party SaaS platforms are included where Albatross holds data ownership; vendor-side infrastructure is referenced via supply-chain controls but not directly tested.

Executive Summary

Albatross Financial Services is a mid-maturity organisation with strong response and identity foundations but a widening exposure gap on AI-specific risk surface. The board has acknowledged the AI threat landscape but is not yet engaging on a measurable cadence, the risk appetite statement still pre-dates the AI era, and AI-augmented detection and response capabilities are early-stage. Foundational hygiene — patch cadence, configuration management, MFA coverage, immutable backups — is broadly in line with the credible 2026 bar for a regulated financial-services organisation of this size, but each of these classical capabilities currently treats AI-supporting infrastructure as adjacent rather than in scope.

The single function meeting its target threshold is **Respond** (3.50 / 3.50). Decision rights and escalation are unusually clear for an organisation of this size, reflecting recent investment after the 2025 industry-wide regulatory review. SOAR maturity and tabletop frequency are credible at L3, though neither has been exercised against an AI-augmented attack scenario.

The largest gaps are concentrated in **Detect** (AI-augmented triage at L2, behavioural detection at L3 with limited coverage of AI-system telemetry) and **Recover** (business continuity has not been tested against an AI-attack scenario). On the Govern side, the **Risk Appetite Recalibration** capability scored L2 — a written acknowledgement that "AI changes things" exists but quantitative thresholds have not yet been set. This is the single highest-leverage gap in the assessment: virtually every downstream control investment depends on a credible answer to "how much AI-era risk does the organisation accept?"

The principal recommendations, in priority order: (1) commission a board-led risk-appetite reset within 90 days; (2) extend detection telemetry into the model-gateway and inference layer and pilot AI-augmented triage on a focused use case; (3) run a tabletop exercise against an AI-augmented social-engineering scenario and feed the findings into both the response playbook and the BC test plan; (4) tighten third-party AI risk monitoring on the four highest-tier model and SaaS providers. Detailed phasing follows in the Roadmap section.

Methodology

This assessment was conducted using the Meridant AI Threat Readiness Assessment, a packaged SKU built on a NIST Cybersecurity Framework v2.0 backbone with 24 hand-authored capabilities. Each capability carries five behavioural anchors (L1 Not Defined → L5 Optimised) and a default target level (L3 or L4 depending on whether the capability represents foundational hygiene or AI-specific extension). Respondents from the Albatross security, IT, risk, and AI engineering teams completed the survey between 15 April and 1 May 2026; 17 respondents in total, with role-based question routing. Scoring used the standard Meridant aggregation: capability scores are the arithmetic mean of scored responses, N/A responses are excluded from the count, function scores are the unweighted mean of constituent capability scores, and the overall AI Threat Readiness score is the weighted mean across all capabilities (weight = 1.0 for all v1.0 capabilities). Cross-walk references to NIST AI RMF, NIST SP 800-53, NIST SP 800-30, NIST SP 800-218 SSDF, CSA MAESTRO, OWASP LLM Top 10, and MITRE ATLAS are sourced from the v1.0 SKU content library.

Function Maturity Summary

Capability	Score	Target	Achieved
GOVERN	**3.25**	**3.50**	No
Board AI threat awareness	3.00	4.00	No
Risk appetite recalibration for AI-era threats	2.00	3.00	No
Third-party AI risk policy & contracting	4.00	4.00	Yes
Security budget posture vs threat trajectory	4.00	3.00	Yes
IDENTIFY	**3.25**	**3.75**	No
Asset inventory including AI assets & data flows	3.00	4.00	No
Vulnerability management coverage & SLA	4.00	4.00	Yes
External attack surface management	4.00	4.00	Yes
SBOM & software supply-chain transparency	2.00	3.00	No
PROTECT	**3.60**	**3.80**	No
Patch & vulnerability remediation cadence	4.00	4.00	Yes
Configuration hardening automation	4.00	4.00	Yes
Phishing-resistant MFA coverage	4.00	4.00	Yes
Privileged access containment & segmentation	4.00	4.00	Yes
Secure-by-design SDLC, including AI-using systems	2.00	3.00	No
DETECT	**3.00**	**3.25**	No
Detection telemetry coverage	4.00	4.00	Yes
Behavioural & anomaly detection	3.00	3.00	Yes
AI-augmented triage & threat hunting	2.00	3.00	No
Detection time discipline & measurement	3.00	3.00	Yes
RESPOND	**3.50**	**3.50**	Yes
SOAR & response automation maturity	3.00	3.00	Yes
Decision rights & escalation paths	4.00	4.00	Yes
Stakeholder & regulator communication playbooks	3.00	3.00	Yes
Tabletop exercise frequency & realism	4.00	4.00	Yes
RECOVER	**3.00**	**3.33**	No
Immutable backup posture	4.00	4.00	Yes
Critical-service RTO realism	3.00	3.00	Yes
Business continuity tested vs AI-attack scenarios	2.00	3.00	No

Overall AI Threat Readiness: 3.27 / 5.00

Findings

Govern

The board acknowledges AI as a strategic risk and has discussed it twice in the past twelve months, but the cadence is event-driven rather than scheduled — there is no standing AI threat item on the risk committee agenda and no documented threshold-based reporting back to directors. Risk appetite has not yet been formally rewritten for the AI era; a working group convened in late 2025 produced draft thresholds but they remain unapproved. By contrast, third-party AI risk policy is a recent strength: a 2025 procurement-led initiative produced tiered diligence, AI-specific contract clauses, and continuous monitoring for the four highest-tier model and SaaS providers. Budget posture is mature, with surge mechanisms used in response to a Q1 2026 industry threat-tier elevation.

- **Board AI threat awareness (L3 vs L4):** Standing AI item on agenda since Nov 2025 but only semi-annual cadence; decisions captured in minutes are not yet tracked to closure.
- **Risk appetite recalibration (L2 vs L3):** Draft AI-era thresholds exist but are not board-approved. This is the highest-leverage gap in the assessment.
- **Third-party AI risk policy (L4):** Meets target. Continuous monitoring of high-tier vendors in place. Renewal-time tier review embedded in procurement workflow.
- **Budget posture (L4):** Exceeds target. Surge mechanism tested and used in March 2026 for accelerated detection investment.

Identify

Foundational asset management and vulnerability programmes are mature and broadly aligned to the credible 2026 bar. The gap is concentrated in two AI-specific extensions. Asset inventory is rated L3 because while infrastructure and applications are tracked authoritatively, AI assets (model endpoints, training and inference data flows, vector stores, prompt logging) are captured ad hoc rather than continuously discovered. SBOM maturity is L2: SBOMs are requested from critical vendors but not consistently produced for in-house software, and AI model cards are rarely captured for the third-party model APIs in production use. External attack surface management is strong, with continuous EASM coverage explicitly including the customer-facing model gateway.

- **Asset inventory inc. AI assets (L3 vs L4):** AI-asset register exists but lags reality; shadow AI surveyed annually rather than continuously.
- **Vulnerability management coverage (L4):** Meets target. AI-supporting infrastructure included in scope with same SLAs.
- **External attack surface management (L4):** Meets target. Model-gateway exposure caught and remediated in Feb 2026.
- **SBOM maturity (L2 vs L3):** Inconsistent in-house SBOM generation; model cards captured for two of seven third-party model endpoints.

Protect

The protective control surface is the most mature function in the assessment in absolute terms. Patch cadence, configuration hardening, phishing-resistant MFA, and privileged access containment all meet their L4 targets, reflecting sustained investment over the past three years. The single gap is **Secure-by-design SDLC for AI-using systems** — internal-use Copilot rollout outpaced the security-review pipeline, the AI-assisted credit decisioning pilot was scoped before secure-by-design checkpoints were defined for AI components, and prompt-injection /

RAG-poisoning threat modelling is not yet routine. This is a known issue inside the AI engineering team and a remediation programme is in flight.

- **Patch remediation cadence (L4):** Meets target. ≥95% adherence on critical and high SLAs; surge process exercised twice in past 12 months.
- **Configuration hardening automation (L4):** Meets target. Drift remediation mean-time-to-correct measured in hours.
- **Phishing-resistant MFA (L4):** Meets target. FIDO2 universal for privileged access; ~82% of standard users on phishing-resistant factors.
- **Privileged access segmentation (L4):** Meets target. JIT in place for all admin tiers; AI model-deployment pipelines treated as privileged scope.
- **Secure-by-design SDLC inc. AI (L2 vs L3):** AI threat modelling not yet routine; remediation programme in flight, target completion Q3 2026.

Detect

Detection telemetry coverage is strong on the classical estate (endpoint, network, identity, cloud control-plane). Behavioural and anomaly detection is L3 — defined and operating — but with limited coverage of AI-system telemetry: model inference logs, prompt streams, embedding-store access patterns, and agent-toolchain activity are not yet fused with the broader behavioural baseline. Detection time discipline is credible at L3, with MTTD measured and reported, though without active reduction targets. The standout gap is AI-augmented triage and threat hunting, rated L2: the SOC has trialled commercial AI-assist tooling but it is not yet embedded in the triage workflow, and hunting hypotheses do not consistently incorporate AI-attack patterns (e.g. ATLAS techniques).

- **Detection telemetry coverage (L4):** Meets target. SIEM coverage across identity, endpoint, network, cloud control-plane; model-gateway telemetry partial.
- **Behavioural anomaly detection (L3):** Meets target. UEBA in production for identity and endpoint; AI-system behavioural signals not yet integrated.
- **AI-augmented triage (L2 vs L3):** Tool trial under way; not yet embedded in SOC workflow.
- **Detection time discipline (L3):** Meets target. MTTD reported monthly; no active reduction programme yet.

Respond

This is the only function meeting its threshold. The Albatross response capability benefited from a 2025 regulatory-driven uplift: decision rights and escalation paths are unusually clear for an organisation of this size, with named decision-makers, time-bound escalation triggers, and a board-level severity ladder. Tabletop exercises run quarterly with realistic scenarios; SOAR automation handles ~40% of high-volume detection types end-to-end. The thinnest part of the function is communication playbooks (L3) — regulator notification timelines are codified but customer and counterparty messaging has not been pre-staged for an AI-incident scenario (e.g. customer deepfake-driven fraud at scale).

- **SOAR maturity (L3):** Meets target. ~40% high-volume detection automation; orchestration extended to identity and cloud responses.
- **Decision rights & escalation (L4):** Meets target. Decision-rights matrix reviewed annually; severity ladder formally board-approved.
- **Communication playbooks (L3):** Meets target. Regulator notification timelines codified; AI-incident customer comms not pre-staged.
- **Tabletop frequency & realism (L4):** Meets target. Quarterly cadence; last AI-augmented social-engineering scenario run Q4 2025.

Recover

Recovery posture is strong on the technical fundamentals — immutable backup posture is L4 with air-gapped restore tested quarterly. Critical-service RTO realism is L3: documented RTOs exist for tier-1 services and

quarterly recovery drills validate the technical path, though restoration of dependent AI services (model endpoints, vector stores, embeddings) is not yet integrated into the drill scope. The most significant gap is **Business continuity tested vs AI-attack scenarios** at L2: BC plans exist and are exercised, but none of the past four exercises specifically simulated an AI-augmented attack (e.g. concurrent deepfake-driven authorisation fraud during a primary-site failover).

- **Immutable backup posture (L4):** Meets target. Air-gapped backup with quarterly restore tests.
- **Critical-service RTO realism (L3):** Meets target. Tier-1 RTOs validated quarterly; AI-service restoration not yet in drill scope.
- **BC tested vs AI-attack scenarios (L2 vs L3):** BC plans exist; AI-augmented scenarios not yet exercised. Recommend incorporating in next quarterly drill.

Risk Summary

The findings above translate into the following enterprise risk posture, prioritised by combined likelihood and impact.

#	Risk	Function	Drivers	Likelihood	Impact	Priority
1	Investment misallocation due to absent AI-era risk appetite	Govern	Risk appetite recalibration L2; no quantitative thresholds	High	High	P1
2	Undetected adversary activity against AI surface	Detect	AI-augmented triage L2; AI-system telemetry not in behavioural baseline	High	High	P1
3	Failed recovery from AI-augmented attack scenario	Recover	BC tested vs AI-attack scenarios L2	Medium	High	P1
4	AI-system vulnerability introduced via insecure development	Protect	Secure-by-design SDLC inc. AI L2	Medium	High	P2
5	Supply-chain compromise via opaque AI component lineage	Identify	SBOM maturity L2; AI model cards partial	Medium	High	P2
6	Slow board response to threat-tier change	Govern	Board AI threat awareness L3 (semi-annual cadence)	Medium	Medium	P2
7	Customer / regulator messaging lag in AI incident	Respond	Communication playbooks L3, AI-specific not pre-staged	Low	High	P2
8	Shadow AI introducing unmanaged data exposure	Identify	Asset inventory inc. AI assets L3 (annual shadow AI survey)	Medium	Medium	P3

Detailed Roadmap

Quick Wins (0 – 30 days)

- **Schedule the board risk-appetite recalibration session.** Calendar block on the June risk committee agenda. Inputs already drafted by the working group; this is a scheduling decision, not a content decision.
- **Add AI-system telemetry sources to the SIEM ingestion backlog.** Model-gateway logs, prompt-stream metadata, vector-store access logs. Engineering work is straightforward; the bottleneck is scope agreement.
- **Document the AI-incident customer communication playbook stub.** Even an unpolished first draft beats the current absence. Use the Q4 2025 tabletop output as the seed.
- **Capture model cards for the remaining five third-party model endpoints.** Procurement already has the template from the four completed.

Phase 1 — Foundation (1 – 3 months)

Initiative	Owner	Function	Expected outcome	Effort
Board-approved AI-era risk appetite statement with quantitative thresholds	CRO	Govern	Risk appetite recalibration L2 → L3	M
Establish quarterly board AI threat posture report with decision-tracking	CISO	Govern	Board AI threat awareness L3 → L4	M
Extend behavioural detection baseline to AI-system telemetry	SOC Lead	Detect	Behavioural anomaly detection L3 → L4 (partial)	L
Run a tabletop exercise against an AI-augmented attack scenario, then refresh BC plan inputs from findings	IR Lead	Respond / Recover	Improves Recover; surfaces concrete BC gaps	M
Activate the AI-incident customer / regulator communication playbook	Comms Lead + Legal	Respond	Communication playbooks L3 → L4	S

Phase 2 — Build (3 – 9 months)

Initiative	Owner	Function	Expected outcome	Effort
Embed AI-augmented triage tooling into the SOC workflow with measured triage-time reduction	SOC Lead	Detect	AI-augmented triage L2 → L3	L
Routine prompt-injection, RAG-poisoning, and model-supply-chain threat modelling in the SDLC for all AI-using systems	AppSec Lead	Protect	Secure-by-design SDLC inc. AI L2 → L3	L
In-house SBOM generation at build time for all production services; AI model card capture mandatory at procurement	Engineering + Procurement	Identify	SBOM maturity L2 → L3	L
Add AI-service restoration to quarterly recovery drill scope (model endpoints, vector stores, embeddings)	DR Lead	Recover	Critical-service RTO realism partial uplift	M
Continuous AI-asset discovery via egress monitoring and SaaS-application telemetry; quarterly shadow AI report	IT Asset Mgr	Identify	Asset inventory inc. AI assets L3 → L4	M

Phase 3 — Optimise (9 – 18 months)

Initiative	Owner	Function	Expected outcome	Effort
Quarterly BC exercise specifically targeting AI-augmented attack scenarios; findings feed back into BC plan and protect-control investments	DR Lead	Recover	BC tested vs AI-attack scenarios L2 → L4	M
Continuous attestation of authentication strength per session with risk-based step-up	IAM Lead	Protect	Phishing-resistant MFA L4 → L5	L
Real-time AI asset and data-flow graph with automated tagging propagating into protect / detect controls	Platform Eng	Identify	Asset inventory inc. AI assets L4 → L5	XL
Vendor AI risk dashboard with real-time signal ingestion	TPRM Lead	Govern	Third-party AI risk policy L4 → L5	L
Director-level AI threat education programme tied to board-pack annual refresh	CISO + CRO	Govern	Board AI threat awareness L4 → L5	M

Effort key: S = ≤ 1 person-month; M = 1–3 person-months; L = 3–6 person-months; XL = ≥ 6 person-months.

Appendix — Framework Reference Crosswalk

Capability ↔ external reference codes from the AI Threat Readiness v1.0 content library. Frameworks cited in canonical precedence order: NIST CSF 2.0, NIST AI RMF, NIST SP 800-53, NIST SP 800-30, NIST SP 800-218 SSDF, CSA MAESTRO, OWASP LLM Top 10, MITRE ATLAS.

Govern

Board AI threat awareness

The board has structured, recurring visibility into the AI-augmented threat landscape relevant to the organisation.
Score: 3.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	GV.OC-01	Organizational mission and stakeholders communicated and used in risk management decisions.
NIST CSF 2.0	GV.RR-01	Organizational leadership establishes and communicates roles and responsibilities for cybersecurity risk.
NIST CSF 2.0	GV.OV-01	Cybersecurity risk management strategy outcomes are reviewed and adjusted.
NIST AI RMF	GOVERN 1.1	Legal and regulatory requirements involving AI are understood and documented.
NIST AI RMF	GOVERN 2.1	Roles and responsibilities related to AI risk are documented and communicated.
NIST SP 800-53	PM-1	Information security program plan with senior leadership sign-off.
NIST SP 800-53	PM-9	Risk management strategy approved by senior leadership.
CSA MAESTRO	L6 Security & Compliance	Governance, auditability, and regulatory alignment across the AI stack.

Risk appetite recalibration for AI-era threats

Enterprise risk appetite explicitly revisited in light of AI-augmented adversary capabilities and AI-use exposure.
Score: 2.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	GV.RM-01	Risk management objectives established and agreed to by stakeholders.
NIST CSF 2.0	GV.RM-02	Risk appetite and risk tolerance statements established, communicated, and maintained.
NIST CSF 2.0	GV.OV-03	Cybersecurity risk management performance is evaluated.
NIST AI RMF	GOVERN 1.3	Processes to determine risk tolerance.
NIST AI RMF	MAP 1.5	Organizational risk tolerances inform AI system risk management.
NIST SP 800-53	PM-9	Risk management strategy with risk tolerance defined.
NIST SP 800-30	§3.1	Risk framing including risk tolerance and risk appetite.
CSA MAESTRO	L6 Security & Compliance	Risk-based prioritisation in the agent context.

Third-party AI risk policy & contracting

Third parties governed by policy and contracts that address AI-specific risk. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	GV.SC-01	Cybersecurity supply chain risk management strategy established.
NIST CSF 2.0	GV.SC-05	Supply chain requirements integrated into contracts.
NIST CSF 2.0	ID.SC-02	Suppliers prioritized and assessed.
NIST AI RMF	GOVERN 6.1	Policies address AI risks associated with third-party entities.
NIST SP 800-53	SR-2	Supply chain risk management plan.
NIST SP 800-53	SR-3	Supply chain controls.
NIST SP 800-53	SA-9	External system services.
OWASP LLM Top 10	LLM05	Supply chain vulnerabilities — third-party model and dataset risk.
CSA MAESTRO	L1 / L3	Compromised pre-trained models + vulnerabilities in ML libraries.

Security budget posture vs threat trajectory

Security investment levels and allocation explicitly informed by the AI-era threat trajectory. **Score:** 4.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	GV.RM-04	Strategic direction describing appropriate risk response established.
NIST CSF 2.0	GV.OV-02	Risk management strategy reviewed and adjusted.
NIST AI RMF	GOVERN 1.6	Inventory of AI systems and resources allocated to AI risk management.
NIST SP 800-53	PM-3	Information security and privacy resources / budget.
NIST SP 800-53	PM-11	Mission and business process definition tying security to mission.

Identify

Asset inventory including AI assets & data flows

Authoritative inventory including AI assets — models, training/inference data flows, prompt stores, embeddings, vector stores. **Score:** 3.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	ID.AM-01	Inventories of hardware managed by the organization.
NIST CSF 2.0	ID.AM-02	Inventories of software, services, and systems.
NIST CSF 2.0	ID.AM-03	Network communication and data flow representations.
NIST CSF 2.0	ID.AM-04	Inventories of services provided by suppliers.
NIST AI RMF	MAP 1.1	Intended purposes and contexts of AI system deployment documented.
NIST AI RMF	MAP 3.1	Potential benefits of AI system functionality documented.
NIST SP 800-53	CM-8	System component inventory.
NIST SP 800-53	PM-5	System inventory.

Vulnerability management coverage & SLA

Vulnerability programme covers the full estate including AI-supporting infrastructure with SLA-bound remediation. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	ID.RA-01	Vulnerabilities in assets are identified, validated, and recorded.
NIST CSF 2.0	ID.RA-08	Processes for vulnerability disclosures are established.
NIST CSF 2.0	DE.CM-09	Computing hardware and software monitored to find potentially adverse events.
NIST SP 800-53	RA-5	Vulnerability monitoring and scanning.
NIST SP 800-53	RA-7	Risk response.
NIST SP 800-53	SI-2	Flaw remediation.
NIST SP 800-30	§3.2	Risk assessment with vulnerability identification step.

External attack surface management

Continuously refreshed view of external attack surface including AI assets and inference endpoints. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	ID.AM-05	Assets prioritized based on classification, criticality, and impact.
NIST CSF 2.0	ID.RA-01	Vulnerabilities in assets are identified.
NIST CSF 2.0	DE.CM-02	Physical and network boundary environment is monitored.
NIST SP 800-53	CA-2	Control assessments.
NIST SP 800-53	CA-8	Penetration testing.
NIST SP 800-53	RA-3	Risk assessment with threat-context integration.
MITRE ATLAS	AML.TA0002	Reconnaissance — adversary discovery of ML assets and victim surface.
OWASP LLM Top 10	LLM10	Model theft via exposed endpoints.

SBOM & software supply-chain transparency

Generate, consume, and act on SBOMs including AI/ML-specific artefacts — model cards, data sheets, training-data provenance. **Score:** 2.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	ID.AM-08	Lifecycle management of systems, hardware, software, services, data.
NIST CSF 2.0	ID.RA-09	Authenticity and integrity of hardware and software assessed pre-acquisition.
NIST CSF 2.0	GV.SC-08	Suppliers included in incident planning, response, and recovery.
NIST AI RMF	MAP 4.1	Approaches for mapping AI tech / legal risks of third-party components.
NIST AI RMF	MEASURE 2.10	Privacy risk including training-data provenance examined.
NIST SP 800-53	SR-4	Provenance.
NIST SP 800-53	SR-11	Component authenticity.
NIST SP 800-53	SA-8(9)	Trusted components via provenance and attestation.
OWASP LLM Top 10	LLM03	Training data poisoning — provenance is the defensive counterpart.
OWASP LLM Top 10	LLM05	Supply chain vulnerabilities.
MITRE ATLAS	AML.T0010	ML supply chain compromise.
CSA MAESTRO	L1 / L2	Foundation Models + Data Operations provenance gap.

Protect

Patch & vulnerability remediation cadence

Vulnerabilities remediated on cadence matched to risk; surge process tested. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	ID.IM-02	Improvements identified from security tests and exercises.
NIST CSF 2.0	PR.PS-02	Software maintained, replaced, removed commensurate with risk.
NIST CSF 2.0	PR.PS-03	Hardware maintained, replaced, removed commensurate with risk.
NIST CSF 2.0	DE.CM-09	Computing hardware and software monitored.
NIST SP 800-53	SI-2	Flaw remediation.
NIST SP 800-53	SI-2(2)	Automated remediation status.
NIST SP 800-53	CM-3	Configuration change control.
NIST SP 800-30	§3.4	Risk monitoring — continuous detection of vulnerabilities.

Configuration hardening automation

Hardened-by-default deployment; continuous drift detection across the estate including AI infrastructure. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	PR.PS-01	Configuration management practices established and applied.
NIST CSF 2.0	PR.PS-04	Log records generated for continuous monitoring.
NIST CSF 2.0	PR.IR-01	Networks and environments protected from unauthorized access.
NIST SP 800-53	CM-2	Baseline configuration.
NIST SP 800-53	CM-6	Configuration settings.
NIST SP 800-53	CM-7	Least functionality.
NIST SP 800-53	SI-7	Software, firmware, and information integrity.

Phishing-resistant MFA coverage

FIDO2 / WebAuthn / hardware-token / cert-based authentication; legacy factors eliminated for sensitive use. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	PR.AA-01	Identities and credentials managed.
NIST CSF 2.0	PR.AA-02	Identities proofed and bound to credentials.
NIST CSF 2.0	PR.AA-03	Users, services, hardware authenticated.
NIST CSF 2.0	PR.AA-04	Identity assertions protected, conveyed, verified.
NIST SP 800-53	IA-2	Identification and authentication.
NIST SP 800-53	IA-2(1)	MFA to privileged accounts.
NIST SP 800-53	IA-2(2)	MFA to non-privileged accounts.
NIST SP 800-53	IA-2(8)	Replay-resistant authentication (phishing-resistant factor mapping).

Privileged access containment & segmentation

Privileged access minimised, vaulted, time-bound, isolated; AI admin interfaces in privileged scope. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	PR.AA-05	Access permissions, entitlements, authorizations managed.
NIST CSF 2.0	PR.AA-06	Physical access managed.
NIST CSF 2.0	PR.IR-01	Network protection from unauthorized access.
NIST SP 800-53	AC-2	Account management.
NIST SP 800-53	AC-5	Separation of duties.
NIST SP 800-53	AC-6	Least privilege.
NIST SP 800-53	AC-6(1)	Authorize access to security functions.
MITRE ATLAS	AML.T0012	Valid accounts — lateral movement via compromised credentials.

Secure-by-design SDLC, including AI-using systems

Threat modelling, secure coding, security testing across the SDLC including AI components (prompt injection, RAG poisoning, model supply chain). **Score:** 2.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	PR.PS-06	Secure software development practices integrated and performance monitored.
NIST CSF 2.0	ID.RA-09	Authenticity and integrity assessed prior to acquisition.
NIST AI RMF	MAP 2.3	Scientific integrity and TEVV considerations across the AI lifecycle.
NIST AI RMF	MEASURE 2.7	AI system security and resilience evaluated and documented.
NIST SP 800-53	SA-3	System development life cycle.
NIST SP 800-53	SA-8	Security and privacy engineering principles.
NIST SP 800-53	SA-11	Developer testing and evaluation.
NIST SP 800-218 SSDF	PW.7	Review and analyze human-readable code for security issues.
NIST SP 800-218 SSDF	PW.8	Test executable code to identify vulnerabilities.
OWASP LLM Top 10	LLM01	Prompt injection.
OWASP LLM Top 10	LLM02	Insecure output handling.
OWASP LLM Top 10	LLM06	Sensitive information disclosure.
MITRE ATLAS	AML.T0051	Prompt injection.
CSA MAESTRO	L1–L3	Foundation Models, Data Operations, Agent Frameworks — design-time integrity.

Detect

Detection telemetry coverage

Telemetry across the AI-affected attack surface — identity, endpoint, network, cloud, model gateway, prompt stream. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	DE.CM-01	Networks and network services are monitored.
NIST CSF 2.0	DE.CM-03	Personnel activity is monitored to find adverse events.
NIST CSF 2.0	DE.CM-06	External service provider activity monitored.
NIST CSF 2.0	DE.CM-09	Computing hardware and software monitored.
NIST SP 800-53	AU-2	Event logging.
NIST SP 800-53	AU-6	Audit record review, analysis, reporting.
NIST SP 800-53	SI-4	System monitoring.
CSA MAESTRO	L4 Deployment & Infrastructure	Observability across the deployed AI surface.

Behavioural & anomaly detection

Behavioural baselines and anomaly detection over identity, endpoint, network, and AI-system telemetry. **Score:** 3.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	DE.AE-02	Potentially adverse events analyzed to better understand activities.
NIST CSF 2.0	DE.AE-03	Information is correlated from multiple sources.
NIST CSF 2.0	DE.CM-03	Personnel activity monitored.
NIST SP 800-53	SI-4	System monitoring.
NIST SP 800-53	SI-4(2)	System monitoring — automated tools for real-time analysis.
NIST AI RMF	MEASURE 2.6	AI system security risks identified and documented.
OWASP LLM Top 10	LLM08	Excessive agency — detecting actions outside policy.
MITRE ATLAS	AML.TA0007	ML attack staging — anomalous AI-system activity.

AI-augmented triage & threat hunting

AI tooling embedded in triage workflow; hunting hypotheses incorporate AI-attack patterns. **Score:** 2.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	DE.AE-04	Estimate impact and scope of adverse events.
NIST CSF 2.0	DE.AE-08	Incidents declared when adverse events meet defined criteria.
NIST CSF 2.0	DE.CM-09	Hardware and software monitored.
NIST SP 800-53	IR-4	Incident handling.
NIST SP 800-53	IR-5	Incident monitoring.
MITRE ATLAS	(full framework)	Adversary techniques against ML systems — hunting hypothesis library.
CSA MAESTRO	L7 Agent Ecosystem	Threat hunting across multi-agent context.

Detection time discipline & measurement

MTTD measured, reported, and trending toward documented reduction targets. **Score:** 3.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	DE.AE-06	Information on adverse events is provided to authorized staff and tools.
NIST CSF 2.0	ID.IM-01	Improvements identified from evaluations.
NIST SP 800-53	IR-5	Incident monitoring.
NIST SP 800-53	IR-6	Incident reporting.
NIST SP 800-53	CA-7	Continuous monitoring.
NIST SP 800-30	§3.4	Risk monitoring — continuous detection.

Respond

SOAR & response automation maturity

Orchestration and automation handle high-volume detection types end-to-end with measurable analyst-time saving. **Score:** 3.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	RS.MA-01	Incident response plan is executed in coordination with relevant third parties.
NIST CSF 2.0	RS.MA-02	Incident reports triaged and validated.
NIST CSF 2.0	RS.AN-03	Analyses are performed to establish what has taken place.
NIST CSF 2.0	RS.AN-06	Incident reports recorded and preserved.
NIST SP 800-53	IR-4	Incident handling.
NIST SP 800-53	IR-4(1)	Automated incident handling processes.
NIST SP 800-53	IR-7	Incident response assistance.

Decision rights & escalation paths

Named decision-makers, time-bound escalation triggers, board-level severity ladder. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	GV.RR-02	Roles, responsibilities, authorities for managing cybersecurity risk are established.
NIST CSF 2.0	RS.MA-01	Incident response plan executed in coordination.
NIST CSF 2.0	RS.CO-02	Internal and external stakeholders informed.
NIST SP 800-53	IR-2	Incident response training.
NIST SP 800-53	IR-8	Incident response plan.
NIST AI RMF	GOVERN 2.1	Roles and responsibilities related to AI risk documented.

Stakeholder & regulator communication playbooks

Pre-staged messaging for regulators, customers, counterparties — including AI-incident scenarios. **Score:** 3.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	RS.CO-02	Stakeholders informed per response plan.
NIST CSF 2.0	RS.CO-03	Information shared with designated stakeholders.
NIST CSF 2.0	GV.RR-04	Cybersecurity included in human resources practices including comms training.
NIST SP 800-53	IR-6	Incident reporting.
NIST SP 800-53	IR-8	Incident response plan including communications.
NIST AI RMF	GOVERN 4.3	Practices and processes that operationalize AI policies prioritise stakeholder feedback.

Tabletop exercise frequency & realism

Regular tabletop cadence with realistic scenarios including AI-augmented adversary techniques. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	ID.IM-02	Improvements identified from security tests and exercises.
NIST CSF 2.0	RS.MA-04	Incident reporting linked to lessons-learned activities.
NIST CSF 2.0	PR.AT-01	Personnel provided with awareness and training.
NIST SP 800-53	IR-2	Incident response training.
NIST SP 800-53	IR-3	Incident response testing.
NIST SP 800-53	CP-4	Contingency plan testing.
MITRE ATLAS	(full framework)	Scenario library for AI-augmented adversary techniques.

Recover

Immutable backup posture

Air-gapped or immutable backup tested under realistic restoration conditions. **Score:** 4.00 / 5.00 **Target:** 4.00

Framework	Code	Description
NIST CSF 2.0	PR.DS-11	Backups of data are created, protected, maintained, and tested.
NIST CSF 2.0	RC.RP-03	Integrity of backups verified before being used.
NIST CSF 2.0	RC.RP-04	Critical mission functions and risk are considered in restoration.
NIST SP 800-53	CP-9	System backup.
NIST SP 800-53	CP-9(1)	Test backup information.
NIST SP 800-53	CP-10	System recovery and reconstitution.
NIST SP 800-53	SI-7	Software, firmware, information integrity (immutability assurance).

Critical-service RTO realism

Documented RTOs validated under drill; AI services restored in scope. **Score:** 3.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	RC.RP-01	Recovery plan executed during or after an incident.
NIST CSF 2.0	RC.RP-02	Recovery actions selected, scoped, prioritized, performed.
NIST CSF 2.0	RC.CO-04	Public updates use approved methods and messaging.
NIST SP 800-53	CP-2	Contingency plan.
NIST SP 800-53	CP-4	Contingency plan testing.
NIST SP 800-53	CP-7	Alternate processing site.

Business continuity tested vs AI-attack scenarios

BC exercises specifically simulate AI-augmented attack scenarios; findings feed back into plan and controls. **Score:** 2.00 / 5.00 **Target:** 3.00

Framework	Code	Description
NIST CSF 2.0	RC.RP-01	Recovery plan executed.
NIST CSF 2.0	RC.RP-05	Integrity of restored assets is verified.
NIST CSF 2.0	ID.IM-02	Improvements identified from exercises.
NIST SP 800-53	CP-2	Contingency plan.
NIST SP 800-53	CP-4	Contingency plan testing.
NIST SP 800-53	IR-3	Incident response testing.
MITRE ATLAS	(full framework)	Scenario library for BC test exercises.
CSA MAESTRO	L4 / L7	Deployment-infrastructure resilience + agent-ecosystem failure modes.

End of sample report. Confidential — sample artifact for design review only.