AI Incident Response Plan

Vesta Mutual Insurance AS · Plan version 1.0 · 29 June 2026 · Owner: AI Governance Lead · Exercise: annual tabletop · Review: annual and after every SEV-1/2

Worked example for portfolio and training purposes. Vesta Mutual Insurance AS is a fictional company; all data, metrics and names are invented. Prepared by Erik Bernath, Furioso AI Consulting OÜ (furiosoaiconsulting.eu), June 2026. Licensed CC BY 4.0: reuse freely with attribution. This document is informational and is not legal advice.

1. Purpose and what counts as an AI incident

This plan defines how Vesta detects, contains, reports and learns from AI incidents. An AI incident is any event where an AI system in the inventory produces harmful, unlawful or materially wrong outcomes, behaves outside its documented bounds, or is used outside policy. Five categories: harmful or discriminatory output; malfunction or drift; misuse (including shadow AI with sensitive data); security events specific to AI (prompt injection, data leakage through a model); and vendor-side incidents affecting a system Vesta deploys. AI incidents overlap with, but are not identical to, security incidents and personal-data breaches; section 5 handles the interplay.

2. Severity ladder

Level	Definition	Vesta examples	Response clock
SEV-1	Serious harm occurred or imminent: unlawful discrimination in force, wrongful denials at scale, or an AI Act serious incident (Art. 3(49))	PriceWise systematically over-pricing a protected group in production; chatbot giving claims-waiver promises at scale; TalentScreen filtering on a prohibited basis	Contain within 2 hours; incident team convenes same day
SEV-2	Material malfunction or compliance breach, harm contained or probable rather than occurred	Red disparate-impact test on TalentScreen; PriceWise drift past red threshold; unsanctioned tool found processing client personal data	Contain within 1 business day
SEV-3	Degradation, near-miss or policy breach without material harm	Aida hallucinating policy terms to 1 customer (corrected); staff pasting internal data into a public tool once; parsing failures above tolerance	Owner resolves within 5 business days; logged and trended

3. Roles

Incident Lead: the AI Governance Lead (deputy: Head of IT) classifies severity, runs the response, owns the record. System owner: executes containment for their system. DPO: rules on personal-data dimensions and GDPR notification within the first assessment call. CISO function (IT Security): handles security-category incidents jointly. Communications and Legal: engaged for every SEV-1 and any incident with customer contact. The COO is informed of every SEV-1 immediately and decides external communications. Any employee can report a suspected incident directly, anonymously if preferred; reporting in good faith is explicitly protected.

4. Response phases

Detect. Channels: monitoring dashboards (drift, fairness, override rates), staff reports, customer complaints flagged by Customer Ops, vendor notices, and the quarterly shadow-AI sweep. Every channel routes to the AI Governance Lead inbox with a 1-business-day triage promise.

Triage and classify. Severity per section 2, recorded in the incident register with system, category, affected persons, and clock start. When in doubt between two levels, take the higher.

Contain. Every high-risk system has a pre-approved fallback: TalentScreen falls back to manual screening of the full applicant pool; PriceWise falls back to the GLM baseline with widened manual review; Aida falls back to human-only chat with adjusted staffing. Containment authority is explicit: the system owner or the Incident Lead can suspend any system without further approval, and nobody is criticized for a defensible suspension.

Assess. Root cause including the model dimension (data, drift, version change, configuration, misuse), scope of affected persons and decisions, and a remediation list for past affected cases, not only future ones. For decisions that may have been wrong (pricing, shortlisting), the affected-case review is mandatory and its scope is approved by the COO.

Notify. Internal: per severity ladder. Vendor systems: Vesta informs the provider without undue delay when it identifies a serious incident or a risk it cannot resolve (Art. 26(5) duty), and suspends use where instructions cannot be followed. Where Vesta is the provider (PriceWise): serious incidents are reported to the market surveillance authority immediately and within the Art. 73 deadlines (15 days general; shorter for the gravest categories), with the report owned by Legal. Personal-data breach: GDPR 72-hour clock runs in parallel, owned by the DPO. Insurance supervisor notification is assessed by Legal for SEV-1 events touching pricing or claims.

Recover. A contained system returns to service only through a re-approval gate: validated fix, re-run of the relevant tests (fairness, calibration, or behavioural), and sign-off by the system owner and AI Governance Lead; SEV-1 returns also need COO sign-off.

Learn. Post-incident review within 10 business days for SEV-1/2: timeline, root cause, control gaps, actions with owners and dates. Findings feed the risk assessments, the policy, and the training curriculum, and the incident register is a standing input to the quarterly AI Governance Committee.

5. Interplay with existing plans

If an AI incident is also a security incident, the security incident process leads and this plan supplies the AI-specific assessment. If personal data is breached, the DPO's GDPR process runs in parallel with its own clock. This plan never extends another plan's deadline; where clocks differ, the shortest governs.

6. Templates and register

The incident register records: ID, date, system, category, severity, summary, affected persons and scale, containment time, notifications made with timestamps, root cause, actions, closure date. A holding-statement template for customer-facing incidents is maintained by Communications (acknowledge, no speculation on cause, human contact path, follow-up commitment). Both live with the AI Governance Lead; the register is evidence, so entries are never edited after closure, only annotated.

7. Exercises

One tabletop exercise per year, alternating scenarios across the high-risk systems (2026: PriceWise miscalibration discovered by an external complaint; 2027: TalentScreen disparate-impact finding mid-campaign). The exercise tests the fallbacks, the notification decision tree, and the affected-case review scoping, and its findings are treated as SEV-3 entries in the register. The first exercise is scheduled before the 2 August 2026 enforcement date.

8. Framework mapping

EU AI Act: Art. 26(5) deployer duties, Art. 73 provider serious-incident reporting, Art. 3(49) definitions. GDPR: Arts. 33-34 in parallel. ISO/IEC 42001: nonconformity and corrective action (cl. 10), incident-related Annex A controls. NIST AI RMF: MANAGE 4 (incident response, recovery, communication). DORA note: Vesta is an insurer, so ICT-incident reporting under DORA may also attach when the AI incident is ICT-related; Legal holds the mapping table.