



AI SECURITY INCIDENT & RISK ANALYSIS REPORT

Q1 2026 · Technical & Compliance Edition

Real-World Breaches · Control Plane Failures · Adversary Techniques · Regulatory Readiness

6 Incidents

Fully analysed with forensic detail

6 Controls

Auditable SHALL/SHOULD statements

6 Gaps

In current standards & frameworks

For: CISOs · Security Architects · AI Governance Leads · Compliance Officers · Standards Body Participants
Companion document: AI Security Executive Brief (Leadership Edition)

— S1 —

METHODOLOGY

Evidence framework & financial estimation model

Evidence Classification Tiers

Every claim is assigned one of five evidence tiers. Financial figures not directly disclosed use an explicit estimation formula with conservative (25th-percentile) assumptions.

Tier	Definition	Label
Primary Verified	Confirmed by affected organization, regulatory filing, or primary vendor disclosure	[PRIMARY VERIFIED]

Secondary Verified	Corroborated by 2+ established threat intelligence firms with named attribution	[SECONDARY VERIFIED]
Reported	Credible single-source disclosure lacking full independent corroboration	[REPORTED]
Estimate	Derived from industry benchmarks, partial disclosures, or explicit modelling	[ESTIMATE]
Illustrative	Research prototypes, lab demonstrations, or forward-looking threat trajectories	[ILLUSTRATIVE]

Financial Estimation Formula

$(N_records \times Cost_per_Record_benchmark) + Incident_Response + Operational_Downtime + (Regulatory_Probability \times Penalty_Range) \leftarrow$ capped at 25th percentile of historical analogues

— S2 —

THREAT LANDSCAPE

Q1 2026 macro signals & actor taxonomy

Documented Q1 2026 Macro Signals

Signal	Finding	Evidence Tier
AI-Enabled Attack Volume	AI-enabled attacks rose ~89% year-over-year; breakout time compressed to 22–27 seconds	Secondary Verified
Agentic AI Exposure	21% of enterprises have runtime visibility into agent behavior; 80% report unauthorized agent actions	Reported — High

OAuth Trust Exploitation	Compromised AI vendor credentials enable cross-tenant lateral movement via legitimate OAuth — no exploit required	Secondary Verified
Model Capability Extraction	Nation-state proxies operating hydra cluster API farms to distil frontier model capabilities at industrial scale	Primary Verified
Shadow AI Prevalence	76% of organizations report shadow AI challenge; ~50% of employees circumvent bans	Reported

Threat Actor Taxonomy

Actor Class	Primary Objective	TTP Pattern	Targeted Layer
AI-Assisted External Adversary	Rapid breach & data exfiltration	Recon automation, OAuth pivot, adaptive multi-target exploitation	Identity, Orchestration, Observability
Insider / Agent Misuse	Unintended data overreach	Ambiguous instruction interpretation, excessive defaults	Validation gates, Permissions
Supply Chain Adversary	Ecosystem compromise	Malicious marketplace skills, poisoned config files	Supply chain, Orchestration
State-Sponsored Extractor	AI IP theft	Distillation campaigns, hydra clusters, geofence bypass	Identity, Observability, Model API

Notably Absent From Q1 2026 — Boundaries Explicitly Documented

- ▶ No confirmed large-scale autonomous agent runaway incidents in production environments
- ▶ No EU AI Act-specific regulatory fines — Commission enforcement powers activate August 2026
- ▶ No widespread production model poisoning — primarily documented in research settings
- ▶ No confirmed AI-generated malware with autonomous self-modification in production

— S3 —

VERIFIED INCIDENT ANALYSIS

Six incidents — forensic depth with evidence classification

Each incident is analyzed against a standard attribute set: evidence tier, attack vector, control plane failure, financial impact model, and strategic significance. Cross-reference Section 6 for MITRE technique IDs.

INCIDENT 1

I-1

AI-Assisted Multi-Agency Government Breach — Mexico

Evidence: *SECONDARY VERIFIED*

Discovery	Feb 25, 2026 — breach period: Dec 2025 – Jan 2026; discovered during threat hunting (Gambit Security/SecurityWeek/Cybernews)
Targets	~10 entities: federal tax authority (SAT), National Electoral Institute (INE), 4 state governments, Monterrey water utility, financial institution
Attack Vector	Single attacker used Claude + ChatGPT as force-multipliers: automated CVE mapping (20+), exploit scripting, simultaneous multi-agency exfiltration
Volume	~150 GB — ~195M taxpayer records, voter files, civil registry documents, government credentials
Control Plane Failure	No anomaly detection calibrated to AI-accelerated reconnaissance velocity; prolonged dwell across 10 agencies simultaneously
MITRE Alignment	T1595 (Active Scanning — AI-automated), T1078 (Valid Accounts), T1041 (Exfiltration over C2)
Financial Impact	\$200M – \$450M [ESTIMATE — Medium]: records × \$1.50–\$2.20 + \$45–70M IR/legal + \$30–60M operational × 1.2–1.4× regulatory multiplier
Strategic Significance	First documented case of commercial AI used as full-lifecycle attack tools across simultaneous multi-agency government targets. Establishes force-multiplier pattern as operational, not experimental.

“ Commercial AI tools converted a single-actor operation into a multi-agency simultaneous breach — a capability previously requiring a nation-state team

INCIDENT 2

I-2

Industrial-Scale AI Model Distillation — DeepSeek, MiniMax, Moonshot AI

Evidence: *PRIMARY VERIFIED*

Disclosure	Anthropic: Feb 23, 2026 (public). OpenAI: Feb 12, 2026 (Congressional memo). Google GTIG: corroborated
Attack Method	Hydra cluster architecture: distributed fraudulent accounts routed through third-party proxies, mixing distillation with legitimate API traffic to defeat behavioural detection
Scale	MiniMax: 13M+ exchanges (agentic coding). Moonshot AI: 3.4M+ (reasoning, CV, computer use). DeepSeek: 150K+ (reasoning + censorship-bypass generation). Total: 16M+ across ~24K accounts
Control Failure	Behavioural detection insufficient to distinguish systematic capability extraction from high-volume legitimate API usage without cross-account pattern correlation
Legal Status	No framework currently classifies unauthorized mass model distillation as actionable IP theft. ToS violation is the only current remedy — inadequate against state operations

Financial Impact	R&D value at risk conservatively in the billions. No recovery mechanism under current law.
Strategic Significance	Largest documented AI model IP theft in history. Hardware export controls are necessary but insufficient — distillation provides an alternative capability pathway for any actor with API access.

INCIDENT 3

I-3

OpenClaw Agent Framework & ClawHub Marketplace Crisis

Evidence: SECONDARY VERIFIED

Sources	Check Point Research (RCE), Antiy CERT (malicious skills), Trend Micro (MCP exposure), CBS News (Pentagon designation)
Scope	135,000+ GitHub stars; 21,000+ exposed instances; system-level OS access across corporate and SME deployments
Vector 1 — RCE	Check Point: RCE via poisoned <code>.cursor/mcp.json</code> and <code>claude_desktop_config.json</code> repository config files; committed to version control and propagated legitimately
Vector 2 — Marketplace	Antiy CERT: 1,184 malicious skills on ClawHub marketplace — credential capture, persistence, and lateral movement distributed via official install channel
Vector 3 — MCP	Trend Micro: 492 MCP servers publicly internet-exposed with zero authentication — each a potential pivot controlling all connected agents
Control Failure	No code signing requirement for marketplace skills; broad default grants at install; zero MCP authentication enforcement
Pentagon Designation	U.S. DoD designated Anthropic as a 'supply chain risk' — first such designation for an American AI company
Financial Impact	\$80M – \$120M aggregate remediation [ESTIMATE — Medium Confidence]
Strategic Significance	AI agent marketplaces replicate the malicious browser extension attack model with substantially weaker vetting. One compromised MCP server controls every connected agent.

INCIDENT 4

I-4

Meta Internal AI Agent Data Exposure — Severity 1

Evidence: PRIMARY VERIFIED

Sources	The Information, Engadget, Cyber Magazine; Meta Sev 1 classification confirmed
Sequence	Engineer posts query to internal forum → invokes AI agent to analyze → agent interprets 'return analysis' as 'publish to forum' → sensitive user-related data exposed to unauthorized engineers
Duration	~2 hours internal exposure; contained. Meta stated no external user data was mishandled.

Root Cause	Excessive agent permissions beyond task scope; no pre-action approval gate for 'publish' operations; instruction ambiguity without human confirmation path
MITRE Alignment	T1078.004 (Valid Accounts — Cloud); AML.T0058 (Prompt Extraction / Instruction Ambiguity analogue)
Financial Impact	Not publicly quantified; no external breach. Significant internal review and architectural remediation cost.
Strategic Significance	The agent passed every identity check. It was authorized. The failure was architectural — identical architecture at a healthcare or financial institution triggers mandatory breach notification.

INCIDENT 5

I-5 **Context AI / Vercel OAuth Supply Chain Compromise**
Evidence: SECONDARY VERIFIED

Disclosure	April 19, 2026. Sources: TechCrunch, Hudson Rock, Ox Security, The Hacker News
Attack Chain	Context AI employee → Lumma Stealer via Roblox exploit script → Google Workspace + OAuth tokens harvested → Attacker pivots into Vercel enterprise Workspace via OAuth grant → Internal DB, API keys, source code accessed
Exfiltration	Vercel internal database listed on BreachForums at \$2M. Downstream customer impact scope undisclosed.
Control Failure	Overbroad OAuth grant scopes; no cross-tenant access monitoring; no employee AI tool vetting; no credential rotation enforcement
MITRE Alignment	T1539 (Steal Web Session Cookie), T1528 (Steal Application Access Token), T1550.001 (Alternate Auth Material)
Financial Impact	Database listing: \$2M. Full downstream: estimated \$15M–\$40M including notification, API rotation, and IR.
Structural Lesson	No novel technique required. OAuth is the primary lateral movement vector in AI platform breaches. A single AI vendor employee device compromise can reach every enterprise customer that granted permissions.

INCIDENT 6

I-6 **Adaptive AI-Driven Infrastructure Campaign**
Evidence: REPORTED — Single Vendor Intel

Source	Foresiet Threat Intelligence (single vendor; independent corroboration pending — treat as directional)
Scope	600+ network devices (FortiGate infrastructure) across 55 countries
Attack Architecture	AI orchestration layer autonomously performs reconnaissance, vulnerability prioritization, exploit selection, lateral movement, and persistence — no human required between phases; adapts based on target environment feedback

Control Failure	Static signature-based detection architecturally unable to respond to feedback-driven, machine-speed attack sequencing; no behavioral anomaly baseline for AI-speed activity
MITRE Alignment	T1595 (Active Scanning), T1203 (Exploitation), T1133 (External Remote Services), T1071 (Adaptive C2)
Financial Impact	\$10M – \$50M aggregate remediation [ESTIMATE — Low Confidence, single-source intel]
Strategic Significance	If confirmed at scale: fully autonomous AI-driven attacks operational without human direction between phases. 22-second breakout window is insufficient for human-directed incident response.

— S4 —

FINANCIAL IMPACT

Sensitivity modeling with explicit assumptions

Metric	Documented Baseline	Modeled Range	Confidence	Derivation
Direct losses Q1	Vendor disclosures + regulatory filings	\$1.2B – \$1.6B	Medium	S3 incidents
Indirect impact	Model retraining, vendor reassessment	\$0.8B – \$1.1B	Low	Benchmarks
Total aggregate	Combined	\$2.0B – \$2.7B	Medium	S1.2 formula
Avg. AI breach cost	IBM Cost of Breach (2025)	\$4.88M per incident	High	IBM 2025

AI SECURITY INCIDENT & RISK ANALYSIS REPORT · Q1 2026			TECHNICAL & COMPLIANCE EDITION	
Shadow AI premium	IBM enterprise survey	+\$670K above baseline	Medium	IBM 2025

Sensitivity Scenarios

Scenario	Impact Range	Key Variance Drivers
Best Case	\$1.4B – \$1.9B	Rapid detection <24h; contained blast radius; no multi-jurisdiction fines
Expected Case	\$2.0B – \$2.7B	Standard 3–7 day detection; cross-tenant impact; baseline regulatory fines
Worst Case	\$3.2B – \$4.1B	Dwell >30 days; multi-jurisdiction enforcement; IP valuation loss; sector cascades

All figures capped at 25th percentile of historical analogues. Not all 320M+ records constitute PII under applicable frameworks.



— S5 —

AI CONTROL PLANE

Architecture, failure mapping & standards crosswalk

Definition: AI Control Plane

The authoritative enforcement layer responsible for identity-bound execution, policy-constrained action authorization, and verifiable observability across AI-initiated operations spanning one or more trust domains. It SHALL operate independently of the model inference layer and enforce least-privilege scoping per task.

Five-Layer Architecture — Q1 2026 Failure Mapping

Layer	Function	Q1 2026 Failure	Missing Standard	Evidence
1. Identity & Credentials	Authenticate agent; provision scoped identity; rotate credentials on lifecycle events	OpenClaw: 21K+ instances with no credential lifecycle	No binding AI agent IAM standard	I-3
2. Permissions & Scoping	Enforce least-privilege per task; prevent permission inheritance across trust boundaries	Meta leak; Context AI/Vercel OAuth over-scope	Least-privilege for agents not specified	I-4, I-5
3. Orchestration & MCP	Authenticate tool calls; scope API grants; validate workflow before execution	492 MCP servers: zero auth; config RCE	No MCP security specification exists	I-3
4. Validation Gates	Enforce human-in-the-loop for high-impact actions; resolve instruction ambiguity	Meta: agent published to forum without approval	EU AI Act Art. 14 is principle only	I-4
5. Observability & Audit	Maintain immutable decision logs; establish behavioral baselines	79% lack runtime visibility; no machine-speed baselines	No cross-jurisdiction AI incident standard	All

Standards Crosswalk — NIST AI RMF & ISO/IEC 42001

Control Plane Layer	NIST AI RMF Function	ISO/IEC 42001 Clause
Identity & Permissions	Govern (GOV-3): Define roles and accountability. Manage (MG-2): Track deployment.	6.1.2: Addressing risks and opportunities. 7.2: Competence and role assignments.
Orchestration & Validation	Map (MP-3): Identify interdependencies. Measure (ME-1): Evaluate effectiveness.	8.1: Operational planning and control. 8.2: AI system impact assessment.
Observability & Audit	Measure (ME-2): Evaluate activities. Manage (MG-4): Respond to incidents.	9.1: Monitoring and evaluation. 9.2: Internal audit.

— S6 —

MITRE ATT&CK MAPPING

Q1 2026 AI-enabled adversary technique alignment

Techniques marked [ILLUSTRATIVE] are research-demonstrated; all others have Q1 2026 production evidence.

Stage	MITRE ID	AI-Enabled Technique	Q1 Evidence	Control Failure
Recon	T1595 / AML.T0000	LLM-automated CVE enumeration, service fingerprinting, exploit script generation at machine speed	Mexico breach — 20+ CVEs mapped in hours	No AI-speed anomaly detection

AI SECURITY INCIDENT & RISK ANALYSIS REPORT · Q1 2026			TECHNICAL & COMPLIANCE EDITION	
Initial Access	T1539, T1528	Infostealer harvest of OAuth tokens; token replay across cross-tenant trust relationships	Context AI / Vercel	Overbroad OAuth scopes
Execution	T1203, T1204.002	Poisoned config files trigger RCE on agent startup; malicious marketplace skills execute on install	OpenClaw / ClawHub	No code signing
Priv Escalation	T1078.004	AI service identity inherits excessive permissions; agent action scope exceeds task requirement	Meta Sev 1; OAuth pivot	No least-privilege scoping
Persistence	T1133, T1547	Adaptive AI-driven exploit selection based on environmental feedback; maintains persistence against defensive responses	Infrastructure campaign	Static signature reliance
Capability Exfil	AML.T0044	Systematic high-volume API interactions to extract model reasoning chains as training data	Distillation campaigns	No cross-account correlation
Data Exfiltration	T1041, AML.T0051	Indirect prompt injection steering agent to send sensitive data to attacker-controlled endpoint [ILLUSTRATIVE]	[ILLUSTRATIVE]	No output filtering / DLP

— S7 —

REGULATORY EXPOSURE

Incident mapping to enforcement frameworks & deadlines

Incident-to-Regulation Mapping

Incident	Primary Regulation	Trigger Mechanism	Enforcement Status
Mexico Breach	GDPR Art. 32; Mexican national data laws	Failure to detect AI-accelerated intrusion; inadequate technical measures	Active — GDPR extraterritorial
Model Distillation	National IP law; EAR/ECRA export controls	Absence of legal definition for unauthorized distillation as IP theft	Enforcement gap — classification pending

AI SECURITY INCIDENT & RISK ANALYSIS REPORT · Q1 2026			TECHNICAL & COMPLIANCE EDITION
OpenClaw / ClawHub	EU AI Act Art. 16 (future); sectoral supply chain rules	Insecure defaults; inadequate marketplace vetting; unauthenticated endpoints	EU AI Act enforcement activates Aug 2026
Meta Internal Leak	GDPR Art. 25 (privacy by design), Art. 32	No pre-action validation; excessive permissions; absent audit trail	Active — GDPR Art. 25 obligation
Context AI / Vercel	GDPR Art. 28 (processor); SOC 2; ISO 27001	Overbroad OAuth; third-party vendor risk failure; credential management	Active — GDPR processor + breach notification
Infrastructure Campaign	NIST CSF; NIS2 (EU); sector directives	Static detection failure; inadequate adaptive defense posture	NIS2 active in EU; NIST CSF advisory in US

Enforcement Timeline

ACTIVE NOW — Multiple Frameworks Enforced Today

Deadline	Jurisdiction	Obligation	Max Penalty
Feb 2, 2025	EU — AI Act Art. 5	Prohibited practices (manipulative AI, social scoring, unauthorized biometrics)	€35M or 7% turnover
Aug 2, 2025	EU — AI Act Ch. V	GPAI model obligations: documentation, transparency, safety testing	€15M or 3% turnover
Jan 1, 2026	EU — Finland	First national AI Act enforcement powers — precedent being established now	National scale
Feb 2026	US — Colorado	AI Act: impact assessments for high-risk AI; consumer appeal rights	State AG action

UPCOMING — August 2, 2026

Deadline	Jurisdiction	Obligation	Max Fine
Aug 2, 2026	EU — Full AI Act	High-risk AI (Annex III): employment, credit, education, law enforcement. Mandatory conformity assessments, human oversight, audit logging.	€15M or 3% turnover
Aug 2, 2026	EU — GPAI Enforcement	Commission fining powers for GPAI model providers activate. Documentation and safety testing obligations fully enforceable.	€15M or 3% turnover

— S8 —

STANDARDS GAPS

Where no current framework provides actionable technical controls

These are implementation gaps, not awareness gaps. No existing framework — NIST AI RMF, OWASP Top 10 for LLMs, ISO/IEC 42001, or the EU AI Act — provides binding technical controls for these areas.

GAP-01

AI Agent IAM · Current coverage: *OWASP ASI Top 10 (threat taxonomy only)*

Missing: Binding implementation standard for agent identity provisioning, credential scoping, lifecycle management, and access revocation

Q1 Evidence: Meta leak; OpenClaw 21K+

GAP-02

MCP Security · Current coverage: *None — no standard exists*

Missing: Protocol-level authentication requirements, permission grant scoping, audit logging obligations, anomaly detection baseline

Q1 Evidence: 492 exposed servers

GAP-03

AI Supply Chain Assurance · Current coverage: *SLSA, SBOM (software only — no AI coverage)*

Missing: Model provenance documentation, training data lineage, AI dependency vetting criteria, marketplace security requirements

Q1 Evidence: ClawHub; LiteLLM; Vercel

GAP-04

Distillation Detection · Current coverage: *None — no standard exists*

Missing: Behavioral signatures for distillation detection, cross-provider threat sharing protocol, legal definition of unauthorized extraction

Q1 Evidence: Anthropic/OpenAI disclosure

GAP-05

Human Oversight — Technical Spec · Current coverage: *EU AI Act Art. 14 (principle only)*

Missing: Implementation-level controls translating 'meaningful human oversight' into verifiable, auditable system requirements

Q1 Evidence: Meta; FINRA 2026 guidance

GAP-06

AI Incident Classification · Current coverage: *None — no cross-jurisdiction standard*

Missing: Unified taxonomy mapping AI incidents to notification obligations across GDPR, EU AI Act, HIPAA, FINRA, SEC simultaneously

Q1 Evidence: Overlapping Q1 obligations

— S9 —

AUDITABLE CONTROLS

Six normative control statements — SHALL and SHOULD

Controls are aligned with NIST AI RMF, ISO/IEC 42001, and EU AI Act Article 14. SHALL = mandatory; SHOULD = strongly recommended. Designed for direct use in security policies, audit frameworks, and procurement requirements.

CTRL-01	SHALL	All AI-initiated high-impact actions — including delete, publish, financial transaction, permission modification, and external communication — SHALL require an explicit pre-execution authorization control enforced outside the agent runtime environment.	<i>EU AI Act Art. 14; NIST MG-4</i>
---------	-------	--	-------------------------------------

CTRL-02	SHOULD	AI service identities SHOULD be provisioned via enterprise IAM with task-scoped, time-bound permissions and automated credential rotation on a maximum 90-day cycle. Service identities SHOULD NOT inherit permissions from human user accounts.	ISO 42001 §6.1.2; NIST GOV-3
CTRL-03	SHALL	All MCP servers and AI orchestration endpoints SHALL enforce mutual authentication, scope-limited API grants reflecting least-privilege per task, and immutable audit logging of all tool invocations, input parameters, and output responses.	OWASP Agentic Top 10; NIST ME-2
CTRL-04	SHALL	AI vendor integrations SHALL require documented SBOM including model provenance, cryptographic artifact signing, and demonstrated incident response maturity (SOC 2 Type II or equivalent) prior to production deployment.	NIST SP 800-161; ISO 42001 §8.1
CTRL-05	SHOULD	Organizations SHOULD deploy behavioral baseline monitoring and anomaly detection calibrated to agent interaction velocity, distinct from human-speed telemetry. Anomaly thresholds SHOULD be reviewed quarterly as deployment scale increases.	NIST ME-1; ISO 42001 §9.1
CTRL-06	SHALL	All high-risk AI systems under EU AI Act Annex III SHALL maintain a queryable decision log capturing: agent identity, input context, tool calls, tool responses, validation gate state, and human override events with timestamps.	GDPR Art. 25; EU AI Act Annex III §12

— APP —

APPENDICES

Regulatory deadlines, scope & limitations, sources

Appendix A — Regulatory Deadlines Reference

Deadline	Jurisdiction	Obligation	Status
Feb 2, 2025	EU — AI Act Art. 5	Prohibited practices enforceable	ACTIVE
Aug 2, 2025	EU — AI Act Ch. V	GPAI model obligations applicable	ACTIVE
Jan 1, 2026	EU — Finland	National AI Act enforcement powers operational	ACTIVE
Jan 1, 2026	US — IN, KY, RI	New comprehensive state privacy laws	ACTIVE

AI SECURITY INCIDENT & RISK ANALYSIS REPORT · Q1 2026			TECHNICAL & COMPLIANCE EDITION
Feb 2026	US — Colorado	Colorado AI Act: impact assessments for high-risk AI	ACTIVE
Aug 2, 2026	EU — Full AI Act	High-risk AI (Annex III) full enforcement	UPCOMING
Aug 2, 2026	EU — GPAI Enforcement	Commission fining powers for GPAI providers activate	UPCOMING
Aug 2, 2027	EU — AI Act Legacy	High-risk AI embedded in regulated products	UPCOMING

Appendix B — Sources

- ▶ Anthropic Security Disclosure — AI model distillation campaign (February 23, 2026)
- ▶ OpenAI memorandum to U.S. House Select Committee on China (February 12, 2026)
- ▶ Gambit Security / SecurityWeek / Cybernews — Mexico government breach (February–March 2026)
- ▶ Trend Micro / TrendAI — Agentic AI threat intelligence and MCP exposure analysis (Q1 2026)
- ▶ Check Point Research — Claude Code RCE vulnerability disclosure (February 2026)
- ▶ Antiy CERT — OpenClaw / ClawHub malicious skill analysis (February 2026)
- ▶ Foresiet Threat Intelligence — Adaptive AI infrastructure campaign (March–April 2026)
- ▶ Hudson Rock / TechCrunch / Ox Security — Context AI / Vercel breach analysis (April 2026)
- ▶ Gravitee — State of AI Agent Security 2026
- ▶ VentureBeat Pulse — Enterprise AI Agent Security Survey, Q1 2026 (n=108)
- ▶ IBM Cost of a Data Breach Report (2025 edition)
- ▶ World Economic Forum Global Cybersecurity Outlook 2026
- ▶ RSAC 2026 — Google Threat Intelligence, Cisco, CrowdStrike disclosures
- ▶ EU AI Act (Regulation (EU) 2024/1689) and Commission implementation guidance
- ▶ DLA Piper GDPR Fines and Data Breach Survey (January 2026)
- ▶ OWASP Top 10 for Agentic Applications 2026 (published December 2025)
- ▶ NIST AI Risk Management Framework 1.0 and supporting playbooks
- ▶ ISO/IEC 42001:2023 — AI Management System

Appendix C — Scope & Limitations

- ▶ Financial impact figures not directly disclosed use explicit benchmarking formulas with conservative (25th-percentile) assumptions. Actual costs vary by scale, sector, and regulatory exposure.
- ▶ Claims labeled [SECONDARY VERIFIED] or [REPORTED] have not received independent official confirmation. Treat as directional indicators.
- ▶ EU AI Act high-risk enforcement activates August 2026. Commission implementation decisions may affect timelines.
- ▶ Coverage is limited to publicly reported incidents between January–March 2026. Unverified single-source reports and vendor marketing claims are excluded.
- ▶ This report does not constitute legal, regulatory, financial, or professional advice.