Incident Response Framework

AI Security Incidents vs. Non-AI Incidents Guidance for Credit Unions

By Barry Lewis

Optiri

Contents ✨

1. Executive Summary

Credit unions are increasingly integrating artificial intelligence into their operations, from fraud detection and loan underwriting to member chatbots and anti-money laundering (AML) screening. While these systems create significant operational value, they also introduce a new class of security incidents that existing incident response (IR) plans may not adequately address.

This document provides a detailed comparison of how incident response procedures should differ when a credit union faces an AI-related security incident versus a traditional, non-AI security incident. AI incidents introduce unique challenges around model integrity, data poisoning, adversarial attacks, bias exploitation and third-party AI vendor risk that require specialized detection capabilities, different escalation paths, additional regulatory considerations and modified remediation strategies.

The goal is not to replace existing IR frameworks but to augment them with AI-specific guidance so credit unions can respond quickly, effectively and in compliance with applicable regulations when their AI systems are compromised or behave in unintended and harmful ways.

2. Scope and Applicability

This topic applies to all credit unions that deploy, host or consume AI/ML-powered tools in any capacity, including but not limited to:

- Fraud detection and transaction monitoring models
- Automated loan origination and credit scoring engines
- Chatbots and virtual assistants used for member interactions
- Anti-money laundering (AML) and Bank Secrecy Act (BSA) screening systems
- Document processing and optical character recognition (OCR) tools
- Marketing personalization and recommendation engines
- Robotic process automation (RPA) with embedded ML components
- Third-party AI services accessed via APIs (e.g., identity verification, voice authentication)

For the purposes of this document, a "non-AI incident" refers to conventional cybersecurity events such as malware infections, phishing compromises, network intrusions, denial-of-service attacks, insider threats and data breaches that do not involve the manipulation, failure or compromise of an AI/ML system.

3. Defining AI Security Incidents at a Credit Union

An AI security incident at a credit union is any event in which the confidentiality, integrity, or availability of an AI/ML system is compromised, or where the AI system produces outcomes that cause financial harm, regulatory noncompliance, member harm or reputational damage due to adversarial action, system failure or undetected model degradation.

3.1 Categories of AI Security Incidents

AI security incidents at credit unions generally fall into the following categories, each of which has no direct equivalent in traditional incident response:

DATA POISONING:
Malicious alteration of training or input data to corrupt model behavior. For example, an attacker subtly modifies transaction records to train the fraud model to ignore certain fraud patterns.

ADVERSARIAL ATTACKS:
Specially crafted inputs designed to cause a model to make incorrect predictions. For instance, manipulating loan application data just enough to trick an automated underwriting model into approving a fraudulent loan.

MODEL THEFT OR EXTRACTION:
Systematic querying of an AI system to reverse-engineer or replicate its decision-making logic, potentially exposing proprietary credit scoring criteria.

PROMPT INJECTION / MANIPULATION:
Exploiting natural language interfaces (e.g., member-facing chatbots) to bypass controls, extract sensitive data or cause the system to produce harmful outputs.

MODEL DRIFT AND SILENT FAILURE:
Gradual degradation of model accuracy due to changing data patterns that go undetected, leading to systematically biased or erroneous decisions affecting members.

BIAS EXPLOITATION:
Deliberate or accidental amplification of discriminatory patterns in AI-driven lending, pricing or service delivery, creating fair lending violations.

AI SUPPLY CHAIN COMPROMISE:
A third-party AI vendor or open-source model component is compromised, introducing vulnerabilities into the credit union's AI pipeline.

4. Phase-by-Phase Comparison: AI vs. Non-AI Incident Response

The following subsections walk through each standard phase of incident response and detail how credit union teams should adjust their approach when the incident involves an AI system.

4.1 Preparation

Traditional (Non-AI)

Standard preparation for non-AI incidents includes maintaining an up-to-date IR plan, training staff on phishing awareness and security hygiene, running tabletop exercises simulating ransomware or data breach scenarios, maintaining relationships with law enforcement and forensic vendors and ensuring backups and disaster recovery plans are tested. The incident response team typically consists of IT security staff, legal counsel, communications and executive leadership.

AI-Specific Additions

For AI incidents, preparation must extend well beyond traditional cybersecurity readiness. Credit unions must maintain a comprehensive AI asset inventory documenting every model in production, including its purpose, data sources, training methodology, vendor, version and risk classification. The IR team must be expanded to include data scientists or ML engineers who understand model architecture and can evaluate whether a model has been compromised. Preparation should include AI-specific tabletop exercises that simulate scenarios like data poisoning, adversarial loan applications or a chatbot prompt injection. Credit unions must also establish model performance baselines and continuous monitoring thresholds so that anomalous model behavior can be detected. Vendor contracts must include AI-specific incident notification requirements, audit rights and model transparency obligations.

4.2 Detection and Analysis

Traditional (Non-AI)

Detection of non-AI incidents relies on well-established tools and techniques: SIEM alerts, intrusion detection systems, antivirus signatures, network traffic analysis, endpoint detection and response (EDR) and user-reported anomalies. Analysis involves correlating indicators of compromise (IOCs) such as malicious IPs, file hashes and known attack signatures. The scope of impact is generally deterministic—either data was exfiltrated or it was not, either a system was compromised or it was not.

AI-Specific Differences

Detection of AI incidents is fundamentally more difficult because AI failures can be subtle, probabilistic and may not trigger any traditional security alerts. A poisoned fraud model, for example, may still appear to function normally on most transactions while systematically failing to flag a specific type of fraud. Detection requires specialized monitoring, including model performance drift detection, statistical analysis of prediction distributions, input data integrity checks and anomaly detection on model outputs. Analysis of AI incidents requires evaluating training data provenance, examining whether model weights have been tampered with, assessing whether adversarial inputs are being systematically submitted and determining whether bias metrics have shifted. Unlike traditional incidents where the damage is usually binary, AI incident impact may be a spectrum – the model may be partially compromised, affecting only certain populations or transaction types, making scoping significantly more complex.

4.3 Containment

Traditional (Non-AI)

Containment for non-AI incidents follows familiar playbooks: isolate affected systems from the network, disable compromised accounts, block malicious IP addresses, quarantine infected endpoints and preserve forensic evidence. The goal is to stop the spread of the attack while maintaining as much business continuity as possible. Fallback to manual processes is generally well-understood.

AI-Specific Differences

Containing an AI incident is considerably more nuanced. Simply taking an AI system offline can have severe operational consequences – if the fraud detection model is disabled, the credit union loses its primary defense against real-time fraud, potentially increasing losses dramatically. Containment decisions must weigh the risk of a compromised model continuing to operate against the risk of operating without it entirely. Options include rolling back to a previous known-good model version, switching to a rule-based fallback system, increasing human review of AI-generated decisions, restricting the model's scope (e.g., limiting it to low-risk transactions only) or isolating the model's data pipeline while keeping the model running on clean cached data. Additionally, if the incident involves a third-party AI service, the credit union may have limited ability to contain the issue directly and must rely on the vendor while potentially activating contract-based SLA provisions.

4.4 Eradication

Traditional (Non-AI)

Eradication in traditional incidents involves removing the threat entirely: wiping and reimaging compromised systems, patching exploited vulnerabilities, removing malware, rotating credentials and closing attack vectors. The process is generally deterministic and can be validated through scanning and testing.

AI-Specific Differences

Eradicating an AI security incident is far less straightforward. If training data has been poisoned, the model itself is compromised and cannot simply be "cleaned" – it must be retrained on validated data, which can take days or weeks. If adversarial inputs are the issue, input validation and preprocessing pipelines must be redesigned and hardened. If model weights were stolen, the competitive and security value of the model may be permanently lost, requiring a fundamental redesign. Eradication must also address the root cause in the data pipeline, feature engineering, model training infrastructure or API security – areas that often sit outside the traditional IT security team's expertise. Full eradication may require retesting the model against fairness and accuracy benchmarks, revalidating it against regulatory standards and obtaining sign-off from both data science and compliance teams before returning it to production.

4.5 Recovery

Traditional (Non-AI)

Recovery from non-AI incidents involves restoring systems from clean backups, bringing patched systems back online, monitoring for reinfection, gradually restoring full operations and verifying data integrity. Recovery timelines are generally predictable and tied to system restoration.

AI-Specific Differences

Recovering from an AI incident requires additional steps that extend well beyond system restoration. The credit union must validate that the retrained or restored model performs within acceptable accuracy and fairness parameters before returning it to production. This involves running the model through a comprehensive test suite including performance metrics, bias testing and adversarial robustness testing. Any decisions the compromised model made during the incident window must be identified and reviewed – this could mean auditing thousands of loan decisions, fraud alerts or member interactions. Members affected by incorrect AI-driven decisions may need to be notified and made whole. Recovery may also require updating model monitoring thresholds based on lessons learned, implementing new input validation controls and renegotiating vendor agreements. The credit union should operate with heightened human oversight of the AI system for a defined period after recovery to ensure stability.

4.6 Lessons Learned

Traditional (Non-AI)

Post-incident reviews for traditional incidents focus on how the attack succeeded, what controls failed, how quickly the team detected and responded, what procedural improvements should be made and whether additional training or tools are needed. Findings are documented and fed back into the IR plan.

AI-Specific Differences

Post-incident analysis for AI incidents must go deeper, examining the model lifecycle for weaknesses that enabled the incident. This includes reviewing data governance practices, model validation procedures, monitoring gaps, vendor oversight and the AI-specific elements of the IR plan itself. The credit union should assess whether its model risk management framework (aligned with SR 11-7/OCC 2011-12 guidance) adequately addresses adversarial threats. Lessons learned should drive updates to the AI asset inventory, model monitoring thresholds, AI-specific tabletop scenarios and vendor contract requirements. Findings should be shared with the board of directors and relevant NCUA examiners as appropriate, particularly if the incident resulted in member harm or regulatory noncompliance.

5. Regulatory and Compliance Considerations

Credit unions face a distinct regulatory landscape when AI systems are involved in a security incident. The following considerations extend beyond standard breach notification obligations.

5.1 NCUA Expectations

The National Credit Union Administration (NCUA) expects credit unions to have robust risk management frameworks for all technology systems, including AI. An AI security incident may trigger examination scrutiny beyond what a traditional cyber incident would provoke. Examiners may evaluate the credit union's AI governance structure, model validation practices and the adequacy of oversight by the board and senior management. Credit unions should be prepared to demonstrate that they identified and managed AI-specific risks prior to the incident and responded appropriately.

5.2 Fair Lending and ECOA / Regulation B

If a compromised AI model was involved in credit decisions, the credit union must evaluate whether the incident resulted in disparate impact or disparate treatment in violation of the Equal Credit Opportunity Act (ECOA) and Regulation B. A traditional data breach does not typically raise fair lending concerns, but a compromised or drifted AI lending model may have systematically disadvantaged protected classes. The credit union may need to conduct a retrospective fair lending analysis, report findings to regulators and provide remediation to affected members.

5.3 BSA / AML Compliance

If the AI system involved in the incident performs transaction monitoring or suspicious activity detection for BSA/AML compliance, a compromise could mean that suspicious activity went undetected and unreported during the incident window. The credit union must assess whether Suspicious Activity Reports (SARs) should have been filed but were not and may need to file retroactive SARs and notify FinCEN. This obligation does not exist for traditional cyber incidents that do not affect BSA/AML monitoring systems.

5.4 State and Federal Breach Notification

Standard breach notification rules apply to both AI and non-AI incidents when member information is compromised. However, AI incidents may create additional notification complexity. For example, if a chatbot was manipulated to disclose member information through prompt injection, the scope of the breach may be difficult to determine. If an AI-driven identity verification system was compromised, the credit union must determine whether synthetic identities were approved or legitimate members were improperly denied service. Emerging state AI-specific legislation may impose additional notification or disclosure requirements related to automated decision-making failures.

5.5 CFPB and Adverse Action Requirements

Under the Consumer Financial Protection Bureau's guidance, if an AI system was involved in adverse credit decisions, the credit union must be able to provide specific and accurate reasons for the adverse action. If the AI model was compromised, the credit union may be unable to demonstrate that the reasons provided were accurate, creating potential violations. Remediation may require revisiting all adverse action notices issued during the incident window.

6. Roles and Responsibilities Differences

AI incidents require an expanded incident response team with specialized expertise that traditional incidents do not demand.

Traditional (Non-AI) Incident	IR PHASE/AREA	AI Security Incident
ISO or IT Security Manager coordinates response across IT, legal, communications.	**IR TEAM LEAD**	Same lead, but must coordinate additional stakeholders including data science, model risk management and potentially AI vendor contacts.
Network engineers, system administrators and security analysts perform forensics.	**TECHNICAL SME**	Data scientists, ML engineers and AI platform administrators are essential for analyzing model behavior, training data and pipeline integrity.
Privacy officer and compliance team assess breach notification and regulatory requirements.	**COMPLIANCE**	Must also assess fair lending, BSA/AML, ECOA and emerging AI-specific regulatory implications.
Contact affected software or service vendors for patches or support.	**VENDOR MGMT.**	Must engage AI-specific vendors with model-level questions, request training data audit trails and potentially invoke AI-specific contract provisions.
Briefed on material incidents per governance policy.	**BOARD/EXEC**	May require additional briefing on AI-specific risks, model risk management adequacy and strategic implications for the credit union's AI roadmap.
Advises on breach notification, liability and law enforcement engagement.	**LEGAL COUNSEL**	Must also advise on AI liability, algorithmic discrimination claims, intellectual property (model theft) and evolving AI regulation.
Handles member inquiries about data breaches.	**MEMBER SERVICES**	Must also handle inquiries about incorrect AI-driven decisions, explain automated decision-making processes and coordinate remediation for affected members.

7. Communication and Member Notification

7.1 Internal Communications

Both AI and non-AI incidents require timely internal communication. However, AI incidents require additional communication to staff who rely on AI-driven tools. If the fraud model is compromised, frontline staff must be alerted to increase manual review. If the chatbot is taken offline, member service representatives must be prepared for increased call volume. If the lending model is rolled back, loan officers need guidance on interim procedures. Clear internal communication about the nature of the AI failure – without causing unnecessary alarm – is critical.

7.2 Member-Facing Communications

Member notification for AI incidents requires additional sensitivity. Members may not be aware that AI was being used in the credit union's processes, and learning about it in the context of a security incident can erode trust. Communications should be transparent about what happened, how members were affected, what remediation is being provided and what the credit union is doing to prevent recurrence. If the AI system made incorrect decisions affecting individual members (e.g., wrongful loan denials, missed fraud alerts), personalized outreach may be necessary. Unlike a traditional data breach where notification can be somewhat standardized, AI incident notifications may need to be tailored based on how each member was affected.

7.3 Regulatory and Examiner Communications

In addition to standard breach notifications, the credit union should proactively communicate with NCUA examiners about material AI incidents, particularly those involving fair lending, BSA/AML or consumer protection implications. Demonstrating transparency and a robust response can mitigate examination findings and enforcement risk.

8. Post-Incident Analysis and Lessons Learned

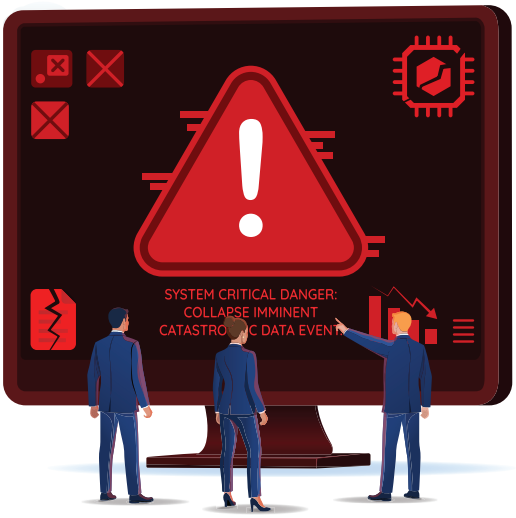
8.1 Standard Post-Incident Review

For both AI and non-AI incidents, the credit union should conduct a thorough post-incident review within 30 days, documenting the timeline, root cause, response effectiveness and improvement actions. This aligns with general IR best practices and NCUA expectations.

8.2 AI-Specific Post-Incident Considerations

AI incidents warrant additional post-incident analysis:

1. **Model Lifecycle Review:** Assess where in the model lifecycle (data collection, training, deployment, monitoring) the vulnerability existed and how it went undetected.
2. **Data Governance Assessment:** Evaluate whether data quality controls, access controls and/or provenance tracking were sufficient to prevent or detect the incident.
3. **Model Validation Gap Analysis:** Determine whether the credit union's model validation processes (per SR 11-7 or equivalent) adequately tested for adversarial robustness and failure modes.
4. **Monitoring Threshold Calibration:** Update model monitoring thresholds, alerts and dashboards based on the specific indicators that were missed or delayed.
5. **Vendor Performance Review**: If a third-party AI system was involved, evaluate the vendor's incident response, transparency and whether contract terms need to be strengthened.
6. **AI Governance Framework Update:** Feed findings into the credit union's broader AI governance framework, including risk appetite statements, acceptable use policies and board reporting.
7. **Regulatory Impact Assessment:** Document any regulatory implications and remediation actions taken, building an audit trail for examiners.



9. AI-Specific Incident Response Playbook Additions

Credit unions should supplement their existing IR playbooks with the following AI-specific runbooks:

9.1 Playbook: Data Poisoning Incident

1. Isolate the data pipeline feeding the affected model.
2. Identify the time window and scope of data contamination.
3. Roll back the model to a version trained before the contamination window.
4. Audit all decisions made by the model during the contamination window.
5. Validate the integrity of training data sources and implement enhanced input validation.
6. Retrain the model on verified clean data.
7. Conduct performance, fairness and robustness testing before redeployment.
8. Notify affected members and regulators as required.

9.2 Playbook: Adversarial Input/Prompt Injection

1. Log and quarantine the adversarial inputs for forensic analysis.
2. Implement emergency input filtering or rate limiting.
3. Assess whether the adversarial inputs resulted in unauthorized access, data leakage or incorrect decisions.
4. If a chatbot is involved, temporarily increase human review or take the bot offline.
5. Work with the AI vendor or internal team to harden input validation and implement adversarial robustness controls.
6. Test the hardened system against known adversarial techniques before restoring full service.

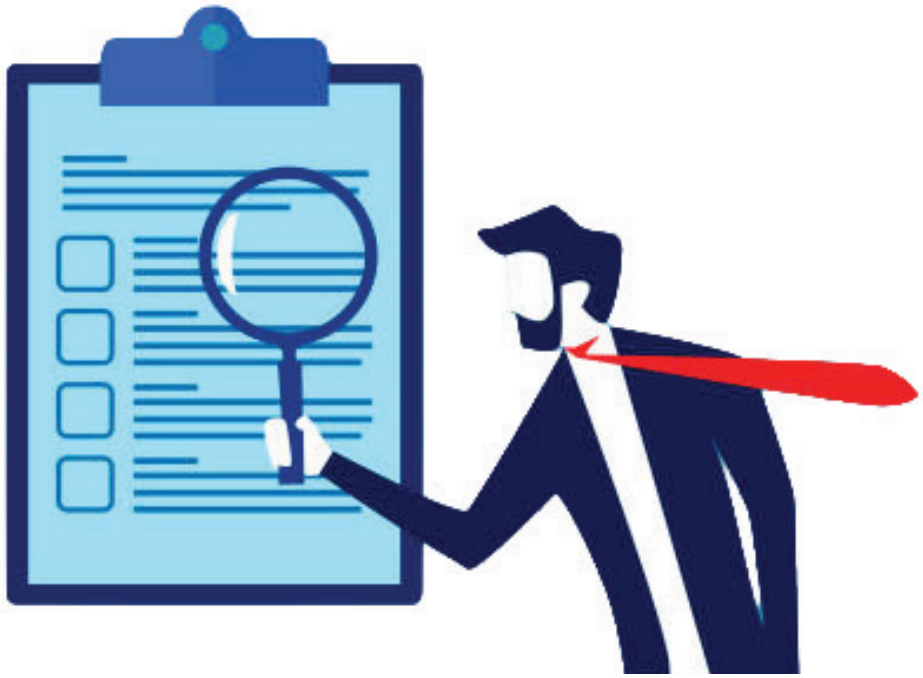
9.3 Playbook: Model Drift/Silent Failure

- Trigger an investigation when model performance metrics fall outside defined thresholds for a sustained period.
- Analyze whether drift is due to changing data patterns (concept drift) or data quality issues (data drift).
- Assess member impact: identify populations or transaction types disproportionately affected.
- Determine whether drift resulted in fair lending, BSA/AML or consumer protection violations.
- Retrain or recalibrate the model and update monitoring thresholds.
- Conduct a retrospective review of all decisions in the drift window and remediate affected members.

9.4 Playbook: AI Vendor Compromise

- Activate vendor incident response provisions in the service contract.
- Assess the scope of the vendor compromise and its impact on the credit union's systems and data.
- Determine whether member data was exposed through the vendor's systems.
- Evaluate whether to continue using the vendor's service, switch to a fallback or operate manually.
- Coordinate with the vendor on root cause analysis, remediation timeline and future prevention.
- Report to NCUA and other regulators as appropriate, particularly regarding third-party risk management.

Update the IR Plan: Incorporate AI-specific incident categories, escalation paths and containment procedures into the existing incident response plan.



10. Recommendations and Next Step

To ensure readiness for AI security incidents, credit unions should take the following actions:

1. **Update the IR Plan:**
   Incorporate AI-specific incident categories, escalation paths and containment procedures into the existing incident response plan.

2. **Build AI Expertise on the IR Team:**
   Ensure data scientists or ML engineers are designated as part of the incident response team with defined roles and on-call procedures.

3. **Conduct AI Tabletop Exercises:**
   Run at least one AI-specific tabletop exercise annually, simulating scenarios like data poisoning, adversarial attacks and vendor compromises.

4. **Maintain an AI Asset Inventory:**
   Document all AI/ML systems in production with risk classifications, data sources, vendors and model validation status.

5. **Implement AI-Specific Monitoring:**
   Deploy model performance monitoring, data drift detection and output anomaly detection tools alongside traditional security monitoring.

6. **Review Vendor Contracts:**
   Ensure all AI vendor agreements include incident notification timelines, audit rights, model transparency requirements and data handling obligations.

7. **Align with Regulatory Expectations:**
   Review NCUA guidance, FFIEC standards and emerging federal and state AI regulations to ensure the IR framework meets evolving expectations.

8. **Report to the Board:** Provide the board of directors with regular updates on AI risk management, including the state of AI incident preparedness and any incidents that have occurred.

Appendix A:
Quick-Reference Comparison Table

The following table provides a high-level comparison of key differences across the incident response lifecycle.

Traditional (Non-AI) Incident	IR PHASE/AREA	AI Security Incident
Standard IR plan, tabletop exercises, staff awareness training, backup and recovery procedures.	**PREPARATION**	AI asset inventory, AI-specific tabletop exercises, data science team integration, model performance baselines, vendor AI clauses.
SIEM, IDS/IPS, EDR, antivirus, network monitoring, user reports.	**DETECTION**	Model drift monitoring, output anomaly detection, data integrity checks, fairness metric tracking, adversarial input detection.
IOC correlation: IPs, file hashes, attack signatures. Binary scope (compromised or not).	**ANALYSIS**	Model behavior analysis, training data provenance review, statistical distribution analysis. Impact is a spectrum, not binary.
Isolate systems, disable accounts, block IPs, quarantine endpoints.	**CONTAINMENT**	Roll back model, switch to rules-based fallback, restrict model scope, increase human review, isolate data pipeline.
Reimage systems, patch vulnerabilities, remove malware, rotate credentials.	**ERADICATION**	Retrain model on clean data, redesign input validation, rebuild data pipeline, revalidate fairness and accuracy.
Restore from backups, bring systems online, monitor for reinfection.	**RECOVERY**	Validate retrained model performance, audit incident-window decisions, remediate affected members, heightened oversight period.
Attack vector analysis, control gap assessment, procedure improvements.	**LESSONS LEARNED**	Model lifecycle review, data governance assessment, vendor performance review, AI governance framework update, regulatory impact assessment.
Breach notification (state/federal), NCUA reporting.	**REGULATORY**	All standard + fair lending review, BSA/AML gap analysis, adverse action notice review, CFPB implications, emerging AI regulation compliance.
Standard breach notification to members.	**COMMUNICATION**	Tailored notifications based on individual AI-decision impact, transparency about AI use, personalized remediation.

Appendix B:
AI Incident Classification Matrix

Use the following matrix to classify AI security incidents by severity and determine the appropriate response level.

Severity	Example	Member Impact	Regulatory Risk	Response Time	Escalation
Critical	Fraud model fully compromised; loans approved for synthetic identities	Significant financial loss or harm to multiple members	High: Fair lending, BSA/AML, CFPB	Immediate (within 1 hour)	CEO, Board, NCUA, Legal
High	Chatbot leaking member PII via prompt injection	Personal data exposure for subset of members	Moderate: Breach notification, privacy	Within 4 hours	ISO, Compliance, Legal
Medium	Lending model drifting, showing bias in certain demographics	Potential disparate impact on protected classes	Moderate: Fair lending review needed	Within 24 hours	ISO, Data Science, Compliance
Low	Minor chatbot misbehavior, no data exposure, no incorrect decisions	Minimal to no direct member impact	Low: Monitor and document	Within 72 hours	IT Security, Data Science


Optiri
A Trellance Company