Synthetic Users Incident Response Plan (IRP)

Version: 1.2

Effective Date: 16 July 2025

Last Updated: March 25, 2026

Owner: CTO — Artur Ventura

Approved By: CTO — Artur Ventura

Change History

Version	Date	Author	Changes
1.1	July 16, 2025	Security & Compliance Lead	Initial release
1.2	March 25, 2026	Artur Ventura, CTO	Added AI/GenAI incident classification and response procedures (Section 7); added external threat intelligence sources (Section 5); updated scope to explicitly include AI/GenAI systems; added JPMC client notification obligations; updated related documents. Per JPMC SCA CRA 14.1.1.

1. Purpose

This plan establishes Synthetic Users' structured approach to detecting, responding to, containing, and recovering from security incidents that could compromise data confidentiality, integrity, or availability.

2. Scope

This plan applies to all Synthetic Users systems, employees, contractors, and third parties handling company or customer data.

It covers cybersecurity events, data breaches, unauthorized access, system compromises, and other incidents affecting business continuity or data protection obligations.

This plan explicitly covers **AI/GenAI systems**, including incidents involving prompt injection, data leakage via model outputs, model misuse, abnormal model behaviour, and third-party LLM provider events. AI/GenAI incidents are subject to the same classification, escalation, testing, and root cause analysis requirements as all other incidents.

3. Roles and Responsibilities

Role	Responsibility
Incident Response Lead (IRL)	Coordinates all response activities, approves external communications, and reports to the CTO.
Security Engineer / IT Team	Investigates, contains, and eradicates threats; preserves forensic evidence.
Legal & Compliance	Ensures adherence to GDPR and breach notification obligations.
Communications Lead	Manages internal and external messaging, including client and public statements.
Executive Team	Provides strategic oversight and approves public disclosures.

Contact details for all roles are maintained in the internal Security Runbook.

4. Phases of Incident Response

4.1 Detection and Identification

- Immediately log and triage all suspected incidents via the internal incident tracking system.
- Validate the event using system logs, intrusion detection alerts, or anomaly reports.
- Classify severity (Low, Medium, High, Critical) based on scope and potential impact.
- Notify the Incident Response Lead within **1 hour** of detection.

4.2 Containment

Short-Term Containment

- Disconnect or isolate affected systems from networks.
- Revoke compromised credentials or access tokens.
- Capture volatile evidence (e.g., memory dumps, logs) before system reboot.

Long-Term Containment

- Apply temporary security controls (e.g., firewall rules, access restrictions) to prevent spread.
- Initiate monitoring for related suspicious activity.

4.3 Investigation and Assessment

- Conduct forensic analysis to determine the cause, entry point, and scope.
- Preserve all evidence securely for legal or compliance purposes.
- Document each action in the **Incident Report Log**.
- Assess affected data types (personal, confidential, operational).
- Engage Legal & Compliance for impact classification under GDPR or other applicable laws.

4.4 Eradication and Recovery

- Eliminate malicious code, unauthorized access, or misconfigurations.
- Validate that compromised accounts and systems are fully remediated.
- Restore systems from verified clean backups.
- Conduct integrity checks before reconnecting systems to production.
- Strengthen controls that failed (patches, access restrictions, monitoring).

4.5 Notification and Compliance

- Notify affected customers (including the Affected Organization) **without undue delay** and within legal timeframes (e.g., 72 hours under GDPR).
- Provide clear, factual information about:
  - Nature and scope of the incident
  - Data involved
  - Mitigation measures taken
  - Recommended actions for affected parties
- Coordinate regulatory notifications through Legal & Compliance.
- **JPMC client notification:** If the incident affects JPMC data, JPMC systems, or services provided under the JPMC engagement, notify JPMC within **72 hours** of confirming the incident. Provide JPMC with incident scope, data involved, containment status, and remediation timeline. Follow up with a written post-incident report within 10 business days of closure.
- Maintain documentation of all communications for audit and compliance review.

4.6 Post-Incident Review

- Conduct a **Post-Incident Review Meeting** within **10 business days** of closure.
- Analyze incident root cause, timeline, and response effectiveness.
- Record lessons learned and assign owners for remediation tasks.
- Update relevant policies (Access Control, Encryption, Password Management, etc.) based on findings.
- File the finalized **Incident Report** in the Security Repository.

5. Continuous Improvement

- Perform at least one **tabletop incident simulation per year** to test this plan. Tabletop scenarios include at least one AI/GenAI-specific scenario annually (e.g., prompt injection, data leakage via LLM output, LLM provider outage).
- Audit incident handling logs quarterly for completeness and accuracy.
- Revise the plan annually or after any major incident, system change, or regulatory update.

5.1 External Threat Intelligence

Synthetic Users actively monitors the following external intelligence sources to stay current on emerging threats and vulnerabilities:

- **Cloud provider security advisories** — AWS Security Bulletins, Render status and security notifications
- **CVE/NVD feeds** — National Vulnerability Database for disclosed CVEs affecting dependencies and infrastructure
- **Vendor vulnerability disclosures** — Security notifications from critical vendors and LLM providers (OpenAI, Anthropic, Google, Firebase, Google Workspace)
- **Open-source dependency alerts** — Dependabot automated alerts for vulnerable open-source packages
- **AI/GenAI threat intelligence** — OWASP LLM Top 10 updates, emerging prompt injection and model exploitation techniques

Intelligence is triaged by the CTO and Engineering Lead and fed directly into vulnerability management and incident response workflows.

6. Incident Severity Classification

Severity	Definition	Initial Response Time	Executive Notification
Critical	Active breach or data exfiltration; JPMC data or customer PII confirmed affected; ransomware or total system compromise	Immediate — within 1 hour	CEO + CTO within 1 hour
High	Suspected breach; unauthorized access to production systems; AI/GenAI prompt injection with confirmed data leakage	Within 4 hours	CTO within 1 hour; CEO within 4 hours
Medium	Anomalous model behaviour; failed injection attempt; access control violation with no confirmed data loss; vendor security advisory requiring action	Within 24 hours	CTO within 4 hours
Low	Near-miss event; dependency vulnerability with no active exploitation; policy violation with no security impact	Within 5 business days	Engineering Lead; CTO if recurring

7. AI/GenAI Incident Response

AI/GenAI incidents are handled under this plan using the same phases (Detection → Containment → Investigation → Eradication → Notification → Post-Incident Review) with the following additional guidance.

7.1 AI/GenAI Incident Types

Incident Type	Description	Initial Containment Action
Prompt Injection	A user-crafted input overrides system prompt instructions, potentially exposing system prompt contents, other tenants' data, or executing unauthorized actions	Disable the affected endpoint or feature; rotate affected API keys; preserve prompt and response logs for forensic review
Data Leakage via Model Output	LLM output contains data from another tenant's RAG index, from the model's training data, or PII not intentionally included in the prompt	Identify affected sessions; notify affected tenants; review RAG retrieval scope and tenant isolation controls
Model Misuse	Platform is used to generate harmful, deceptive, or policy-violating content in violation of the Acceptable Use Policy	Suspend offending account; preserve session logs; review content filtering configuration
LLM Provider Security Event	A third-party LLM provider discloses a security breach, data retention issue, or vulnerability affecting API data	Activate LLM Shuffle fallback to alternate provider; obtain written confirmation from provider on data exposure scope; assess DPA compliance
Abnormal Model Behaviour	LLM outputs deviate significantly from expected behaviour in ways that suggest model compromise, poisoning, or provider-side incident	Switch affected model to fallback provider via LLM Shuffle; log anomalous outputs; notify provider; notify CTO immediately
AI Service Disruption	LLM provider unavailability causes platform degradation for customers including JPMC	Activate fallback provider; communicate estimated resolution to affected customers; notify JPMC if JPMC engagement is affected

Prompt/Context Window Exfiltration	Evidence that system prompt contents, JPMC data, or internal context have been extracted by a user or external party	Treat as High/Critical breach; follow full breach response; notify JPMC within 72 hours if JPMC data was involved

7.2 AI/GenAI Forensic Preservation

When investigating an AI/GenAI incident, the following artefacts must be preserved:

- Full prompt and response logs for the affected session(s), including system prompt and context window contents
- RAG retrieval logs showing which documents were retrieved for the affected query
- LLM provider API request and response metadata (timestamps, model version, token counts)
- Application-level access logs for the affected user/tenant
- Content filtering decisions and any flagged outputs

Logs must be captured before any remediation action that could alter or delete them.

7.3 AI/GenAI Annual Testing

The annual tabletop exercise must include at least one AI/GenAI scenario. Suggested scenarios include:

- A user successfully performs a prompt injection that retrieves another tenant's data
- An LLM provider discloses a security incident affecting API data during an active JPMC engagement
- Abnormal output patterns are detected across multiple research sessions suggesting model-level compromise

Test findings and remediation actions are documented and retained per the Information Governance & Records Management Standard.

---

8. Related Documents

- SDLC AI/GenAI Addendum — AI/GenAI model lifecycle and security controls including adversarial testing
- AI/GenAI Algorithm Design Document — System architecture and data flow relevant to AI/GenAI forensic investigation
- Third-Party Risk Management Policy — LLM provider risk classification and vendor incident response SLAs
- Access Rights Review Policy — Access control policy referenced in eradication and recovery
- Encryption Policy — Encryption standards relevant to data breach assessment
- Password Management Policy — Credential management referenced in containment
- Business Continuity Plan — Continuity procedures for prolonged AI/GenAI service disruption
- Information Governance & Records Management Standard — Incident record retention requirements