

SIMULATED CASE STUDY

# AI Incident Response Playbook

Sensitive-data exposure and incorrect customer response

|           |                                                                 |
|-----------|-----------------------------------------------------------------|
| Project   | AI governance (Project 4)                                       |
| Company   | Northstar Financial Technologies (fictional)                    |
| Status    | Simulated playbook                                              |
| Scenarios | Sensitive-data exposure; materially incorrect customer response |

This document is part of a portfolio exercise. Northstar Financial Technologies, its people, systems, records and findings are invented. No real audit was performed, no assurance is given and nothing here is a certification or compliance claim.

# 1. Purpose and scope

Steps to follow when an AI system used by Northstar exposes sensitive data or gives a customer materially incorrect output. It supports the general incident process and does not replace it. Times below are Northstar internal targets, not legal deadlines.

## 2. Severity

| Severity | Meaning                                                                                      | Examples                                                                                               |
|----------|----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| Sev 1    | Confirmed exposure of Restricted or personal data to an unauthorised party, or customer harm | Customer ticket data visible to another tenant; wrong instruction followed by a customer to their loss |
| Sev 2    | Suspected exposure, or a materially incorrect reply reached a customer                       | Support reply with the wrong fee sent; confidential document uploaded to an unapproved tier            |
| Sev 3    | Near miss or an error caught before it reached anyone                                        | Agent catches a wrong draft; prompt injection blocked                                                  |

## 3. Roles

| Role                                | Duty                                                                                    |
|-------------------------------------|-----------------------------------------------------------------------------------------|
| Incident lead (IT Security Manager) | Runs the response, keeps the timeline                                                   |
| Business owner                      | Decides on customer-facing actions with Legal and the Privacy Officer                   |
| Privacy Officer                     | Assesses whether personal information is involved and leads the notification assessment |
| Legal                               | Advises on legal and contractual obligations, privilege and communications              |
| Model or system owner               | Suspends the tool, preserves logs, applies fixes                                        |
| Head of Procurement                 | Contacts the vendor under the contract                                                  |
| Chief Risk Officer                  | Informed of every Sev 1 and Sev 2; decides on risk acceptance and escalation            |

## 4. Steps

### Step 1. Detect and report

Anyone who sees a suspected incident reports it to the IT Security Manager at once and does not delete anything.

### Step 2. Triage (target: within 4 hours)

Confirm what happened, which use case, what data or customers are involved, and the severity.

### Step 3. Contain (target: tool suspended within 1 hour of a Sev 1 or Sev 2 confirmation)

Suspend the tool or the affected feature, revoke keys and uploads access if needed, stop sending AI-drafted replies, and ask the vendor to stop processing and preserve records.

### Step 4. Preserve evidence

Export prompts, outputs, retrieved sources, user and approver ids, timestamps and vendor logs; record file hashes; restrict access to the copies; keep the original logs unchanged; place a hold on related tickets.

### Step 5. Decide on notification

The Privacy Officer and Legal assess whether personal information was exposed and what reporting or customer notice is required. This playbook does not state legal deadlines; the applicable obligations and timelines are confirmed with counsel when the incident occurs. Check contractual notice duties to and from the vendor.

## Step 6. Communicate

Business owner and Legal agree the message to affected customers or staff; support agents receive a script; the vendor is notified in writing.

## Step 7. Recover

Correct the cause, re-run the evaluation and security tests, get approval from the AI Governance Committee to resume, and monitor closely for 30 days.

## Step 8. Lessons learned (target: within 10 business days)

Blameless review; update controls, thresholds, training and this playbook; record in the incident log and the risk register.

## 5. Scenario A: sensitive-data exposure

Trigger: confidential or personal data is found in a place it should not be, for example an uploaded document is visible to another user, or the vendor reports its handling of our content. Extra actions: identify every document or ticket involved from upload logs, request the vendor's deletion confirmation, determine who could have seen the data, and review whether the use case's data class rules were followed.

## 6. Scenario B: materially incorrect customer response

Trigger: a customer received an AI-drafted reply that is wrong in a way that matters, such as fees, deadlines or service status. Extra actions: contact the customer with a correction, search sent replies for the same article or pattern, check whether the approving agent reviewed it, and decide whether to pause the use case based on the error-rate threshold.

## 7. Evidence checklist

| Item                                              | Source                     | Keeper              |
|---------------------------------------------------|----------------------------|---------------------|
| Prompt, retrieved sources, draft and final output | Tool logs                  | System owner        |
| Approver and timestamp                            | Tool logs                  | System owner        |
| Vendor processing and deletion confirmation       | Vendor                     | Head of Procurement |
| Ticket or document ids involved                   | Support or document system | Incident lead       |
| Timeline of decisions and who made them           | Incident log               | Incident lead       |

## 8. Limitations

A fictional, simulated playbook. It has not been rehearsed. Legal and regulatory duties are deliberately not stated and must be confirmed by qualified advisers at the time.