**AI Incident Response Plan (AIRP)**

**Formål:** At sikre hurtig identifikation, inddæmning og udbedring af hændelser specifikt relateret til AI-modeller og deres output.

**1. Klassificering af AI-Hændelser**

Vi skelner mellem tre typer af AI-specifikke hændelser:

| Type | Beskrivelse | Eksempel |
| --- | --- | --- |
| **Model Drift** | Modellens præcision falder drastisk over tid. | En sagsbehandlings-AI begynder pludselig at give forkerte anbefalinger pga. nye lovkrav. |
| **Algoritmisk Bias** | Modellen udviser diskriminerende adfærd. | SHAP-analysen viser, at postnummer vægter uforholdsmæssigt højt i afgørelser. |
| **Adversarial Attack** | Forsætlig manipulation af input eller model. | "Prompt Injection" eller manipulation af træningsdata (Data Poisoning). |

Eksportér til Sheets

**2. Response Workflow (5-trins modellen)**

**Trin 1: Identifikation (Monitoring & Alerts)**

* **Trigger:** MLOps-overvågning slår alarm pga. bias-afvigelser, eller en bruger rapporterer en "mærkelig" afgørelse.
* **Handling:** **Account Security Manager (ASM)** indkalder straks en **Security Champion** fra det relevante team.

**Trin 2: Inddæmning (The Kill Switch)**

For at beskytte borgerens retssikkerhed (KOMBIT Princip 9), har vi tre inddæmningsniveauer:

1. **Level 1 (Soft Rollback):** Skift til den forrige stabile version af AI-modellen via K8s.
2. **Level 2 (Shadow Mode):** Modellen kører stadig, men dens output skjules for brugeren og logges kun til analyse.
3. **Level 3 (Kill Switch):** AI-modulen deaktiveres helt, og systemet skifter til manuel sagsbehandling eller fallback-logik.

**Trin 3: Analyse (Root Cause)**

* **Audit-spor:** ASM og teamet trækker **Inference Logs** for den pågældende hændelse.
* **XAI-Tjek:** Kør en **SHAP/LIME-analyse** på de fejlagtige input for at forstå, hvorfor modellen traf den pågældende beslutning.
* **Data-tjek:** Blev modellen trænet på korrupt eller biased data? (Tjek **CycloneDX ML-BOM**).

**Trin 4: Udbedring (Recovery)**

* **Gen-træning:** Modellen gen-trænes med korrigerede data eller nye vægtninger (Constraints).
* **Validering:** Den nye model skal gennemgå en fuld **SAI-SDLC testcyklus** (SonarQube, Bias-test, Champion-review) før gen-deployment.

**Trin 5: Læring (Post-Mortem)**

* Resultaterne af hændelsen føres ind i det næste **Sikkerhedsretrospektiv**.
* Tjeklister i Azure DevOps (DoD) opdateres for at forhindre gentagelse.

**3. Eskalationsmatrix (Hvem gør hvad?)**

* **Teknisk ansvarlig:** Security Champion (Udfører Rollback).
* **Strategisk ansvarlig:** Account Security Manager (Vurderer risiko jf. FAIR).
* **Kommunikationsansvarlig:** Projektlederen (Informerer kunden/KOMBIT).

**4. Værdi for kunden (KOMBIT)**

Med denne plan kan du fortælle KOMBIT:

*"Vi ved, at AI ikke er ufejlbarlig. Derfor har vi ikke bare bygget en model, men også en 'nødbremse'. Hvis vores overvågning spotter bias eller fejl, kan vi øjeblikkeligt rulle tilbage eller deaktivere AI-delen uden at lægge hele systemet ned. Det er jeres garanti for borgerens retssikkerhed."*