AI Incident Reporting

Reporting an AI incident protects everyone: the team, the users and the organization.

Why is it important?

- Detect flaws, misuse and harmful biases before they escalate.
- It complies with regulatory frameworks and facilitates auditing and accountability.
- Protect people, data and business continuity.
- Turn every incident into learning: corrective actions and system improvements.

What is reported?

- Wrong or dangerous outputs (critical hallucinations, toxic or biased content).
- Data leaks, unauthorized access, credential/API exposure.
- Unexpected agent behaviors (actions outside of limits...)
- Security attacks: prompt injection, jailbreaks, shadow AI (unauthorized BYOAI).
- Impacts on customers or employees (discrimination, reputational damage, impact).

You can create a traffic light with illustrative examples.

How to report?

1. Open an easy-to-access channel: intranet, dedicated email, form + QR code.
2. Complete the mini-form: what happened, where, system, impact, level of impact.
3. Send and receive immediate acknowledgment: folio/ticket and defined response times.

Internal response flow

Detection the employee notices an anomaly.

Report send a report through the secure channel.

Assessment IA Resp team reviews the case.

Answer corrective action is communicated.

Resp. AI team

1. Triage
2. Research
3. Remediation
4. Closure and learning

Benefits of reporting incidents

Strengthens the integrity of the system

Avoid reputational damage

Reduce regulatory risks

Promotes continuous learning and improvement

Protection for the reporter

Non-retaliation policy and anonymous reporting options.

Created by Karine Boucher - karine.ai
Passionate about human-centered Responsible AI and AI Governance

linkedin.com/in/karineboucher/

“Conscience is the key to turning code into responsible action.”