Generative AI Risk Tracker & Response Toolkit

A Practical Framework for Identifying, Tracking, and Managing Generative AI Risks

Introduction

As generative AI becomes embedded across industries, it brings transformative potential—alongside significant risk.

From security vulnerabilities and misinformation to ethical dilemmas and regulatory uncertainty, organizations are increasingly exposed to risks they may not yet be prepared to manage.

This toolkit is designed to equip decision-makers, security professionals, product owners, and compliance teams with the tools and frameworks necessary to actively manage generative AI risks.

Whether you're overseeing the deployment of AI tools internally or developing customer-facing solutions, this guide offers a structured approach to anticipate, track, mitigate, and respond to AI-related threats.

Section 1: Risk Logging Template

The first step in responsible AI adoption is risk visibility. The following template helps you log, categorize, and monitor generative AI risks across business units or product teams.

Risk ID	Risk Category	Description	Likelihood	Impact	Current Controls	Owner	Status
GAIR-001	Cybersecurity Threats	Risk of prompt injection exposing user credentials via chatbot interaction	High	High	Input filtering, user access control	InfoSec Team	Open
GAIR-002	Intellectual Property	AI-generated content reproducing licensed material from training datasets	Medium	High	Output screening, content moderation tools	Legal	Monitoring
GAIR-003	Misinformation	Public-facing AI tool provides medically inaccurate responses	Medium	Medium	Retrieval-augmented generation (RAG), expert QA	Product Team	Mitigated

How to Use: Maintain this log centrally and update it during model training, deployment, and feedback cycles. This ensures visibility across stakeholders and supports informed decision-making.

Section 2: Incident Response Flowchart for Generative AI Failures

When a risk escalates into an incident, the ability to respond quickly and effectively is critical. The following seven-step framework provides a blueprint for addressing generative AI incidents.

1. Detect

Monitor AI system behavior continuously. Use automated alerts and user reports to detect unusual or harmful outputs.

2. Classify

Determine the nature of the risk:

- Is it a hallucination?
- A security breach?
- A violation of ethical or legal boundaries?

Classify by category and severity.

3. Contain

Temporarily disable or isolate the affected model, application, or component. Prevent further use until investigation is complete.

4. Investigate

Conduct root cause analysis. Evaluate:

- Prompts and user inputs
- Model responses and confidence levels
- Training data and contextual sources

5. Resolve

Implement a fix based on findings. This may include:

- Prompt adjustment
- Model retraining
- Policy or access control updates

6. Document

Capture a full incident report including timeline, impact, response actions, and lessons learned. Archive for internal and external audits.

7. Prevent

Update risk controls, team training, and development protocols to prevent recurrence.

Section 3: Mitigation Strategy Matrix

Use this reference to map specific generative AI risks to appropriate mitigation strategies.

Risk Type	Example	Recommended Mitigation
Bias Amplification	Resume screening tool favors male applicants for leadership roles	Implement bias audits and diverse training data sources
Misinformation	AI assistant gives outdated or incorrect tax advice	Use retrieval-augmented generation (RAG) with verified data
Prompt Injection	Malicious prompt tricks model into revealing private user data	Sanitize inputs; restrict prompt chaining
Copyright Infringement	Generated content closely resembles licensed material	Implement IP detection; use licensing filters during output generation
Model Inversion Attacks	Sensitive data reconstructed from model outputs	Employ differential privacy techniques and monitor model behavior
Data Poisoning	Malicious actor manipulates training data to bias model output	Validate datasets; secure data pipelines; run pre-training integrity checks

Section 4: AI Risk Monitoring Checklist

This checklist is intended to support continuous monitoring of generative AI systems and mitigate emerging issues early.

Weekly Review Activities

- Review AI-generated outputs for anomalies or offensive content.
- Monitor user-submitted prompts for unusual behavior.
- Check system access logs for unauthorized interactions.
- Audit performance metrics and model drift indicators.

Monthly Review Activities

- Update the organizational AI risk register.
- Review risk statuses and mitigation outcomes from the past month.
- Conduct qualitative reviews of any flagged incidents.
- Re-evaluate the AI usage policy based on new regulatory or industry developments.

Section 5: Safer Prompt Design Guidelines

Poorly constructed prompts can unintentionally result in hallucinations, sensitive disclosures, or unethical responses. Use these guidelines to strengthen prompt design.

Use Case	Risk-Prone Prompt	Improved Prompt (Safe and Clear)
Legal Drafting	"Write a termination clause for this contract"	"List commonly used termination clauses in general contracts without offering legal advice."
Health Recommendations	"Suggest a treatment for migraine headaches"	"List typical treatments used for migraines and include a disclaimer about consulting a doctor."
Performance Feedback	"Write a negative performance review"	"Create a professional feedback example focused on constructive improvement."

These improvements help minimize legal liability, offensive outputs, and hallucinated recommendations.

Section 6: Integration and Deployment Best Practices

To ensure organizational readiness, this toolkit should be embedded into core workflows across departments:

- **Security & IT:** Integrate risk tracking with your SIEM, IAM, or logging infrastructure.
- **Legal & Compliance:** Use the incident documentation and mitigation matrix for IP audits and regulatory reviews.
- **HR & Training:** Provide AI safety awareness and prompt-writing training to internal teams.
- **Product Teams:** Incorporate risk reviews as part of your development sprints and deployment pipeline.

Section 7: Toolkit Resources (Suggested Format)

The following materials are suggested to complement this content in downloadable form:

1. **Risk Tracker Template** – Excel or Google Sheets version of the risk logging table
2. **Incident Response Checklist** – Printable PDF for cross-functional teams
3. **Mitigation Matrix Poster** – High-level view for product teams and AI leads
4. **Monitoring Checklist** – Weekly and monthly task lists in editable document format
5. **Prompt Guidelines Sheet** – Recommended prompts and rewrite examples for safe use
6. **Governance Policy Template** – Starter document for internal use or external compliance

Conclusion

Generative AI is unlocking innovation at unprecedented speed—but without proper oversight, it poses serious risks to data security, brand integrity, user safety, and public trust.

This Generative AI Risk Tracker & Response Toolkit provides the practical foundation to manage these risks proactively.

By establishing formal tracking processes, clear response protocols, and robust mitigation strategies, your organization can not only minimize exposure but also demonstrate leadership in responsible AI adoption.

A secure, ethical, and future-ready AI program begins with visibility—and ends with accountability.

GSDC
Global Skill Development Council

CERTIFICATION IN GENERATIVE AI IN RISK AND COMPLIANCE

Get global recognition and stand out as a leader in the field of Generative AI In Risk And Compliance.

ABOUT GSDC CERTIFICATION

LIFETIME VALIDITY

GSDC Certification is an globally accreditted certification with lifetime validity.

EBOOK

Extensive and exclusive Ebook created by world's experts to help you with understanding core concepts.

CREATED BY EXPERTS

GSDC certifications are created and authored by world's leading experts in the field.

LEARNING MATERIALS

Get access to learning materials such as videos, ebooks, templates, and practice exams, which will help you clear the certification exam.

LEARNING OBJECTIVE

- **It helps with policy management and coordinates it with businesses' current policies and processes.**
- **Generative AI successfully stimulates various scenarios and allows risk managers to assess potential impacts and plans.**
- **It can be used in the various operations of risk mitigation and its implementation strategies.**
- **It contributes to better scanning and evaluates pending legislation.**

Enroll now with the code **LEARN20** To avail **20%** discount

Enroll Now

www.gsdcouncil.org