safeshield®

The Step-by-Step AI Risk Assessment Guide

A condensed reference for AI Governance, Risk, and Compliance professionals

March 2026

The Five Phases of Risk Assessment (Abridged)

Phase 1: Understanding the AI System

A risk assessment begins with clarity. You need a complete picture of what the AI system is, what it is intended to do, and how it fits into existing business processes. This includes understanding its purpose, the decision it supports or automates, and who may be affected by its outcomes. The goal is to establish context so that later risk decisions are grounded in real operational impact.

Begin by clarifying the system's intended use, expected benefits, and limitations. Document how it integrates with workflows and what assumptions developers have made. This early understanding creates the foundation for every risk decision that follows.

Summary Checklist

- □ Define purpose, scope, and intended users
- □ Document assumptions, limitations, and context
- □ Map how the system integrates into workflows

Phase 2: Identifying AI Risks

Once the context is established, the next step is to identify where risks may emerge. Technical risks relate to the model's performance, stability, security, and susceptibility to drift. Ethical and societal risks include fairness concerns, potential harm, explainability limitations, or decisions that may undermine trust.

Operational risks arise from how the system is deployed and used, including misunderstandings, misuse, or failures in human oversight. Legal and compliance risks stem from regulatory obligations, documentation gaps, or alignment with standards such as GDPR, the EU AI Act, or ISO/IEC 42001. The goal in this phase is to create a full picture of all relevant risks before prioritisation begins.

Summary Checklist

- □ Technical risks (performance, drift, robustness, security)
- □ Ethical risks (bias, fairness, explainability)
- □ Operational risks (oversight, misuse, process failure)
- □ Legal and compliance risks (GDPR, EU AI Act, ISO/IEC 42001 alignment)

Phase 3: Evaluating and Prioritising Risks

Once risks have been identified, you assess their likelihood and impact. Some risks may be highly probable but low in impact; others may be unlikely but carry significant organisational, legal, or societal consequences. This evaluation helps determine overall severity.

The priority assigned to each risk guides what happens next. Critical risks demand immediate action before deployment, while medium and low risks may be managed through monitoring or routine oversight. During this stage, you also determine where human oversight is required and how it should function in practice. This step transforms a long list of risks into a clear, actionable set of priorities.

Summary Checklist

- □ Assess likelihood and impact
- □ Classify severity
- □ Determine where human oversight is required

Phase 4: Designing and Implementing Controls

With priorities established, the next stage is designing the controls that will manage the risks. Preventative controls aim to stop issues from occurring, detective controls help identify when something goes wrong, and corrective controls define how problems will be addressed.

Controls must be proportionate to the level of risk and aligned with regulatory expectations. This phase also requires documenting why specific controls were chosen and how they address the risks identified earlier. Finally, the controls must be integrated into the AI lifecycle so that governance is built into development, deployment, and ongoing operations.

Summary Checklist

- □ Select preventative, detective, and corrective controls
- □ Document risk treatment decisions
- □ Integrate controls into the AI lifecycle

Phase 5: Monitoring and Continuous Assessment

AI systems evolve, which means risks evolve as well. Ongoing monitoring ensures you can detect drift, fairness issues, performance changes, misuse, or unexpected system behaviours. Clear triggers should be established for when a new assessment is required, such as major model updates, data changes, or regulatory developments.

Regular reporting keeps stakeholders informed and ensures governance decisions remain aligned with business objectives, risk appetite, and regulatory requirements. Continuous monitoring is paramount for a risk assessment to become a living process, rather than a one-time exercise.

Summary Checklist

- □ Establish monitoring metrics and thresholds
- □ Define triggers for reassessment
- □ Report findings to stakeholders

Next Steps

Use this guide as a compact reference when conducting or contributing to AI risk assessments. For deeper implementation skills—including how to apply frameworks like ISO/IEC 42001, the EU AI Act, and the NIST AI RMF—explore our AI GRC training catalogue.

safeshield®

About Safeshield

Safeshield specializes in cybersecurity and AI governance, risk management, and compliance training, providing comprehensive solutions to help businesses safeguard against potential cyber threats and compliance risks. Our services include:

- Security training for employees and executives
- AIMS and ISMS Implementation
- Risk assessments
- Security audits
- More

Visit us at https://www.safeshield.cloud for more information or to view our library of training courses and personalized services.