Artificial Intelligence Risk Management Policy. V1.0

1. Purpose and Objective

The purpose of this AI Risk Management Policy ("Policy") is to establish a structured and consistent approach for identifying, assessing, mitigating, and monitoring risks associated with the development, deployment, and maintenance of Artificial Intelligence (AI) systems. This Policy ensures that the organisation's AI activities remain compliant with legal and regulatory requirements, uphold ethical standards, and align with the organisation's broader strategic objectives and defined risk appetite. *(RM-1)*

2. Scope

2.1 Applicability

This Policy applies to all AI initiatives, projects, products, and services developed or utilised by the organisation, including but not limited to:

- Proprietary AI models built in-house.
- Third-party AI tools integrated into operational processes.
- Pilot or proof-of-concept AI deployments within any business unit.

2.2 Exclusions

Certain low-risk or experimental AI activities may be granted limited exceptions if they do not pose significant material, legal, ethical, or operational risks. All requests for exceptions must follow the procedures outlined in Section 12 of this Policy. *(RM-1)*

2.3 Related Policies

This Policy complements and should be read in conjunction with the AI Governance Policy and any other corporate policies pertaining to data protection, information security, ethics, and compliance. *(GL-3)*

3. Definitions

For the purposes of this Policy, the following definitions apply:

- **AI System**: Any software application employing machine learning, deep learning, natural language processing, or other advanced computational techniques to simulate human intelligence or decision-making processes.
- **Risk**: The combination of the likelihood of an event and its potential consequences—financial, operational, reputational, or otherwise—that could adversely affect the organisation's objectives. For avoidance of doubt, this includes the potential for harm to a customer, employee or other stakeholder as a direct result of the organisations development or use of an AI system.
- **Risk Appetite**: The level and type of risk the organisation is willing to accept in pursuit of its strategic objectives.
- **Risk Tolerance**: The acceptable variance or thresholds around the organisation's risk appetite for specific categories of AI risk (e.g., legal, operational, reputational).
- **Risk Register**: A centralised repository (document or system) in which identified risks, their assessments, and their respective controls are recorded and maintained.
- **Control Owner**: An individual or team assigned responsibility for implementing and maintaining specific risk controls to ensure ongoing risk mitigation.

Additional AI-related terminology, if not defined here, shall align with definitions provided in the organisation's AI Governance Policy or corporate glossary.

4. Risk Appetite and Risk Tolerance

4.1 Risk Appetite Statement

The Organisation recognizes the transformative potential of AI but also acknowledges the inherent risks associated with advanced data-driven technologies. To balance innovation with responsible deployment, the Organisation maintains a **moderate to low** risk appetite for AI-related initiatives. This means the Organisation is open to adopting AI solutions that offer substantial operational or strategic benefits, provided that associated risks—such as bias, privacy breaches, regulatory non-compliance, or reputational harm—are clearly identified, assessed, and effectively controlled. *(RS-5, RO-1)*

4.2 Risk Tolerance Levels

Risk tolerance levels provide clarity on how much risk is acceptable within specific categories. For AI projects, the following guidelines apply:

- **Regulatory and Legal Compliance**: **Minimal Tolerance.** The Organisation will not engage in AI activities that carry a high risk of violating applicable laws or regulations. *(RO-1)*
- **Data Privacy and Security**: **Low Tolerance.** The Organisation emphasizes strong data governance practices. Use of personal or sensitive data in AI systems must follow stringent privacy and security controls. *(PR-1)*
- **Ethical and Reputational Impact**: **Low to Moderate Tolerance.** AI systems that could result in significant ethical dilemmas or reputational damage must undergo thorough review and require approval from the Risk Management Committee. *(RS-5, CO-2)*
- **Operational Disruption**: **Moderate Tolerance.** The Organisation allows a reasonable level of experimentation with AI solutions, but projects that present a high risk of severe operational disruption are subject to heightened controls and monitoring. *(OM-1)*

4.3 Ongoing Assessment of Appetite and Tolerance

The Risk Management Committee (RMC) shall review the stated risk appetite and tolerance levels at least annually or in response to major organisational or regulatory changes. Revisions to these thresholds require formal endorsement by Executive Leadership.

5. Policy Statement

The Organisation is committed to systematically identifying, assessing, and mitigating risks arising from AI systems at every stage of their lifecycle. All AI initiatives shall be subject to rigorous risk evaluation processes, aligned with the Organisation's risk appetite (Section 4.1) and tolerance levels (Section 4.2). This Policy establishes clear accountabilities and controls to ensure AI-related risks are managed proactively, transparently, and in compliance with applicable regulations and internal guidelines. *(RM-1)*

6. Risk Management Framework

This Policy implements a structured and proactive framework for managing AI risk, covering the full lifecycle of AI risk activities—risk categorisation, identification, assessment, and control selection. The framework ensures that all AI-related risks are surfaced early, assessed consistently, treated appropriately, and monitored effectively throughout the AI System lifecycle. The core components of this framework are described below, supported by additional guidance published as required by the RMC.

6.1 Risk Identification

AI risks should be identified systematically, beginning as early as the ideation or proposal stage of a project and continuing through development, deployment, and operations. Teams are expected to inquire beyond generic checklists to uncover AI-specific risk types, including technical, ethical, operational, and strategic risks. *(RM-2, IM-1)*

Recommended methods for risk identification include:

- **Pre-mortem analysis**: Envisioning plausible future failures and working backward to identify contributing factors.
- **Incident pattern mining**: Learning from historical AI failures and external incident databases.
- **Horizon scanning**: Identifying risks across short, medium, and long-term timeframes.
- **Red-teaming**: Simulating adversarial attacks or misuse scenarios.
- **Dependency chain analysis**: Mapping out upstream and downstream components that influence the AI system's stability or integrity.

At least one structured technique (e.g. a pre-mortem or red-teaming workshop) must be applied for each major AI project. Identified risks shall be recorded in the AI Risk Register along with their source, description, and associated context. Risks must be revisited periodically and updated following incidents or system changes.

6.2 Risk Assessment

Each identified risk must be formally assessed according to guidance published by the RMC. This includes four key elements:

- **Likelihood**: How probable it is that the risk will occur, scored on a five-point qualitative scale.
- **Impact severity**: The potential harm or consequence of the risk materialising, considering legal, operational, financial, or reputational damage.
- **Impact velocity**: How quickly the risk could unfold and affect the organisation, if triggered.
- **Dynamic feedback effects**: Whether the risk could self-amplify or create cascading failures, such as model feedback loops or emergent behaviours.

A 5x5 Risk Matrix is used to combine likelihood and impact. Risks with fast velocity or feedback potential must be prioritised even if their raw risk score appears moderate. These dynamic factors are particularly relevant to AI due to the self-adjusting nature of many systems and their potential for non-linear effects.

The overall risk classification (e.g., Low, Medium, High, Critical) will determine the required treatment pathway and level of governance oversight. *(RM-2, RM-3)*

6.3 Risk Control and Mitigation

For all risks rated Medium or higher, appropriate control measures must be selected. Control options and selection criteria are described in guidance developed by the organisation's AI Risk Management Committee. Controls should be proportionate to the risk's assessed level and aligned with the organisation's stated risk appetite and tolerance thresholds.

Examples of standard control types include:

- **Technical controls**: Fairness testing, drift detection, model explainability enhancements, rate-limiting, or fail-safes.
- **Process controls**: Human-in-the-loop mechanisms, approval gates, governance reviews.
- **Organisational controls**: Role-based access, escalation procedures, and training.
- **Third-party risk controls**: SLAs and contractual safeguards where AI systems are sourced externally.

Each control must be assigned to a **Control Owner** who is responsible for implementation and for updating the risk's treatment status. Where mitigation is not feasible, risks may be escalated for acceptance or avoidance decisions in accordance with defined approval thresholds.

Residual risk—after controls have been applied—must be reassessed to confirm whether it falls within acceptable tolerance with the application of the selected controls. *(RM-3, RS-5)*

6.4 Risk Monitoring

AI systems must be monitored on an ongoing basis for any indicators that an identified risk is emerging or that existing controls are no longer effective. This includes monitoring:

- Input data characteristics (e.g. distribution shifts, data quality degradation)
- Model performance metrics (e.g. accuracy, fairness, precision/recall)
- System behaviour (e.g. anomalies, unexpected user outputs)
- User feedback, complaints, or error reports

Monitoring procedures must be documented for each AI system and aligned with the controls outlined in Section 6.3. Where possible, automated monitoring tools should be implemented. Thresholds and triggers must be established for key indicators, and any breaches must be escalated in accordance with the organisation's Incident Management procedures. *(RM-4, IM-1, IM-2)*

6.5 Continuous Improvement

The risk management framework must evolve with changing technologies, use cases, and regulatory expectations. The Risk Management Committee shall review emerging AI risk types and control approaches on a regular basis. Lessons learned from internal incidents, external case studies, audits, and stakeholder feedback shall inform periodic revisions to:

- The risk taxonomy and identification methods
- The scoring and prioritisation model
- The catalogue of control strategies and treatment guidelines

This ensures the Organisation maintains a forward-looking, adaptive posture to AI risk, avoiding static frameworks that fail to reflect the real-world behaviour of complex AI systems. *(RM-4, GL-1)*

7. Roles and Responsibilities

- **Project Leaders**: Oversee initial and ongoing risk identification, ensuring all relevant risks are recorded and assessed against established risk tolerance levels.
- **Risk Management Committee (RMC)**: Provides oversight for AI risk management activities, reviews high-risk items or risks that exceed tolerance thresholds, and approves or denies exception requests. Responsible for revisiting the organisation's AI risk appetite in collaboration with executive leadership.
- **Control Owners**: Implement and maintain assigned controls, reporting on their effectiveness to the RMC as required.
- **Legal and Compliance**: Monitors AI initiatives for regulatory compliance, alerts the RMC to significant legal or ethical concerns, and recommends additional controls if necessary.
- **Data Protection Officer (DPO)**: Evaluates data-related risks, including privacy, data retention, and access controls for AI models that process personal or sensitive data.

8. Governance and Oversight

The RMC, functioning under the overarching AI Governance Policy (GL-1), shall:

- Conduct regular reviews of the Risk Register to ensure alignment with the organisation's risk appetite.
- Evaluate escalated risks (particularly those exceeding defined tolerance levels) and determine appropriate mitigation strategies.
- Communicate significant findings or policy updates to executive leadership and relevant business units, ensuring alignment with organisational objectives specified in the AI Governance Policy.

9. Implementation and Maintenance

- **Policy Distribution**: This Policy shall be disseminated to all relevant stakeholders, including project teams, data scientists, and management personnel involved in AI initiatives.
- **Adoption Timeline**: AI initiatives must conform to this Policy within three months of the effective date or within a timeframe set by the RMC.
- **Ongoing Maintenance**: The RMC shall review this Policy at least annually or more frequently if warranted by technological, regulatory, or organisational changes.

10. Monitoring, Review, and Continuous Improvement

- **Performance Indicators**: The RMC shall monitor metrics such as the frequency of escalated risks, resolution times, and overall compliance with risk tolerance thresholds. *(RM-4)*
- **Annual Audit**: An annual audit of AI risk management practices shall be conducted by Internal Audit or an external auditor to evaluate compliance with this Policy and recommend enhancements where necessary.
- **Policy Revisions**: Substantive changes to this Policy, including updates to risk appetite statements or role assignments, require RMC endorsement and Executive Leadership approval.

11. Training and Awareness

Employees, contractors, and relevant stakeholders involved in AI initiatives are required to complete training on this Policy and associated relevant guidance *(RO-2)*. The organisation's Learning and Development team, in coordination with the RMC, shall develop and deliver training materials focusing on:

- Understanding the organisation's risk appetite and tolerance.
- Recognising and reporting potential AI risks.
- Applying risk controls and adhering to escalation protocols.

12. Compliance and Enforcement

All personnel involved in AI projects are expected to comply with this Policy. Non-compliance or disregard for established risk management procedures may result in disciplinary action, in accordance with HR policies. The RMC and the Legal & Compliance department shall investigate and address any violations of this Policy, ensuring that corrective actions are taken promptly. *(RO-1)*

13. Exceptions and Exemptions

All requests for exceptions to this Policy must be documented and submitted to the RMC. Each request must include a clear justification of the business need, an analysis of potential risks, alignment with the organisation's risk appetite, and any compensating controls to be implemented. The RMC may grant temporary or conditional exemptions at its discretion if the overall risk remains acceptable.

14. References and Related Documents

- **AI Governance Policy** – Outlines the Organisational structure, roles, and ethical principles for AI initiatives.
- **Data Protection Policy** – Addresses data privacy and security requirements for AI systems.
- **Risk Assessment Methodology** – Detailed process for identifying, classifying, and monitoring AI risks, published and maintained by the RMC.

15. Effective Date

This AI Risk Management Policy is effective as of the date approved by Executive Leadership. All AI initiatives must comply with the requirements of this Policy within the timeframe determined by the RMC.

16. Version Control and Revisions

Revisions to this Policy shall be documented in a version control log maintained by the RMC. Major changes require review and approval by Executive Leadership prior to implementation.

Approved by:

[Name], Chief Technology Officer

Date:
[Date]