Overview

This document contains **Responsible AI Framework,** a comprehensive governance, risk, and oversight model designed to guide organisations in the safe, ethical, and compliant adoption of AI systems.

This framework integrates global standards such as:

- **NIST AI RMF**
- **ISO/IEC 42001 (AI Management System)**
- **OECD AI Principles**
- **Microsoft, Google & Meta Responsible AI practices**

1. Responsible AI Principles

1.1 Human-Centered Design
1.2 Ethical Data Use
1.3 Transparency & Traceability
1.4 Safety & Security by Design
1.5 Responsible Deployment
1.6 Continuous Monitoring & Auditability
1.7 Accountability & Governance Alignment

2. AI Governance Structure

2.1 Strategic Governance Layer
- AI Governance Board
- RAI Executive Sponsor
- CISO / Legal / Compliance
- Ethics Review Teams

2.2 Tactical Governance Layer
- AI Program Manager
- AI Risk Manager
- Data Stewards
- Model Owners

2.3 Operational Layer
- Data Scientists
- ML Engineers

- MLOps & Monitoring Teams

3. AI Model Lifecycle

3.1. GOVERN

- Use case classification
- Ethical/legal review
- Data access governance
- Approval workflows

3.2. MAP

- Data analysis & lineage
- Bias detection
- Model assumptions
- Documentation (Model Cards, Data Cards)

3.3. MEASURE

- Fairness testing
- Explainability evaluation
- Adversarial robustness
- Privacy & security testing

3.4. MANAGE

- Deployment guardrails
- Monitoring (drift, misuse, performance)
- Incident response
- Audit cycles & retraining

4. Responsible AI Controls Catalogue

4.1. Data Controls

- Data minimisation
- Provenance validation
- Bias detection
- Consent governance

4.2. Model Controls

- Explainability thresholds
- Fairness metrics
- Human-in-the-loop
- Robustness validation

4.3. Security Controls

- MITRE ATLAS threat mapping
- Adversarial testing
- Access management
- Logging & anomaly detection

4.4. Deployment Controls

- Deployment gates
- Rollback strategy
- Versioning
- Reproducibility

4.5. Monitoring Controls

- Drift alerts
- Fairness monitoring
- Incident logging
- SLA tracking

4.6. Organizational & Governance Controls

- RACI for AI roles
- Audit readiness
- Compliance mapping
- Governance decisions

5. AI RISK SCORING MODEL

Risk Score = Impact * Likelihood * Detectability (Produces a Red-Amber-Green Risk Matrix)

5.1. Impact Dimensions

- Harm to individuals
- Financial risk
- Legal/regulatory exposure
- Security impact
- Reputational harm

5.2. Likelihood Dimensions

- Bias probability
- Data vulnerability
- Model instability
- Attack feasibility

5.3. Detectability

- Ease of identifying anomalies
- Monitoring strength

6. Documentation Templates