Responsible AI Policy

1. Purpose & Scope

This Responsible AI Policy establishes the principles, controls, and governance mechanisms required to ensure that all Artificial Intelligence systems are developed, deployed, and monitored in a safe, ethical, transparent, secure, and compliant manner.

This policy applies to:

- Machine learning models
- Large language models (LLMs)
- Generative AI systems
- AI-powered decision support systems
- Third-party or vendor AI systems
- Internal and external AI use cases

2. Definitions

AI System: Any system that performs tasks requiring perception, prediction, reasoning, or decision-making using data-driven or algorithmic approaches.

High Risk AI: AI systems whose failure may cause significant harm, including:

- Legal violation
- Bias or unfair outcomes
- Privacy breach
- Safety or security risk
- Societal or ethical harm

Responsible AI (RAI): Practices ensuring safe, ethical, compliant, transparent, and human centered AI.

AI Lifecycle:

From Ideation to retirement
Govern - Map - Measure - Manage

Human Oversight:

Human in the loop (HITL), human on the loop (HOTL), and human in command (HIC).

3. Responsible AI Principles

3.1 Human-Centered Design: AI systems must be designed to augment human capability, respect human dignity, and provide clear user control.

3.2 Ethical Data Use: Data must be collected, used, and retained with fairness, minimality, consent, and transparency.

3.3 Transparency & Traceability: AI systems must be explainable, documented, and auditable across the lifecycle.

3.4 Safety & Security by Design: Security, robustness, and adversarial resistance must be embedded into model architecture, testing, deployment, and monitoring.

3.5 Responsible Deployment: AI must undergo approvals, controls validation, and operational readiness checks before production.

3.6 Continuous Monitoring & Auditability: Models must be monitored for performance, drift, bias, misuse, and anomalies with structured audit cycles.

3.7 Accountability & Governance Alignment: Clear roles, decision rights, escalation paths, and oversight structures are mandatory.

4. AI Governance Structure

4.1 Strategic Governance Layer

Responsibilities:
Set AI principles, approve policies, define risk appetite, oversee high-risk models.

- AI Governance Board

- RAI Executive Sponsor
- Chief Information Security Officer (CISO)
- Legal & Compliance
- Ethics Review Council

4.2 Tactical Governance Layer

Responsibilities: Assess risks, review documentation, map MITRE ATLAS threats, ensure controls, prepare audit evidence**.**

- AI Program Manager
- AI Risk Manager
- Data Stewards
- Model Owners
- Security/Privacy Leads

4.3 Operational Layer

Responsibilities: Model development, testing, deployment readiness, monitoring, incident reporting.

- ML Engineers
- Data Scientists
- MLOps
- Testing & QA
- Monitoring Teams

5. AI Risk Assessment Requirements

Every AI system must be classified using:

5.1 Risk Scoring Model:
Impact × Likelihood × Detectability

5.2 RAG Classification:
Red - High Risk (Board Approval Required)
Amber - Moderate Risk (Mitigate Before Deployment)
Green - Low Risk (Standard Deployment)

5.3 High-risk AI must complete:

- Full RAI assessment
- MITRE ATLAS threat assessment
- Privacy Impact Assessment
- Model Card + Data Card
- Explainability report
- Security testing (adversarial)

6. Mandatory AI Controls

6.1 Data Controls

- Data minimization
- Provenance verification
- Bias checks
- Sensitive data protection

6.2 Model Controls

- Explainability thresholds
- Fairness metrics
- Robustness & stress testing
- HITL requirements

6.3 Security Controls (MITRE ATLAS integrated)

- Adversarial input testing
- Data poisoning prevention
- Prompt injection safeguards
- Model inversion & membership inference protections

6.4 Deployment Controls

- Approval gates
- Access governance
- Versioning
- Rollback plans

6.5 Monitoring Controls

- Drift detection
- Performance tracking
- Hallucination detection
- Anomaly alerts

6.6 Organizational Controls

- RACI for AI roles
- Ethical review
- Regulatory mapping
- Annual policy review

7. Transparency & Explainability Requirements

All AI systems must:

- Provide meaningful explanations
- Declare limitations
- Document model assumptions
- Provide confidence levels where applicable

8. Human Oversight Requirements

HITL is required when:

- Decisions affect individuals or groups
- High-risk models operate
- Safety critical tasks are involved

Note: All oversight activity must be logged

9. Privacy & Data Protection Requirements

AI systems must follow:

- PDPA / GDPR

- Data minimization
- Purpose limitation
- Consent tracking
- Encryption & anonymization

10. Monitoring & Audit Requirements

10.1 Monitoring includes:

- Drift
- Fairness
- Performance
- Misuse
- Input attack attempts
- Output anomalies

10.2 Audit includes:

- Controls effectiveness
- Model transparency
- Documentation completeness
- ATLAS threat review

11. Incident Response Requirements

Teams must report incidents within 24 hours for:

- Hallucination causing harm
- Data leakage
- Prompt injection compromise
- Bias escalation
- Security breach
- Regulatory non-compliance

12. Compliance Requirements

Models must map to:

- ISO/IEC 42001
- NIST AI RMF
- OECD Principles
- EU AI Act
- Local jurisdiction laws

13. Policy Exceptions

Exceptions require approval by:

- AI Governance Board
- RAI Sponsor
- CISO
- Legal

14. Policy Review Cycle

This policy must be reviewed every 12 months or earlier if:

- New regulations emerge
- Technology shifts
- New risks arise