

Field	Value
Document ID	GOV-TVR-001
Version	1.0
Effective Date	January 2024
Document Owner	Chief AI Officer
Classification	Internal Use
Review Cycle	Annual

Table of Contents

Testing, Validation & Red Teaming Standard.....1

1. Purpose..... 1

2. Scope..... 1

 2.1 System Coverage..... 1

 2.2 Testing Phases..... 1

 2.3 Segment Analysis Requirements..... 2

3. Regulatory and Framework Alignment..... 2

 3.1 Regulatory Requirements..... 2

 3.2 Industry Framework Alignment..... 2

4. Testing Requirements by Risk Tier..... 3

 4.1 Critical Risk Tier Requirements..... 3

 4.2 High Risk Tier Requirements..... 3

 4.3 Medium Risk Tier Requirements..... 4

 4.4 Low Risk Tier Requirements..... 4

5. Fairness and Bias Testing Requirements..... 4

 5.1 Demographic Cohort Analysis..... 4

 5.2 Bias Detection Methodologies..... 5

 5.3 Bias Mitigation Validation..... 6

6. Adversarial Testing and Red Team Requirements..... 6

 6.1 Abuse and Prompt Injection Testing..... 6

 6.2 Red Team Engagement Requirements..... 7

 6.3 Red Team Methodology Requirements..... 7

7. Implementation Requirements..... 8

 7.1 Test Suite Development Requirements..... 8

 7.2 Testing Environment Specifications..... 9

 7.3 Automated Testing Pipeline Integration..... 10

8. Acceptance Criteria and Approval Requirements..... 10

 8.1 Quantitative Acceptance Thresholds..... 10

 8.2 Qualitative Assessment Requirements..... 11

 8.3 Documentation and Evidence Requirements..... 12

9. Organizational Roles and Responsibilities..... 13

 9.1 Comprehensive RACI Matrix..... 13

9.2 Detailed Role Specifications.....	14
9.3 Training and Competency Development.....	16
10. Required Templates and Artifacts.....	17
10.1 Validation Plan Template Structure.....	17
10.2 Bias Test Specification Template.....	18
10.3 Red Team Scenarios Documentation.....	19
10.4 Test Results Documentation Standards.....	20
11. Monitoring and Continuous Improvement Framework.....	20
11.1 Testing Program Effectiveness Metrics.....	20
11.2 Continuous Process Improvement Requirements.....	21
11.3 Organizational Learning and Knowledge Management.....	22
12. Compliance Monitoring and Audit Framework.....	23
12.1 Internal Audit Program Structure.....	23
12.2 External Audit and Regulatory Examination Coordination.....	23
12.3 Non-Compliance Response and Corrective Action.....	24
12.4 Regulatory Reporting and Communication.....	25

Testing, Validation & Red Teaming Standard

1. Purpose

This standard establishes comprehensive requirements for testing, validating, and conducting adversarial assessments of artificial intelligence systems to ensure performance, robustness, bias mitigation, and safety throughout the AI system lifecycle. The standard mandates systematic approaches to identify vulnerabilities, assess fairness across demographic cohorts, and validate system behavior under both normal and adversarial conditions.

The standard serves to:

- Ensure AI systems meet defined performance benchmarks before deployment
- Identify and mitigate biases that could result in discriminatory outcomes
- Validate system robustness against adversarial attacks and edge cases
- Establish consistent testing methodologies across all AI system development
- Provide evidence of due diligence for regulatory compliance and risk management

2. Scope

This standard applies to all AI systems within the organization's portfolio, encompassing:

2.1 System Coverage

- Machine learning models in development, staging, and production environments
- Natural language processing systems including conversational AI and content generation
- Computer vision systems for image recognition, classification, and analysis