



AI Risk Management Framework GOVERN Function

Procedural Manual & Implementation Guide

AI RISK MANAGEMENT FRAMEWORK

GOVERN Function

Procedural Manual & Implementation Guide

AI RMF 2026 – ISO Standards Integrated Edition

Integration with ISO/IEC 42001 | ISO/IEC 27001 | ISO/IEC 23894

Additional Source Information

Singapore Model AI Governance Framework for Agentic AI

World Economic Forum- AI Agents in Action: Foundations for Evaluation and Governance

Agentic AI Risk-Management Standards Profile (UC Berkeley CLTC)

New Updates: Version 1.5 - COMPLETED EDITION

February 2026

Document Control

Field	Details
Document Title	Bluefox Consulting Services, LLC. U.S. Virginia AI RMF 2026 GOVERN Procedural Manual
Version	5 - COMPLETED EDITION
Date	February 2026
Owner	Chief AI Officer
Approval	AI Governance Committee
Classification	Confidential - Internal Use Only
Review Cycle	Annual
Status	Completed - Includes Agentic AI, Environmental Sustainability, Third-Party Governance, Risk Appetite, Registry, and End-User Redress

Version History

Version	Date	Author	Changes
1.0	Feb 2026	AI Governance Team	Initial release - INCOMPLETE (missing agentic AI and environmental sustainability governance)
1.1	Feb 2026	AI Governance Team	CORRECTED - Added GOVERN 1.5 (Agentic AI), GOVERN 1.6 (Environmental Sustainability), ISO 23894 mappings, Singapore MGF integration, new templates
1.2	Feb 2026	AI Governance Team	COMPLETED - Added GOVERN 1.7, 1.8, 5.2, and 5.3; Integrated AI Impact Assessment; and Appendices L-P.
1.3	Feb 7	AI Governance Team	<ul style="list-style-type: none"> Appendix T: AI Governance KPI Dashboard Appendix U: AI Governance Case Studies Appendix V: Agent Identity and Credential Management Protocol
1.4	Feb 10	AI Governance Team	<ul style="list-style-type: none"> 1: Added Singapore Model AI Governance Framework for Agentic AI overview, key practices, and critical gap analysis (environmental sustainability omission) GOVERN 1.1: Enhanced with multi-agent ecosystem regulatory requirements - cross-jurisdictional compliance, multi-party accountability, governor/auditor agent regulatory status, and data governance across agent boundaries GOVERN 5.1: New subsection on Governor and Auditor Agent Patterns including implementation requirements and oversight mechanisms per Singapore Framework guidance Multi-Agent Ecosystems Section: Updated introductory content to explicitly reference Singapore Framework alongside Singapore MGF and other international guidance

			<ul style="list-style-type: none"> • Appendix X (NEW): Comprehensive Singapore Framework Crosswalk mapping AI RMF 2026 functions to Singapore Framework governance practices with detailed implementation guidance and gap analysis • RACI Matrices: Added multi-agent ecosystem regulatory mapping activities with Singapore Framework-specific role assignments • Control Mappings: Added Singapore Framework column to all control mapping tables • Success Metrics: New Singapore Framework-specific metrics including agent regulatory classification tracking, cross-jurisdictional compliance mapping, and governor/auditor intervention rates • Bibliography: Added Singapore Model AI Governance Framework citation and reference materials
--	--	--	---

Disclaimer Notice:

- 1) This AI ("Framework and/or Procedural Manual and Implementation Guide") is provided for "FREE" for use as **general informational purposes only**. It does not constitute professional, legal, or technical advice, nor does it create a client-consultant relationship. You should consult with a qualified professional before making any decisions based on this Framework.
- 2) **As-Is" Basis:** This ("Framework and/or Procedural Manual and Implementation Guide") is provided "as-is." The author makes no representations or warranties, express or implied, about the accuracy, completeness, or reliability of the content. Any reliance you place on such information is strictly at your own risk—you are required to implement your own level of due diligence with using it..
- 3) **Limitation of Liability:** In no event will the author be liable for any loss or damage, including without limitation, indirect or consequential loss or damage, arising from the use of this framework.
- 4) **User Responsibility:** Users are responsible for their own due diligence and for tailoring this framework to their specific business needs and legal jurisdictions.

What “Implementation-Ready” Actually Means

The gap between governance principles and governance practice has always been the graveyard of well-intentioned frameworks. Organizations adopt a standard, write a policy, and then discover that the distance between “define roles and responsibilities” and actually operationalizing those roles is measured in months and budget lines nobody anticipated.

The AI RMF 2026 GOVERN Manual was built to eliminate that gap. Every GOVERN subcategory includes step-by-step implementation procedures, RACI matrices identifying who is responsible, accountable, consulted, and informed for each activity, control mappings across ISO 42001, ISO 27001, ISO 23894, Singapore MGF, WEF Framework, and EU AI Act, specific deliverables with corresponding templates in the appendices, and measurable success metrics.

The appendices alone include an AI Compliance Register template, Decision Authority Matrix, complete RACI matrix templates, governance committee charters, agent identity management frameworks, agent risk assessment templates, carbon footprint assessment templates, third-party governance toolkits, AI system and agent registry templates, risk appetite and acceptance templates, integrated AI impact assessment templates, end-user redress toolkits, framework compliance matrices, phased implementation roadmaps (from first 90 days through 18-month optimization), industry-specific adaptations for healthcare, financial services, government, manufacturing, and retail, a full KPI dashboard with leading and lagging indicators, real-world case studies, and comprehensive crosswalks to every major international framework.

This is not a document you read and then figure out how to implement. It’s a document you implement by following its procedures, using its templates, and measuring against its metrics.

The GOVERN Function Is the Foundation. Everything Else Depends on It.

The MAP function can’t profile agents that don’t exist in a registry. The MEASURE function can’t track oversight effectiveness without defined oversight patterns. The MANAGE function can’t respond to incidents without established escalation paths and decision authority.

Everything flows from GOVERN. If the governance architecture is structurally sound—with separated authority, documented agent identities, proportionate oversight patterns, anti-automation-bias mechanisms, and environmental sustainability as a core function rather than an afterthought—then the rest of the framework has a foundation that holds under pressure.

If GOVERN is absent, incomplete, or designed for a generation of AI systems that no longer reflects what organizations are actually deploying, then every downstream function is building on sand.

The question for 2026 isn't whether your organization has an AI governance policy. It's whether that policy knows your agents exist, knows what they're authorized to do, can verify that oversight is actually functioning, accounts for the environmental cost of scaling them, and can demonstrate all of the above to a regulator who asks.

The AI RMF 2026 GOVERN Implementation Manual was built to answer every one of those questions. Not with principles. With procedures, templates, metrics, and the structural architecture that makes governance operational.

Contents

Document Control	2
Version History	2
What “Implementation-Ready” Actually Means	4
The GOVERN Function Is the Foundation. Everything Else Depends on It.	4
Enhancements in Version 1.2.....	18
1. Introduction	19
1.1 Purpose and Scope	19
1.2 How to Use This Manual.....	19
1.3 Framework Integration.....	20
1.4 What's New in AI RMF 2026.....	21
Agentic AI Governance	21
Environmental Sustainability Governance.....	22
ISO 23894 Integration	22
**Singapore Model AI Governance Framework for Agentic AI	23
World Economic Forum AI Agents in Action: Foundations for Evaluation and Governance	23
2. GOVERN Function Overview	24
2.1 GOVERN Function Structure.....	24
2.2 Key Governance Bodies	25
3. GOVERN 1: Organizational AI Governance	26
3.1 GOVERN 1.1: Legal and Regulatory Requirements	26
Framework Integration.....	26
Multi-Agent Ecosystem Regulatory Requirements (Singapore).....	27
Key Regulatory Frameworks for AI Systems	27
Implementation Procedures.....	28
Procedure 1.1.1: Establish Compliance Register	28
Procedure 1.1.2: Monitor Regulatory Developments.....	28
Procedure 1.1.3: Conduct Regulatory Gap Analysis	28
Procedure 1.1.4: Conduct Compliance Assessments	28
Legal and Regulatory Requirements	29
Roles & Responsibilities (RACI Matrix)	29
Control Mappings.....	29
Key Deliverables.....	29

Success Metrics..... 30

1.3 SINGAPORE-Specific Metrics - Detailed Specifications..... 30

Implementation Notes..... 31

3.2 GOVERN 1.2: Roles and Responsibilities 32

 Key Organizational Roles..... 32

 Implementation Procedures 32

 Roles & Responsibilities (RACI Matrix) 33

 Control Mappings..... 33

 Key Deliverables 33

3.3 GOVERN 1.3: Diversity, Equity, Inclusion, and Accessibility (DEIA) 34

 Implementation Procedures 34

 Roles & Responsibilities (RACI Matrix) 35

 Control Mappings..... 36

 Key Deliverables 36

3.4 GOVERN 1.4: Organizational AI Risk Management Culture 37

 Implementation Procedures 37

 Roles & Responsibilities (RACI Matrix) 37

 Control Mappings..... 38

 Key Deliverables 38

3.5 GOVERN 1.5: Agentic AI Governance 39

 Context: Why Agentic AI Requires Specialized Governance 39

 Implementation Procedures 40

Agent Profile Card and Tier Linkage..... 41

 Roles & Responsibilities (RACI Matrix) 44

 Control Mappings..... 45

 Key Deliverables 45

 Integration with Singapore MGF for Agentic AI 46

3.6 GOVERN 1.6: Environmental Sustainability Governance..... 47

 Context: Why Environmental Sustainability Requires Governance 47

 Implementation Procedures 48

 Roles & Responsibilities (RACI Matrix) 52

 Control Mappings..... 53

 Key Deliverables 53

 Integration with ISO 23894 Environmental Extensions 54

- 3.7 GOVERN 1.7: Third-Party & Tool Ecosystem Governance 55
 - Implementation Procedures 55
 - Roles & Responsibilities (RACI Matrix) 57
 - Control Mappings 57
 - Key Deliverables 58
- 3.8 GOVERN 1.8: AI Risk Appetite, Risk Criteria, and Risk Acceptance 59
 - Implementation Procedures 59
 - Roles & Responsibilities (RACI Matrix) 61
 - Control Mappings 61
 - Key Deliverables 61
- 4. GOVERN 2: Accountability and Responsibility 62
 - 4.1 GOVERN 2.1: Accountability Structures 62
 - Implementation Procedures 62
 - Roles & Responsibilities (RACI Matrix) 64
 - Control Mappings 65
 - Key Deliverables 65
- 5. GOVERN 3: Workforce Diversity and Team Composition 66
 - 5.1 GOVERN 3.1: Diverse and Multidisciplinary AI Teams 66
 - Team Composition Requirements 66
 - Implementation Procedures 67
 - Roles & Responsibilities (RACI Matrix) 68
 - Control Mappings 68
 - Key Deliverables 68
- 6. GOVERN 4: Organizational Culture and AI Risk Management 69
 - 6.1 GOVERN 4.1: Training and Awareness Programs 69
 - Training Program Framework 69
 - Implementation Procedures 70
 - Roles & Responsibilities (RACI Matrix) 71
 - Control Mappings 71
 - Key Deliverables 71
- 7. GOVERN 5: Oversight and Monitoring 72
 - 7.1 GOVERN 5.1: Governance Bodies and Oversight Mechanisms 72
 - Key Oversight Mechanisms 72
 - Human Oversight Patterns for Agentic and Multi-Agent Systems 73

GP AIS and Open-Weights Decommissioning Constraints.....	73
Implementation Procedures	74
Roles & Responsibilities (RACI Matrix)	76
Control Mappings	76
Key Deliverables	76
7.2 GOVERN 5.4: AI System & Agent Registry and Portfolio Monitoring	77
Implementation Procedures	77
Roles & Responsibilities (RACI Matrix)	78
Control Mappings	78
Key Deliverables	78
7.3 GOVERN 5.5: End-User Enablement, Feedback, Complaints, and Redress	79
Implementation Procedures	79
Roles & Responsibilities (RACI Matrix)	80
Control Mappings	81
Key Deliverables	81
8. GOVERN 6: Trustworthy AI Characteristics	82
8.1 GOVERN 6.1: Trustworthy AI Policy.....	82
Seven Trustworthiness Characteristics	82
Implementation Procedures	83
Roles & Responsibilities (RACI Matrix)	86
Control Mappings	87
Key Deliverables	87
Appendices	88
Appendix A: AI Compliance Register Template	89
Appendix B: RACI Matrix Template - Complete GOVERN Function.....	90
Appendix C: Decision Authority Matrix.....	91
Appendix D: AI Risk Management Training Curriculum.....	92
Appendix E: Governance Committee Charters and Policy Templates.....	93
E.1: AI Governance Committee Charter Template.....	93
E.2: AI Ethics Review Board Charter Template	93
E.3: Trustworthy AI Policy Template Outline.....	94
E.4: Complete Control Mapping Summary.....	95
Appendices F: Agentic AI and Environmental Sustainability	96
Appendix F: Agentic AI Committee Charter Template	97

- 1. Committee Name and Authority 97
- 2. Scope of Oversight..... 97
- 3. Committee Membership 98
- 4. Meeting Frequency and Procedures 98
- 5. Key Responsibilities 99
- 6. Decision Criteria for Agentic AI Approval 99
- Appendix G: Environmental Sustainability Committee Charter Template 100
 - 1. Committee Name and Authority 100
 - 2. Scope of Oversight..... 100
 - 3. Committee Membership 100
 - 4. Meeting Frequency and Procedures 101
 - 5. Key Responsibilities 101
 - 6. Sustainability Approval Criteria 101
- Appendix H: Agent Identity Management Framework Template..... 102
 - 1. Agent Identity Principles..... 102
 - 2. Agent Identity Schema 102
 - 3. Agent Authentication Methods 103
 - 4. Agent Authorization Framework 103
 - 5. Agent Identity Lifecycle 103
 - 6. Agent Identity Monitoring 104
- Appendix I: Agent Risk Assessment Template 105
- Appendix J: Carbon Footprint Assessment Template..... 106
- Appendix K: Sustainability Metrics Tracking Template 107
- Appendix L: Third-Party & Tool Ecosystem Governance Toolkit 93
- Appendix M: AI System & Agent Registry Template..... 97
- Appendix N: AI Risk Appetite and Risk Acceptance Templates 98
- Appendix O: Integrated AI Impact Assessment (AIIA) Template 102
- Appendix P: End-User Enablement, Complaint Intake, and Redress Toolkit..... 105
- Appendix Q: Framework Compliance Matrix 109
 - Q.1: ISO/IEC 42001:2023 Compliance Matrix 110
 - Q.2: EU AI Act Compliance Matrix..... 111
 - Q.3: ISO 23894 Risk Management Extensions 112
 - Q.4: Singapore MGF Four Dimensions Detailed Mapping..... 113
- Appendix R: Implementation Roadmap..... 114

R.1: Quick Start Guide (First 90 Days) 114

R.2: Phased Implementation Roadmap 115

 Phase 1: Foundation (Months 1-3)..... 115

 Phase 2: Expansion (Months 4-6)..... 116

 Phase 3: Maturity (Months 7-12)..... 117

 Phase 4: Optimization (Months 13-18)..... 118

R.3: AI Governance Maturity Model..... 119

R.4: Resource Planning Guide 120

 R.4.1: Staffing Requirements by GOVERN Subcategory 120

 R.4.2: Technology and Tools Budget 121

 R.4.3: Training Budget 122

R.5: Common Implementation Challenges and Solutions..... 123

R.6: Implementation Success Metrics..... 124

APPENDIX S: INDUSTRY-SPECIFIC ADAPTATIONS..... 125

S.1 PURPOSE AND STRUCTURE 125

 Purpose of This Appendix..... 125

 How to Use This Appendix 125

 Structure Overview 126

S.2 HEALTHCARE AND LIFE SCIENCES 127

 S.2.1 Sector Overview 127

 Key AI Use Cases 127

 Unique Governance Challenges 127

 S.2.2 Sector-Specific Regulations 128

 S.2.3 High-Risk AI Use Cases 129

 S.2.4 GOVERN Procedure Adaptations 130

 GOVERN 1.1.1: Legal and Regulatory Compliance Register 130

 GOVERN 2.1.2: AI System Approval Requirements 130

 GOVERN 5.1: AI Oversight Mechanisms..... 131

 S.2.5 Template Modifications 132

 S.2.6 Implementation Priorities: First 90 Days..... 133

 S.2.7 Healthcare Case Study: Clinical AI Deployment..... 134

S.3 FINANCIAL SERVICES 135

 S.3.1 Sector Overview 135

 Key AI Use Cases 135

Unique Governance Challenges 135

S.3.2 Sector-Specific Regulations 136

S.3.3 High-Risk AI Use Cases 137

S.3.4 GOVERN Procedure Adaptations 138

 GOVERN 1.3: Diversity, Equity, Inclusion, and Accessibility 138

 GOVERN 5.1.2: Independent Review and Validation 138

 GOVERN 6.1.6: Explainability and Interpretability 139

S.3.5 Template Modifications 140

S.3.6 Implementation Priorities: First 90 Days 141

S.3.7 Financial Services Case Study: Fair Lending Analysis 142

S.4 GOVERNMENT AND PUBLIC SECTOR 143

 S.4.1 Sector Overview 143

 S.4.2 Sector-Specific Regulations 144

 S.4.3 High-Risk Use Cases 145

 S.4.4 Key Procedure Adaptations 145

 S.4.5 Government Template Modifications 146

 S.4.6 Implementation Priorities: First 90 Days 146

S.5 MANUFACTURING AND INDUSTRIAL 147

 S.5.1 Sector Overview 147

 S.5.2 Sector-Specific Regulations 147

 S.5.3 High-Risk Use Cases 148

 S.5.4 Key Procedure Adaptations 148

 S.5.5 Template Modifications 149

 S.5.6 Implementation Priorities: First 90 Days 149

S.6 RETAIL AND E-COMMERCE 150

 S.6.1 Sector Overview 150

 S.6.2 Sector-Specific Regulations 150

 S.6.3 High-Risk Use Cases 151

 S.6.4 Key Procedure Adaptations 151

 S.6.5 Template Modifications 151

 S.6.6 Implementation Priorities 151

 S.6.7 Retail Case Study: Privacy-Preserving Personalization 152

CONCLUSION 153

T.1 Purpose and Structure 154

Structure Overview 154

T.2 KPI Framework Overview 154

 T.2.1 Leading vs. Lagging Indicators 154

 T.2.2 Measurement Frequency 154

 T.2.3 Data Collection and Governance 155

 Data Sources and Systems 155

 Data Quality Standards 155

 Roles and Responsibilities 155

T.3 GOVERN Function KPIs by Subcategory 156

 T.3.1 GOVERN 1.1: Legal/Regulatory Compliance KPIs 156

 T.3.2 GOVERN 1.2: Roles and Responsibilities KPIs 157

 T.3.3 GOVERN 1.3: DEIA KPIs 158

 T.3.4 GOVERN 1.4: Risk Culture KPIs 159

 T.3.5 GOVERN 1.5: Agentic AI Governance KPIs 160

 T.3.6 GOVERN 1.6: Environmental Sustainability KPIs 161

 T.3.7 GOVERN 2.1: Accountability and Transparency KPIs 162

 T.3.8 GOVERN 3.1: Diverse and Skilled Teams KPIs 163

 T.3.9 GOVERN 4.1: Awareness and Training KPIs 164

 T.3.10 GOVERN 5.1: Oversight Functions KPIs 165

 T.3.11 GOVERN 6.1: Trustworthy AI Characteristics KPIs 166

T.4 Executive Dashboard Template 167

 T.4.1 Dashboard Components 167

 T.4.2 Status Indicators 167

 T.4.3 Sample Dashboard Metrics 168

 T.4.4 Dashboard Refresh and Distribution 168

T.5 Board Reporting Template 169

 T.5.1 Quarterly Report Structure 169

 T.5.2 Sample Board Report Excerpts 169

 Executive Summary Example 169

 Risk and Incident Summary Example 169

T.6 Governance Maturity Scoring 170

 T.6.1 Maturity Levels 170

 T.6.2 Scoring Methodology 170

T.7 Benchmark Ranges and Industry Standards 171

T.7.1 Maturity Benchmarks by Organization Type 171

T.7.2 Functional Area Benchmarks..... 171

T.7.3 Industry-Specific Adjustments 171

Conclusion and Implementation Guidance 172

 Phased Implementation Approach..... 172

Appendix U: AI Governance Case Studies..... 173

 U.1 Purpose and Structure 173

 U.2 Case Study 1: High-Risk AI System Approval Process 174

 Section 1: Case Overview..... 174

 Section 2: Background and Context..... 174

 Section 3: Governance Process Applied 175

 Section 4: Key Documents and Artifacts 176

 Section 5: Outcome and Results..... 177

 Section 6: Lessons Learned..... 177

 Section 7: Takeaway for Practitioners..... 178

 U.3 Case Study 2: Agentic AI Deployment and Oversight 179

 Section 1: Case Overview..... 179

 Section 2: Background and Context..... 179

 Section 3: Governance Process Applied 180

 Section 4: Key Documents and Artifacts 180

 Section 5: Outcome and Results..... 182

 Section 6: Lessons Learned..... 182

 Section 7: Takeaway for Practitioners..... 183

 U.4 Case Study 3: Environmental Impact Assessment and Optimization 184

 Section 1: Case Overview..... 184

 Section 2: Background and Context..... 184

 Section 3: Governance Process Applied 185

 Section 4: Key Documents and Artifacts 186

 Section 5: Outcome and Results..... 187

 Section 6: Lessons Learned..... 187

 Section 7: Takeaway for Practitioners..... 188

 U.5 Case Study 4: AI Ethics Review Board Decision..... 189

 Section 1: Case Overview 189

 Section 2: Background and Context..... 189

Section 3: Governance Process Applied 190

Section 4: Key Documents and Artifacts 191

Section 5: Outcome and Results 192

Section 6: Lessons Learned 192

Section 7: Takeaway for Practitioners 193

U.6 Case Study 5: Third-Party AI Supplier Risk Assessment 194

 Case Summary 194

U.7 Case Study 6: AI Incident Investigation and Response 195

 Case Summary 195

U.8 Case Study 7: Bias Detection and Mitigation in Credit Scoring 196

 Case Summary 196

U.9 Case Study 8: ISO 42001 Certification Journey 197

 Case Summary 197

Appendix U Summary 199

Appendix V: Agent Identity and Credential Management Protocol 200

 V.1 Overview and Purpose 200

 V.2 Agent Identity Core Principles 200

 V.3 Agent Identity Registry Implementation 202

 V.4 Credential Management Framework 203

 V.5 Continuous Monitoring and Audit Requirements 204

 V.6 Implementation Checklist 205

 V.7 Related Resources and Cross-References 206

APPENDIX X: AI RMF 2026 CROSSWALK TO SINGAPORE MODEL AI GOVERNANCE FRAMEWORK FOR AGENTIC AI 207

 Overview 207

 Singapore Framework Background 207

 Framework Comparison: AI RMF 2026 vs SINGAPORE 208

 Key Integration Points 208

 Singapore Framework Core Governance Practices 209

 1. Governance Boundaries Definition 209

 2. System-Level Oversight Patterns 209

 3. Multi-Agent Risk Assessment 210

 4. Cross-Organizational Participation Rules 210

 5. Multi-Agent Incident Management 210

 Singapore Framework Gap: Environmental Sustainability 210

- Universal Gap Across Frameworks..... 211
- Implementation Recommendation 211
- Conclusion..... 212
- Appendix Y: AI RMF 2026 – MULTI-AGENT ECOSYSTEMS..... 213
 - 1. Overview..... 213
 - 2. Characteristics and Example Use Cases..... 213
 - 2.1 Key Characteristics of Multi-Agent Ecosystems..... 213
 - 2.2 Example Use Cases..... 214
 - 3. System-Level Profiling for Multi-Agent Ecosystems..... 215
 - 3.1 System-Level Agentic Profile Template 215
 - 4. Applying GOVERN-MAP-MEASURE-MANAGE to Multi-Agent Ecosystems..... 216
 - GOVERN – System-Level Oversight and Accountability..... 216
 - MAP – System-Level Context and Risk Assessment..... 216
 - MEASURE – System-Level Monitoring Metrics..... 217
 - MANAGE – Multi-Agent Incident Response..... 217
 - 5. Framework Integration Summary..... 218
 - 6. Critical Finding: Environmental Sustainability Gap 218
 - Conclusion..... 218
- Appendix Z: Your AI Agents Don’t Have Identities. 220
 - Your Governance Framework Doesn’t Know They Exist. 220
 - The Question Nobody Is Asking About Their AI Agents 220
- The Agent Identity Problem: Autonomous Actors Without Credentials..... 220
- The Automation Bias Trap: When Oversight Becomes Rubber-Stamping 221
- Why Operational Teams Cannot Certify Their Own Governance 222
- Three Oversight Patterns for Agentic AI—and the Criteria That Determine Which One Applies 223
- Governor and Auditor Agents: When AI Oversees AI 224
- The Gap Every Framework Shares—Except One..... 224
- From Binary Classification to Risk-Proportionate Governance..... 225
- Glossary..... 227
- References..... 233
 - Primary Framework 233
 - International Standards..... 233
 - Regulatory Frameworks..... 234
 - Governance Frameworks 234

Technical Guidance 234
Research and Best Practices 235
U.S. Sector-Specific Privacy and Data Protection Regulations 236
Additional Resources..... 236

Enhancements in Version 1.2

This version includes governance-completeness enhancements and operational templates:

- **Added GOVERN 1.7:** Third-Party & Tool Ecosystem Governance - supplier due diligence, contracting, monitoring, and exit controls
- **Added GOVERN 1.8:** AI Risk Appetite, Risk Criteria, and Risk Acceptance - residual risk acceptance and policy exception governance
- **Added GOVERN 5.2:** AI System & Agent Registry and Portfolio Monitoring - portfolio-level traceability and oversight reporting
- **Added GOVERN 5.3:** End-User Enablement, Feedback, Complaints, and Redress - user disclosure, intake, triage, and remediation workflow
- Added new appendices (L-P) with templates and embedded guidance notes to operationalize the above requirements

1. Introduction

1.1 Purpose and Scope

This Procedural Manual provides comprehensive, actionable guidance for implementing the GOVERN function of the AI Risk Management Framework (AI RMF 2026). It is designed to help organizations establish, maintain, and continually improve their AI governance structures, policies, and risk management culture, with particular emphasis on agentic AI systems and environmental sustainability.

Key Objectives:

- Establish organizational structures for AI governance and oversight, including specialized governance for agentic AI systems
- Define clear roles, responsibilities, and accountabilities across the AI lifecycle
- Develop and implement comprehensive AI risk management policies addressing autonomous AI and environmental impact
- Foster a culture of responsible AI development and deployment
- Ensure compliance with regulatory requirements and industry standards (ISO 42001, ISO 27001, ISO 23894, Singapore MGF, EU AI Act)
- Provide practical procedures, templates, and tools for implementation

1.2 How to Use This Manual

This manual is structured to support both sequential implementation and targeted reference use:

- For new AI governance programs: Follow sections sequentially from GOVERN 1 through GOVERN 6
- For existing programs: Use the control mappings to identify gaps and enhancements
- For agentic AI deployments: Pay special attention to GOVERN 1.5 (Agentic AI Governance)
- For environmental sustainability: Focus on GOVERN 1.6 (Environmental Sustainability Governance)
- For specific needs: Reference individual procedures and templates as needed
- For compliance: Use the ISO and Singapore MGF mappings to demonstrate conformance

1.3 Framework Integration

This manual integrates requirements and best practices from multiple authoritative sources:

Framework/Standard	Integration Approach
NIST AI RMF 1.0 (2023)	Core structure and trustworthiness characteristics
ISO/IEC 42001:2023	Management system requirements and certifiable controls
ISO/IEC 27001:2022	Information security controls applicable to AI systems
ISO/IEC 23894:2023	AI risk management process, techniques, and extensions for agentic AI and sustainability
Singapore Model AI Governance Framework for Agentic AI (2026)	Four dimensions of agentic AI governance: assess/bound risks, accountability, technical controls, end-user responsibility
EU AI Act (2024)	Regulatory compliance requirements for high-risk AI systems
World Economic Forum (WEF) AI Agents in Action: Foundations for Evaluation and Governance	Key features, integration points, and critical gap analysis (environmental sustainability)

1.4 What's New in AI RMF 2026

The AI RMF 2026 Integrated Edition extends the original NIST AI RMF 1.0 (2023) with critical additions:

Agentic AI Governance

Agentic AI systems are AI systems that can plan across multiple steps to achieve specified objectives using AI agents. These systems possess independent planning and action-taking capabilities (e.g., searching the web, creating files, executing transactions) over multiple steps to achieve user-defined goals.

Key Characteristics of Agentic AI:

- **Action-Space:** Range of actions the agent is permitted to take, determined by tools, systems access, and transaction authority
- **Autonomy:** Degree to which an agent can decide when and how to act toward a goal
- **Multi-Agent Systems:** Multiple agents working together in sequential, supervisor, or swarm patterns
- **Dynamic Behavior:** Agents adapt to new information and interact with other agents and systems
- **Real-World Impact:** Agents take actions that directly affect databases, systems, and external services

Agentic AI Risk Categories:

- **Erroneous Actions:** Incorrect actions such as scheduling on wrong dates or producing flawed code
- **Unauthorized Actions:** Actions taken outside permitted scope or without required human approval
- **Biased or Unfair Actions:** Actions leading to unfair outcomes across different groups
- **Data Breaches:** Exposure or manipulation of sensitive data through agent access
- **Disruption to Connected Systems:** Agent malfunctions causing cascading failures in connected systems

The core AI RMF 2026 document introduces an Agent Profile Card in the MAP Function to systematically profile agentic systems by function, role, autonomy, authority, environment and predictability, and adds a dedicated section on multi-agent ecosystems (MAEs) including a System-Level Agentic Profile for interacting agents. This GOVERN implementation manual assumes those artefacts exist and focuses on the governance procedures required to approve and maintain autonomy/authority levels, assign agentic tiers, choose appropriate human oversight patterns (HITL/HOTL/human-in-command), and ensure accountability for both individual agents and multi-agent ecosystems.

Environmental Sustainability Governance

Environmental sustainability has become a critical governance concern as AI systems, particularly large language models and agentic systems, consume significant computational resources and energy. Organizations must track and manage the environmental impact of AI systems throughout their lifecycle.

Key Environmental Metrics:

- Energy Consumption: Data center energy usage for model training and inference
- Carbon Footprint: CO2 emissions associated with AI system operation
- Resource Optimization: Computational resource utilization and efficiency
- Hardware Lifecycle: Environmental impact of hardware manufacturing, use, and disposal
- Sustainability Reporting: Transparent disclosure of environmental impact to stakeholders

ISO 23894 Integration

ISO/IEC 23894:2023 provides guidance on AI risk management. This manual fully integrates ISO 23894 requirements with extensions for agentic AI and environmental sustainability:

ISO 23894	Base Requirement	AI RMF 2026	Extensions
Clause 5.1	Establish risk management framework	AI Governance Board	Agentic AI Committee, Environmental Sustainability Committee
Clause 5.2	Define risk management policy	AI risk policy	Agentic AI risk assessment, mandatory sustainability metrics
Clause 5.3	Assign roles and responsibilities	RACI matrix for AI risks	Agent ownership (Agent Owner role), environmental accountability
Clause 5.4	Allocate resources	Budget for AI risk management	Resources for agent monitoring, energy measurement tools
Clause 5.5	Establish communication channels	Risk reporting processes	Real-time agent incident reporting, sustainability dashboards

****Singapore Model AI Governance Framework for Agentic AI**

Published February 2026, The Singapore AI Governance Framework for Agentic AI (2026) provides comprehensive guidance for governing autonomous AI agents and multi-agent ecosystems, addressing governance boundaries, system-level oversight patterns, and cross-organizational coordination challenges. The framework introduces governor and auditor agent concepts for automated oversight and emphasizes establishing clear accountability mechanisms when agents from different vendors and organizations interact.

Dimension	Key Practices
1. Singapore Framework focuses on operational governance patterns rather than numbered "dimensions" like Singapore MGF	The framework provides excellent technical guidance for agentic AI governance but contains zero coverage of environmental sustainability (carbon footprint, energy consumption, climate impact) - this gap validates AI RMF 2026's unique value as the only framework integrating both agentic AI governance AND environmental considerations.
2. Assess and Bound Risks Upfront	Determine suitable use cases, define agent limits and permissions, implement agent identity management, threat modeling for agentic systems
3. Make Humans Meaningfully Accountable	Clear allocation of responsibilities, design for meaningful human oversight, address automation bias, adaptive governance
4. Implement Technical Controls	Controls for planning/tools/protocols, pre-deployment testing, continuous monitoring, gradual rollout
5. Enable End-User Responsibility	Transparency about agent capabilities, user education and training, prevent loss of tradecraft and foundational skills

World Economic Forum AI Agents in Action: Foundations for Evaluation and Governance

Published November 2025, the WEF Framework for Agentic AI provides emerging best practices for governing autonomous AI systems.

Dimension	Key Practices
1. Technical foundations	Lay the groundwork
2. Functional Classification	Define the agent's role
3. Evaluation and Governance	Scale with confidence

2. GOVERN Function Overview

The **GOVERN Function** establishes and nurtures a culture of AI risk management throughout the organization. It emphasizes organizational structures, policies, processes, and accountability mechanisms that enable responsible AI development and deployment across all stages of the AI lifecycle, with specific provisions for agentic AI systems and environmental sustainability.

GOVERN provides the foundation for all other functions:

- Establishes leadership commitment and strategic direction for AI governance
- Creates the organizational culture necessary for effective AI risk management
- Defines policies that guide MAP, MEASURE, and MANAGE activities
- Sets governance structures for oversight and accountability, including specialized structures for agentic AI
- Allocates resources and assigns responsibilities across the organization
- Ensures environmental sustainability is integrated into AI governance

2.1 GOVERN Function Structure

Category	Focus Area
GOVERN 1	Organizational AI Governance - Policies, legal compliance, roles, culture, agentic AI, environmental sustainability
GOVERN 2	Accountability and Responsibility - Clear accountability structures
GOVERN 3	Workforce Diversity and Team Composition - Diverse, multidisciplinary teams
GOVERN 4	Risk Management Culture - Training, awareness, and cultural support
GOVERN 5	Oversight and Monitoring - Governance bodies and review processes
GOVERN 6	Trustworthy AI Characteristics - Comprehensive trustworthiness policies

2.2 Key Governance Bodies

Organizations implementing this framework will establish the following governance bodies:

Governance Body	Purpose and Scope
AI Governance Board	Overall AI strategy, policy approval, resource allocation, executive oversight
Agentic AI Committee (NEW)	Specialized governance for autonomous AI systems, agent risk assessment, agent identity management oversight
Environmental Sustainability Committee (NEW)	Track and manage environmental impact of AI systems, set sustainability targets, report carbon footprint
AI Ethics Review Board	Ethical assessment of high-risk AI systems, fairness review, value alignment
Internal Audit Function	Independent assessment of AI management system effectiveness

3. GOVERN 1: Organizational AI Governance

Expected Outcome: Policies, processes, procedures, and practices across the organization related to the mapping, measuring, and managing of AI risks are in place, transparent, and implemented effectively.

3.1 GOVERN 1.1: Legal and Regulatory Requirements

Updated: February 2026 (v1.4 – Singapore Framework Integration)

Objective: Organizations identify, track, and manage applicable legal and regulatory requirements for AI systems throughout their lifecycle. This includes traditional AI regulations and emerging requirements for agentic AI and multi-agent ecosystems as defined in Singapore Model AI Governance Framework.

Expected Outcome:

Policies, procedures, and controls to ensure compliance with applicable legal and regulatory requirements

Framework Integration

- **AI RMF 2026:** GOVERN-1.1 establishes legal/regulatory baseline for all functions
- **ISO/IEC 42001:** Clause 4.2 (Understanding needs and expectations of interested parties), Clause 9.2 (Internal audit)
- **ISO/IEC 27001:** Clause 4.1 (Understanding context), A.18.1 (Compliance with legal requirements)
- **Singapore Framework:** Governance Boundaries - defines regulatory obligations across multi-agent ecosystems and organizational boundaries
- **Singapore MGF:** Dimension 2 (Govern for Transparency) requires clear regulatory compliance documentation

Multi-Agent Ecosystem Regulatory Requirements (Singapore)

For organizations deploying multi-agent ecosystems, Singapore Model AI Governance Framework requires additional regulatory considerations:

1. Cross-Jurisdictional Compliance

- Agents operating across multiple jurisdictions subject to different regulatory regimes
- Example: Agent coordinating between EU (AI Act), US (state laws), and Singapore (MGF) operations
- Requirement: Map each agent's operational geography to applicable legal frameworks

2. Multi-Party Accountability

- When agents from different organizations interact, who is legally responsible?
- SINGAPORE guidance: Establish clear contractual liability allocation before agent deployment
- Document in Agent Profile Card (Section 1.6): Legal entity responsible, liability limits

3. Governor/Auditor Agent Regulatory Status

- Are governor agents considered "AI systems" subject to regulation?
- If governor agents have veto authority over financial transactions, do they require financial services licensing?
- Action Required: Legal review of each governor/auditor agent's regulatory classification

4. Data Governance Across Agent Boundaries

- GDPR/privacy law compliance when agents share data across organizational boundaries
- Requirements: Data Processing Agreements (DPAs) between organizations deploying interacting agents
- See EU AI Act Article 28 for specific multi-party AI system requirements

Key Regulatory Frameworks for AI Systems

Framework	Jurisdiction	Agentic AI Coverage	Multi-Agent Guidance
EU AI Act	European Union	Limited	Not Addressed
Singapore MGF Agentic AI	Singapore	Comprehensive	Basic Guidance
**Singapore Model Framework	International	Comprehensive	Comprehensive
US State AI Laws	US (State-level)	Varies by State	Not Addressed

Implementation Procedures

Procedure 1.1.1: Establish Compliance Register

- Identify applicable regulatory frameworks (EU AI Act, state AI laws, sector-specific regulations, SINGAPORE multi-agent guidance)
- Create centralized compliance register using template (Appendix A)
- Assign ownership for each regulatory requirement to specific role
- Document interpretation and applicability assessment for each requirement
- Establish review cycle (quarterly minimum) for regulatory landscape monitoring
- **NEW - Multi-Agent Ecosystems:** For each agent in ecosystem, document regulatory classification and responsible legal entity (see Agent Profile Card Section 1.6)
- **SINGAPORE-specific:** Cross-organizational agent ecosystems require mapping regulatory authority across all participating entities

Procedure 1.1.2: Monitor Regulatory Developments

- Subscribe to regulatory update services and official government channels (including Singapore Model AI Governance Framework updates)
- Designate compliance monitoring lead responsible for tracking changes
- Conduct quarterly regulatory horizon scanning sessions
- Document new or changed requirements in compliance register within 30 days
- Assess impact of regulatory changes on existing AI systems and governance processes

Procedure 1.1.3: Conduct Regulatory Gap Analysis

For each AI system in inventory:

- Map system to applicable regulatory requirements
- Assess current compliance status (compliant/partial/non-compliant)
- Document gaps and required remediation actions
- **SINGAPORE-specific:** If system is multi-agent ecosystem, perform gap analysis at BOTH individual agent level AND system level

Prioritize remediation based on:

- Legal/regulatory deadline (e.g., EU AI Act compliance date)
- Risk severity (high-risk systems first)
- Business impact (customer-facing systems prioritized)

Create remediation roadmap with ownership and timelines

Report gap analysis to AI Governance Committee quarterly

Procedure 1.1.4: Conduct Compliance Assessments

- Perform initial compliance assessment for each new AI system during MAP phase
- Conduct annual compliance audits for all deployed AI systems
- Document compliance evidence and gap analysis
- Create remediation plans for identified compliance gaps
- Track remediation progress in compliance register

Legal and Regulatory Requirements

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Maintain compliance register	AI Governance Office	Chief Compliance Officer	Legal Counsel	AI Governance Committee
Monitor regulatory changes	Legal Counsel	Chief Compliance Officer	AI Governance Office	Business Units
Conduct compliance assessments	AI Governance Office	Chief Compliance Officer	Legal Counsel, System Owners	AI Governance Committee
Multi-agent ecosystem regulatory mapping (SINGAPORE)	Legal Counsel	Chief Compliance Officer	AI Governance Office	AI Governance Committee

- **Note:** The highlighted row (Multi-agent ecosystem regulatory mapping) is the NEW v1.4 addition integrating Singapore Model AI Governance Framework requirements.

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	Singapore Framework
4.2, 9.2	4.1, A.18.1	Risk Management Process	Dimension 2: Govern for Transparency	Governance Boundaries

Key Deliverables

- AI Compliance Register (Template: Appendix A)
- Regulatory Monitoring Plan
- Compliance Assessment Reports
- Gap Analysis and Remediation Plans
- **NEW:** Multi-Agent Ecosystem Regulatory Mapping (for organizations with agentic AI)

Success Metrics

- 100% of applicable regulations identified and documented in Compliance Register
- Compliance Register updated within 30 days of regulatory change
- Zero regulatory violations or enforcement actions
- Gap analysis completed quarterly for all AI systems
- **NEW:** For multi-agent ecosystems: 100% of agents have documented regulatory classification and responsible legal entity (Agent Profile Card Section 1.6)
- **NEW:** Cross-jurisdictional compliance mapping completed for all multi-agent ecosystems

1.3 SINGAPORE-Specific Metrics - Detailed Specifications

Metric	Definition	Target	Frequency
Agent Regulatory Classification Tracking	Percentage of agents with documented regulatory status (e.g., "subject to EU AI Act", "exempt", "requires financial services licensing")	100%	Quarterly
Legal Entity Accountability	Percentage of agents with clearly documented responsible legal entity and liability limits in Agent Profile Card	100%	Quarterly
Cross-Jurisdictional Compliance Mapping	Number of multi-agent ecosystems with completed regulatory mapping across all operating jurisdictions	100% of MAEs	Quarterly
Jurisdictional Coverage Completeness	Percentage of agent operational geographies with identified applicable legal frameworks (EU AI Act, US state laws, Singapore MGF, etc.)	100%	Quarterly

Governor Agent Intervention Rate	Frequency of governor agent interventions (veto, rollback) as percentage of total agent transactions	< 5%	Monthly
Auditor Agent Anomaly Detection	Number of governance violations or anomalies detected by auditor agents per month	Track trend	Monthly
Governor/Auditor Intervention Response Time	Average time from anomaly detection to governance action (veto, escalation, or remediation)	< 5 minutes	Monthly

Implementation Notes

- **Dashboard Integration:** Add SINGAPORE-specific metrics to KPI Dashboard (Appendix T)
- **Governance Review:** AI Governance Committee reviews SINGAPORE metrics quarterly alongside traditional compliance metrics
- **Escalation Thresholds:** Governor intervention rates >10% or response times >15 minutes trigger immediate AI Governance Committee review
- **Data Sources:** Metrics collected from Agent Profile Cards (Section 1.6), System-Level Agentic Profiles, governor/auditor agent logs, and Compliance Register
- **Applicability:** SINGAPORE-specific metrics only apply to organizations deploying multi-agent ecosystems. Organizations with single-agent deployments use traditional metrics only.

3.2 GOVERN 1.2: Roles and Responsibilities

Objective: Clear assignment of roles, responsibilities, and authorities for AI governance across the organization.

Key Organizational Roles

Role	Key Responsibilities
Board/Executive Leadership	Strategic oversight, accountability, resource allocation, policy approval
Chief AI Officer (CAIO)	Overall AI strategy, governance coordination, cross-functional alignment
Chief Risk Officer (CRO)	AI risk assessment framework, risk appetite, escalation management
Chief Information Security Officer (CISO)	AI security architecture, resilience, cybersecurity controls
Data Protection Officer (DPO)	Privacy compliance, data governance, GDPR/data protection requirements
AI Ethics Committee	Ethical review, fairness assessment, value alignment oversight
Agent Owner (NEW)	Individual agent behavior, actions within defined scope, agent monitoring oversight
Environmental Sustainability Officer (NEW)	Track environmental metrics, carbon footprint reporting, sustainability compliance
AI Developers/Engineers	System design, implementation, technical controls, testing
Business Units	AI deployment, operational monitoring, user feedback, incident reporting

Implementation Procedures

Procedure 1.2.1: Define and Document Roles

- Establish AI Governance Committee with executive-level authority
- Define charter for each governance role including scope, authority, and accountability
- Create role description documents with specific AI risk management responsibilities
- Map roles to organizational structure and reporting lines
- Obtain Board approval for governance structure and role assignments

Procedure 1.2.2: Assign Role Owners

- Identify qualified individuals for each defined role
- Document role assignments in AI Governance Register
- Communicate role assignments and expectations organization-wide
- Establish succession planning for critical AI governance roles
- Review role assignments annually and update as needed

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Define governance structure	Chief AI Officer	CEO	Board, CRO, CISO	All Employees
Document role descriptions	Chief AI Officer	HR	Legal, Risk	Department Heads
Assign individuals to roles	HR	Chief AI Officer	Department Heads	Board
Communicate role assignments	Chief AI Officer	CEO	Communications	All Employees

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
5.3 - Organizational roles, responsibilities, and authorities	5.3 - Organizational roles and responsibilities	5.3 - Assign roles and responsibilities (RACI matrix for AI risks, Agent Owner role)	Dimension 2 - Clear allocation of responsibilities within and outside organization	Art. 16 - Obligations of providers of providers (must designate responsible persons)

Key Deliverables

- AI Governance Committee Charter
- Role Description Documents for each governance role
- AI Governance Organizational Chart
- Role Assignment Matrix (RACI) - Appendix B

3.3 GOVERN 1.3: Diversity, Equity, Inclusion, and Accessibility (DEIA)

Objective: AI design and development teams reflect diversity across multiple dimensions to reduce bias and improve outcomes.

Implementation Procedures

Procedure 1.3.1: Establish DEIA Goals

- Define specific, measurable DEIA targets for AI teams (gender, race/ethnicity, disability, age, etc.)
- Incorporate DEIA metrics into hiring and retention KPIs
- Document DEIA goals in AI Governance Policy
- Review progress quarterly with executive leadership
- Publish annual DEIA report for AI teams

Procedure 1.3.2 Establish GPAIS-Specific Risk-Management Intensity and Unacceptable Risk Thresholds

- Classify all general-purpose AI systems (GPAIS), foundation models, and frontier models in scope and explicitly flag them in the AI System and Agent Registry Appendix M.
- For these systems, define heightened minimum expectations for risk management, including:
 - earlier and more frequent risk assessments across the lifecycle (pre-training, post-training, pre-deployment, post-deployment);
 - expanded stakeholder engagement, including impacted communities and external experts;
 - mandatory red-team and adversarial testing focused on dangerous capabilities, misuse/abuse, and systemic/catastrophic scenarios.
 - Define explicit “unacceptable risk” thresholds for GPAIS consistent with NIST AI RMF Map 1.5 and the Berkeley GPAIS Profile, including cases where significant negative impacts are imminent, severe harms are occurring, or catastrophic risks are present.
 - Mandate that development, training, deployment, or scale-up of a GPAIS must cease in a safe manner when those thresholds are met, until risks are sufficiently managed and independently reviewed.
 - Document GPAIS-specific thresholds and cease-development conditions in the AI Risk Appetite Statement Appendix N.1 and link them to the Decision Authority Matrix Appendix C for enforceability.

Procedure 1.3.3: Implement Inclusive Practices

- Conduct bias training for all AI team members (minimum annually)
- Use structured interview processes to reduce hiring bias
- Ensure diverse representation on AI project teams and review committees
- Implement accessibility requirements in all AI system design
- Create inclusive work environment with accommodations and support

Procedure 1.3.4 Incorporate Diverse Perspectives

- Include diverse stakeholders in AI design review sessions
- Conduct user research with representative populations
- Ensure AI Ethics Committee includes diverse membership
- Seek external perspectives through advisory boards and consultants
- Document how diverse perspectives influenced design decisions

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Establish DEIA goals	Chief AI Officer	CEO	HR, DEIA Officer	All Employees
Conduct bias training	HR/Learning & Development	Chief AI Officer	DEIA Experts	AI Teams
Ensure diverse hiring	HR	Department Heads	DEIA Officer	Chief AI Officer
Track DEIA metrics	HR Analytics	Chief AI Officer	DEIA Officer	Board

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
6.1.3 - AI impact assessment (considers fairness); 7.2 - Competence (includes diversity)	7.2 - Competence	6.1 - AI system context (diversity in development teams reduces bias risks)	Dimension 1 - Assess risks (diverse teams identify broader range of risks)	Art. 10 - Data and data governance (representative datasets require diverse perspectives)

Key Deliverables

- DEIA Goals and Metrics for AI Teams
- Inclusive Hiring and Retention Procedures
- Bias Awareness Training Program
- Annual DEIA Progress Report

3.4 GOVERN 1.4: Organizational AI Risk Management Culture

Objective: Foster a culture where identifying and addressing AI risks is encouraged, valued, and rewarded.

Implementation Procedures

Procedure 1.4.1: Demonstrate Leadership Commitment

- Executive team communicates AI risk management priorities in organization-wide forums
- Include AI risk management objectives in executive performance goals
- Allocate budget and resources demonstrating prioritization of AI risk management
- Leadership participates in AI risk review sessions and governance meetings
- Celebrate and publicize AI risk management successes

Procedure 1.4.2: Establish Speak-Up Culture

- Create multiple channels for reporting AI risks and concerns (hotline, email, web form)
- Implement anonymous reporting option with protection against retaliation
- Train managers on receiving and responding to AI risk concerns
- Publicly share how reported concerns led to improvements (anonymized)
- Track and monitor all reported AI risks through to resolution

Procedure 1.4.3: Reward Risk Identification

- Incorporate AI risk identification into performance review criteria
- Establish recognition program for exemplary AI risk management
- Share success stories in company communications
- Provide professional development opportunities related to AI risk management
- Never penalize individuals for raising legitimate AI concerns

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Communicate AI risk priorities	CEO/Executive Team	Board	Chief AI Officer	All Employees
Establish reporting channels	Chief Risk Officer	Chief AI Officer	IT, HR, Legal	All Employees
Train managers on risk culture	HR/Learning & Development	Chief AI Officer	Risk Team	All Managers
Recognize risk identification	Department Heads	Chief AI Officer	HR	All Employees

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
5.1 - Leadership and commitment; 7.3 - Awareness	5.1 - Leadership and commitment; 7.3 - Awareness	5.1 - Establish risk management framework (organizational culture component)	Dimension 2 - Make humans meaningfully accountable (speak-up culture enables accountability)	Art. 9 - Risk management system (must include organizational measures)

Key Deliverables

- AI Risk Reporting Channels and Procedures
- Leadership Communication Plan for AI Risk Management
- AI Risk Recognition and Rewards Program
- Annual Culture Assessment Survey

3.5 GOVERN 1.5: Agentic AI Governance

NEW IN AI RMF 2026: This is a critical new subcategory for governing autonomous AI systems.

Expected Outcome: Specialized governance structures, policies, and processes are established for agentic AI systems to ensure they operate safely, transparently, and within appropriate boundaries while maintaining human accountability.

Objective: Establish comprehensive governance framework specifically for agentic AI systems, including committee structure, agent identity management, risk assessment processes, and continuous oversight mechanisms.

Context: Why Agentic AI Requires Specialized Governance

Agentic AI systems differ fundamentally from traditional AI applications:

- **Action-Taking Capability:** Agents can execute actions that directly impact databases, systems, and external services
- **Multi-Step Planning:** Agents plan and execute sequences of actions without human intervention at each step
- **Dynamic Adaptation:** Agents respond to environmental changes and new information in real-time
- **Multi-Agent Interactions:** Multiple agents may work together, creating complex interaction patterns
- **Autonomous Decision-Making:** Agents make decisions within their authorized scope without predefined workflows

These capabilities create new governance challenges:

- **Unpredictability:** Agent behavior emerges from interactions rather than fixed logic
- **Cascading Risks:** Errors in one agent can propagate to connected agents and systems
- **Accountability Gaps:** Traditional responsibility models tied to static workflows don't fit autonomous systems
- **Attack Surface:** New components (planning, tools, protocols) create additional security vulnerabilities

Implementation Procedures

Procedure 1.5.1: Establish Agentic AI Committee

- Define committee charter specifying scope: All agentic AI systems across the organization
- Appoint committee members with diverse expertise: AI engineering, security, risk management, legal, ethics, domain experts
- Establish meeting frequency: Monthly minimum, with ad-hoc meetings for critical agentic deployments
- Define decision authority: Approval required for all high-risk agentic AI deployments, agent identity frameworks, agentic incident response
- Create escalation path to AI Governance Board for critical agentic risks
- Document all committee decisions including rationale and risk assessments

Procedure 1.5.2: Implement Agent Identity Management Framework

Agent Identity Principles:

- Unique Identification: Each agent has a unique identity distinguishable from human users
- Hierarchical Ownership: Agent identity is linked to supervising agent, human user, or organizational unit
- Capacity Recording: Document whether agent acts independently or on behalf of specified human
- Permission Inheritance: Agents cannot receive permissions exceeding those of authorizing human
- Audit Trail: All agent identity delegations and permission grants are logged

Implementation Steps:

- Develop agent identity schema compatible with existing identity management systems
- Implement agent authentication mechanisms (API keys, tokens, certificates)
- Create agent authorization framework with role-based and dynamic permissions
- Establish agent identity lifecycle management (creation, modification, deactivation)
- Deploy agent identity monitoring and audit logging
- Integrate with existing IAM (Identity and Access Management) systems where possible

Procedure 1.5.3: Define Agent Boundaries and Permissions

For each agentic AI system, establish:

Action-Space Boundaries:

- Define permitted tools and APIs agent can access
- Specify data sources agent can read from
- Identify systems agent can write to or modify
- Determine if agent can access external systems or only internal/sandboxed environments
- Document reversibility of actions agent can take

Autonomy Boundaries:

- Define whether agent follows detailed Standard Operating Procedures (SOPs) or uses independent judgment
- Specify checkpoints requiring human approval (high-stakes actions, irreversible changes, outlier behavior)
- Establish escalation triggers for unexpected situations
- Document level of human involvement (human approves all steps, collaboration, human approves critical steps only, observation only)
- Define emergency shutdown conditions and procedures

Environmental Boundaries:

- Implement sandboxed execution environments for high-risk tasks
- Establish network access restrictions
- Define data access limitations (by classification level, department, purpose)
- Create isolated test environments mirroring production
- Implement kill switches and fallback procedures

Procedure 1.5.4: Conduct Agentic AI Risk Assessments

Agent Profile Card and Tier Linkage

The Agent Risk Assessment MUST be informed by the Agent Profile Card defined in the AI RMF 2026 MAP Function. Before completing this template, the following fields SHALL be captured for each agentic system in the Agent Profile Card: function, role (specialist/generalist), autonomy level, authority level, environment complexity, predictability, and key dependencies/interactions.

GOVERN is responsible for approving the initial autonomy and authority levels recorded in the Agent Profile Card and for confirming the resulting agentic risk tier (for example, Tier 1 – Automated Tools; Tier 2 – Advisory Agents; Tier 3 – Operational Agents; Tier 4 – Critical Agentic Systems). The approved tier **MUST** be recorded on both the Agent Profile Card and the Agent Risk Assessment Template, and **SHALL** be used by MAP, MEASURE and MANAGE functions to determine evaluation depth, monitoring expectations and oversight models (HITL/HOTL/human-in-command).

Use the Agentic AI Risk Assessment Template (Appendix H) to evaluate:

- **Domain and Use Case Risk:** Assess tolerance for error in deployment domain (financial transactions vs. meeting summaries)
- **Data Access Risk:** Evaluate sensitivity of data agent can access (PII, confidential, public)
- **System Access Risk:** Determine if agent accesses external systems and their trustworthiness
- **Action Scope Risk:** Assess range of actions (read-only vs. write, few tools vs. many, computer use)
- **Reversibility Risk:** Evaluate if agent actions can be undone (schedule meeting vs. send email)
- **Autonomy Risk:** Determine unpredictability level (follows SOP vs. independent planning)
- **Task Complexity Risk:** Assess number of steps and analysis depth required
- **Threat Modeling:** Identify specific attack vectors (memory poisoning, tool misuse, privilege compromise)
- **Cascading Impact:** Evaluate potential for agent errors to propagate to other systems

Risk Classification:

Risk Level	Characteristics	Governance Requirements
Critical	Financial transactions, healthcare decisions, legal actions, external communications	Agentic AI Committee approval required. Extensive human oversight. Real-time monitoring. Staged rollout. External audit.
High	Write access to sensitive data, customer-facing, many tools available, high autonomy	Department head approval. Human-in-the-loop at critical steps. Continuous monitoring. Regular audits.
Moderate	Internal tools only, read access to sensitive data, defined SOPs, reversible actions	Technical lead approval. Human approval for significant actions. Standard monitoring.
Low	Sandboxed only, public data, very limited tools, follows strict procedures	Standard approval process. Human oversight available. Basic monitoring.

Procedure 1.5.5: Establish Agentic Monitoring and Oversight

- Deploy real-time agent activity monitoring for all production agents
- Implement logging of all agent actions (tool calls, data access, system modifications)
- Create monitoring dashboards showing: agent status, actions taken, errors/failures, policy violations, resource consumption
- Define alert thresholds for: unauthorized access attempts, repeated failures, outlier behavior, policy violations, resource overconsumption
- Establish incident response procedures specific to agentic malfunctions
- Conduct regular audits of agent behavior and human oversight effectiveness
- Implement agents-monitoring-agents for anomaly detection
- Review agent audit logs weekly minimum, daily for high-risk agents

Procedure 1.5.6: Address Automation Bias in Human Oversight

Automation bias is the tendency to over-trust automated systems, especially after they perform reliably. For agentic AI, implement:

- Train human overseers to identify common agent failure modes: hallucinated reasoning, tool misuse, outdated policy references
- Randomize oversight checks: Require detailed review of randomly selected agent workflows, not just failures
- Implement red team testing: Have security teams attempt to make agents take harmful actions
- Regular audit of human approvals: Review whether human overseers are rubber-stamping agent requests
- Rotate oversight responsibilities: Prevent complacency by rotating which humans oversee which agents
- Feedback loops: Provide human overseers with outcomes of approved agent actions
- Maintain tradecraft: Ensure humans retain skills being automated by agents through regular practice

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Establish Agentic AI Committee	Chief AI Officer	CEO	Board, CRO, CISO	All Stakeholders
Implement agent identity management	CISO, AI Engineering	Chief AI Officer	Security Team, IAM Team	Agentic AI Committee
Define agent boundaries and permissions	Agent Owner	Agentic AI Committee	Security, Risk, Legal	Business Units
Conduct agentic risk assessments	AI Risk Team	Agentic AI Committee	Agent Owner, Security	Chief AI Officer
Monitor agent behavior	Agent Owner	System Owner	Security Operations	Agentic AI Committee
Respond to agent incidents	Agent Owner	Incident Response Lead	Security, Engineering	Agentic AI Committee, Executives

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
5.3 - Roles and authorities	5.3 - Roles	5.1 - Establish framework (with Agentic AI Committee)	Dimension 1 - Assess and bound risks upfront	Art. 9 - Risk management
8.1 - Operational planning	8.1 - Operational planning	5.2 - Define policy (with agentic risk requirements)	Dimension 2 - Make humans accountable	Art. 14 - Human oversight
6.1.3 - AI impact assessment	A.5.1 - Security policies	5.3 - Assign roles (Agent Owner)	Dimension 3 - Technical controls	Art. 16 - Provider obligations
9.1 - Monitoring	9.1 - Monitoring	5.4 - Allocate resources (agent monitoring tools)	Dimension 4 - End-user responsibility	Art. 72 - Post-market monitoring

Key Deliverables

- Agentic AI Committee Charter (Appendix F)
- Agent Identity Management Framework Documentation
- Agent Boundaries and Permissions Policy
- Agentic AI Risk Assessment Template (Appendix H)
- Agent Monitoring Dashboard Specifications
- Agentic Incident Response Playbook (Appendix I)
- Automation Bias Mitigation Training Program

Integration with Singapore MGF for Agentic AI

This subcategory directly implements the Singapore Model AI Governance Framework for Agentic AI (January 2026):

Singapore MGF Dimension	Implementation in GOVERN 1.5
1. Assess and Bound Risks Upfront	Procedure 1.5.3 (Define Agent Boundaries), Procedure 1.5.4 (Conduct Agentic Risk Assessments)
2. Make Humans Meaningfully Accountable	Procedure 1.5.1 (Establish Committee), Procedure 1.5.6 (Address Automation Bias), Agent Owner role in RACI
3. Implement Technical Controls	Procedure 1.5.2 (Agent Identity Management), Procedure 1.5.5 (Monitoring and Oversight)
4. Enable End-User Responsibility	Training programs, transparency requirements, tradecraft maintenance in Procedure 1.5.6

3.6 GOVERN 1.6: Environmental Sustainability Governance

NEW IN AI RMF 2026: This is a critical new subcategory for governing the environmental impact of AI systems.

Expected Outcome: Comprehensive environmental sustainability governance structures, policies, and processes are established to track, manage, and minimize the environmental impact of AI systems throughout their lifecycle, with transparent reporting to stakeholders.

Objective: Establish governance framework for environmental sustainability of AI systems, including committee structure, carbon footprint tracking, energy monitoring, resource optimization, and transparent environmental impact reporting.

Context: Why Environmental Sustainability Requires Governance

AI systems, particularly large language models and agentic AI systems, consume significant computational resources and energy. Environmental sustainability has become a critical governance concern:

- **Energy Consumption:** Training large AI models can consume megawatt-hours of electricity, equivalent to the annual consumption of multiple households
- **Carbon Footprint:** Data center operations for AI inference generate substantial CO2 emissions, especially when powered by non-renewable energy
- **Resource Utilization:** AI systems require specialized hardware (GPUs, TPUs) with significant manufacturing environmental impact
- **Scalability Impact:** As AI deployment scales across organizations, cumulative environmental impact grows exponentially
- **Stakeholder Expectations:** Investors, regulators, customers, and employees increasingly demand environmental accountability

These factors create governance imperatives:

- **Regulatory Compliance:** Emerging regulations require environmental impact disclosure and reduction targets
- **Corporate Responsibility:** Organizations must demonstrate environmental stewardship as part of ESG commitments
- **Cost Management:** Energy costs represent significant operational expenses for AI systems
- **Competitive Advantage:** Environmental leadership differentiates organizations in the marketplace
- **Risk Mitigation:** Climate-related risks must be managed proactively

Implementation Procedures

Procedure 1.6.1: Establish Environmental Sustainability Committee

- Define committee charter specifying scope: Environmental impact of all AI systems across the organization
- Appoint committee members with diverse expertise: Environmental science, energy management, sustainability reporting, AI engineering, finance, legal
- Establish meeting frequency: Quarterly minimum, with ad-hoc meetings for significant AI deployments
- Define decision authority: Approval required for high-energy AI deployments, sustainability targets, environmental policies, reporting commitments
- Create escalation path to AI Governance Board and executive leadership
- Establish coordination with existing corporate sustainability functions
- Document all committee decisions including environmental impact assessments and mitigation strategies

Procedure 1.6.2: Define Sustainability Metrics and Targets

Establish comprehensive metrics framework:

- Define baseline measurements: Current energy consumption, carbon emissions, resource utilization for all AI systems
- Set quantitative reduction targets: Specific percentage reductions over defined timeframes (e.g., 30% reduction in carbon intensity by 2027)
- Establish measurement methodologies: Standardized approaches for calculating energy use (kWh), carbon footprint (tonnes CO₂e), computational efficiency (FLOPs/Watt)
- Define reporting frequency: Monthly internal tracking, quarterly executive reporting, annual public disclosure
- Create benchmarking framework: Compare performance against industry standards and best practices
- Align with corporate sustainability goals: Integrate AI-specific targets with broader organizational environmental commitments
- Document metrics in Environmental Sustainability Policy

Key Environmental Metrics:

Metric Category	Measurement Unit	Description
Energy Consumption	kWh, MWh	Total electricity consumed for model training and inference operations
Carbon Footprint	tonnes CO2e	Total greenhouse gas emissions (Scope 2 and Scope 3) from AI operations
Carbon Intensity	gCO2e/inference	Carbon emissions per model inference or prediction
Computational Efficiency	FLOPs/Watt	Floating point operations per watt of energy consumed
Renewable Energy %	Percentage	Percentage of AI workload energy from renewable sources
Hardware Utilization	Percentage	Average GPU/TPU utilization rate (higher is more efficient)
Water Usage (PUE)	Liters	Water consumed for data center cooling

Procedure 1.6.3: Implement Energy Monitoring Systems

- Deploy energy monitoring tools: Implement software and hardware solutions to track real-time energy consumption of AI workloads
- Instrument data center infrastructure: Install meters and monitoring systems on AI-dedicated compute resources (GPU clusters, TPU pods)
- Tag AI workloads: Implement cost allocation tags to attribute energy consumption to specific AI systems and projects
- Create monitoring dashboards: Real-time visibility into energy consumption by system, team, and project
- Establish alert thresholds: Automated alerts when systems exceed energy budgets or efficiency targets
- Integrate with cloud provider tools: Leverage cloud provider energy tracking and carbon reporting capabilities
- Monitor model training separately from inference: Different optimization strategies apply to each phase

Procedure 1.6.4: Track and Report Carbon Footprint

Comprehensive carbon accounting approach:

- Calculate Scope 2 emissions: Indirect emissions from purchased electricity for AI operations
- Calculate Scope 3 emissions: Embodied carbon in hardware manufacturing, cloud provider upstream emissions
- Apply carbon intensity factors: Convert energy consumption to CO2e using regional grid carbon intensity
- Track by AI system lifecycle phase: Separate accounting for development, training, deployment, inference, decommissioning
- Document calculation methodology: Transparent approach using recognized standards (GHG Protocol, ISO 14064)
- Conduct third-party verification: Annual independent audit of carbon accounting
- Report internally and externally: Executive dashboards, sustainability reports, regulatory filings

Carbon Footprint by AI Lifecycle Phase:

Lifecycle Phase	Carbon Accounting Approach
Development & Experimentation	Track energy consumption of development environments, prototype training runs, experiment iterations. Often highest per-model carbon cost due to exploration.
Model Training	Measure energy consumption during final model training. For large models, this is typically the single largest carbon cost. Document training duration, hardware type, location.
Deployment & Infrastructure	Account for energy to maintain deployed models, infrastructure overhead, load balancers, monitoring systems. Ongoing operational emissions.
Inference at Scale	Calculate total inference energy across all requests. For high-traffic models, cumulative inference carbon can exceed training carbon within months.
Hardware Embodied Carbon	Include Scope 3 emissions from GPU/TPU manufacturing. Amortize hardware embodied carbon across expected useful life.
Decommissioning & Disposal	Account for responsible hardware disposal, data deletion energy costs, final model archiving.

Procedure 1.6.5: Optimize Resource Utilization

Implement technical and operational optimization strategies:

Model Optimization:

- Model compression: Apply quantization, pruning, and distillation to reduce model size and computational requirements
- Efficient architectures: Select model architectures optimized for inference efficiency (e.g., MobileBERT, DistilBERT for appropriate use cases)
- Right-sizing: Match model capacity to task complexity, avoid over-provisioning
- Caching and reuse: Cache model outputs and intermediate results to avoid redundant computation
- Batch optimization: Optimize batch sizes for throughput and energy efficiency

Infrastructure Optimization:

- Hardware selection: Choose energy-efficient hardware for deployment (TPUs, specialized inference accelerators)
- Geographic placement: Deploy AI workloads in data centers with renewable energy availability
- Workload scheduling: Schedule energy-intensive training during low-carbon grid periods
- Dynamic scaling: Implement auto-scaling to match compute resources to demand
- Hardware lifecycle management: Extend hardware life through efficient utilization, responsible disposal

Operational Optimization:

- Training efficiency: Use transfer learning, few-shot learning to reduce training compute
- Model sharing: Share trained models across teams to avoid redundant training
- Experiment management: Rigorous experiment tracking to avoid unnecessary model iterations
- Decommissioning: Promptly decommission unused AI systems to eliminate ongoing energy consumption
- Continuous monitoring: Regular review of energy efficiency metrics and optimization opportunities

Procedure 1.6.6: Report Environmental Impact

- Prepare internal reports: Monthly energy and carbon reports to AI leadership and Environmental Sustainability Committee
- Executive dashboard: Real-time sustainability metrics visible to C-suite and Board
- Annual sustainability report: Comprehensive disclosure of AI environmental impact in corporate sustainability report

- Regulatory filings: Comply with mandatory environmental disclosure requirements (SEC climate rules, EU CSRD)
- Stakeholder communications: Transparent communication of environmental performance to employees, customers, investors
- Industry reporting: Participate in industry benchmarking and transparency initiatives
- Third-party verification: Annual independent assurance of environmental data and reporting

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Establish Environmental Sustainability Committee	Chief Sustainability Officer	CEO	Chief AI Officer, CFO	Board, Stakeholders
Define sustainability metrics and targets	Environmental Sustainability Officer	Environmental Sustainability Committee	AI Engineering, Finance	Executive Team
Implement energy monitoring	AI Operations Team	Environmental Sustainability Officer	Data Center Operations, Cloud Team	Environmental Committee
Track carbon footprint	Environmental Sustainability Officer	Environmental Sustainability Committee	Finance, AI Engineering	Executive Team
Optimize resource utilization	AI Engineering Teams	System Owners	Environmental Officer, Cloud Architects	Environmental Committee
Report environmental impact	Environmental Sustainability Officer	Chief Sustainability Officer	Finance, Legal, Communications	Board, Investors, Public

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
4.1 - Understanding organization and context	4.1 - Context of organization	4.1 - Understand context (data center environmental context)	Dimension 1 - Environmental impact as risk factor	Art. 9 - Risk management (environmental risks)
5.1 - Leadership and commitment	5.1 - Leadership	5.1 - Establish framework (Environmental Sustainability Committee)	Dimension 2 - Leadership accountability for environmental impact	Recital 60 - Environmental sustainability considerations
5.3 - Organizational roles	5.3 - Organizational roles	5.3 - Assign roles (Environmental Sustainability Officer)	Dimension 2 - Clear environmental accountability	Art. 16 - Provider obligations include environmental transparency
6.1.3 - AI impact assessment	8.1 - Operational planning	5.4 - Allocate resources (energy measurement tools)	Dimension 3 - Technical controls for monitoring	Art. 72 - Post-market monitoring includes environmental metrics
9.1 - Monitoring, measurement, analysis	9.1 - Monitoring and measurement	5.5 - Communication (sustainability dashboards, environmental reporting)	Dimension 4 - Stakeholder transparency on environmental impact	Art. 13 - Transparency obligations include environmental information

Key Deliverables

- Environmental Sustainability Committee Charter (Appendix G)
- Environmental Sustainability Policy
- Carbon Footprint Assessment Template (Appendix J)
- Energy Monitoring Dashboard Specifications
- Sustainability Metrics Framework
- Environmental Impact Report Template
- Resource Optimization Guidelines

Integration with ISO 23894 Environmental Extensions

This subcategory directly implements the AI RMF 2026 environmental sustainability extensions to ISO/IEC 23894:2023:

ISO 23894 Clause	AI RMF 2026 Environmental Extension Implementation
5.1 - Establish framework	Procedure 1.6.1: Environmental Sustainability Committee with governance authority over AI environmental impact
5.2 - Define policy	Procedure 1.6.2: Mandatory sustainability metrics in AI risk policy, quantitative reduction targets
5.3 - Assign roles	Environmental Sustainability Officer role in RACI matrix with clear accountability
5.4 - Allocate resources	Procedure 1.6.3: Energy monitoring tools, infrastructure investment for sustainability tracking
5.5 - Communication	Procedure 1.6.6: Sustainability dashboards, transparent reporting to stakeholders
4.1 - Organizational context	Data center environmental context, geographic carbon intensity considerations
4.2 - Stakeholder needs	Environmental impact communities, investor ESG requirements, regulatory disclosure obligations

3.7 GOVERN 1.7: Third-Party & Tool Ecosystem Governance

NEW IN AI RMF 2026: This subcategory establishes governance requirements for external vendors, model providers, agent tools, and cloud services used to build or operate AI systems.

Objective: Ensure third-party AI components and tool ecosystems are selected, contracted, monitored, and exited in a controlled manner with clear shared responsibility, auditability, security, privacy, and performance expectations.

Implementation Procedures

Procedure 1.7.1: Establish Third-Party AI and Agent Tool Inventory

Maintain a centralized inventory of all third-party AI components, including: foundation model APIs, hosted model endpoints, agent tools and plugins, vector databases, data labeling services, monitoring services, and cloud AI services (Template: Appendix L.1). Assign an internal owner for each third-party component (System Owner and/or Agent Owner) and record the business purpose, data categories, environments, and integrations.

Classify each third-party component by criticality and risk tier based on: autonomy level supported, data sensitivity, external action authority, and operational dependency. Link each inventory entry to the applicable contracts, security assessments, and monitoring obligations.

Procedure 1.7.2: Perform AI Supplier Due Diligence and Risk Tiering

Conduct pre-engagement due diligence using the Supplier Due Diligence Checklist (Appendix L.2) and document evidence (SOC reports, ISO certifications, security whitepapers, model documentation, incident history where available). Evaluate supplier controls across: security, privacy, compliance, model update/change management, logging and traceability, data use limitations, performance SLAs, and business continuity.

Assess agentic-specific risks: tool permissioning, external action execution safeguards, rate limits, transaction controls, and human approval capabilities. Assign a supplier risk tier (Low/Moderate/High) and determine required approvals using the Decision Authority Matrix (Appendix C).

Procedure 1.7.3: Contracting and Shared Responsibility Requirements

Include minimum contract requirements for AI suppliers (Appendix L.3), including: audit rights, incident notification SLAs, data processing restrictions, model change notification, vulnerability disclosure, and termination/exit support. Define shared responsibility for security, privacy, monitoring, and incident response using the Shared Responsibility Matrix (Appendix L.4).

Require operational transparency appropriate to risk tier: logging access, retention commitments, and support for investigation (e.g., request tracing, model/version identifiers).

For High risk suppliers or agentic toolchains, require an exit/transition plan prior to production use (Appendix L.5).

Procedure 1.7.4: Continuous Monitoring of Supplier Performance and Changes

Establish monitoring for supplier SLAs and operational changes (service availability, latency, rate limits, pricing, policy changes, model updates).

Perform periodic reassessment based on supplier risk tier: at least annually for Moderate, semi-annually for High, and after material changes (new model version, new tool permissions, new data categories).

Track third-party incidents and near-misses in the AI incident register and require root-cause analysis for supplier-driven events.

Integrate supplier monitoring into the AI System and Agent Registry (Appendix M) to ensure portfolio visibility.

Procedure 1.7.5: Exit, Transition, and Termination Controls

Maintain a documented exit plan for each High risk supplier, including: data export, key rotation, credential revocation, replacement options, and migration steps.

Ensure data deletion/return obligations are executed and evidenced upon termination.

Retain contractual and operational records in accordance with the organization's retention policy and applicable regulatory requirements.

Conduct a post-exit review to capture lessons learned and update supplier criteria.

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Maintain third-party AI/tool inventory	Procurement Lead	Chief AI Officer	CISO; Legal Counsel; System Owners	Executive Team
Conduct supplier due diligence and tiering	AI Risk Team	Chief Risk Officer	CISO; Legal Counsel; Procurement Lead; Data Protection Officer	Chief AI Officer
Negotiate AI supplier contract clauses	Legal Counsel	Chief Procurement Officer	Chief AI Officer; CISO; Chief Risk Officer	System Owners
Approve High risk suppliers for production	AI Governance Committee	Chief AI Officer	Chief Risk Officer; Legal Counsel; CISO	Board
Monitor supplier performance and changes	System Owners/Agent Owners	Chief AI Officer	AI Risk Team; Procurement Lead	AI Governance Committee

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
8.1 - Operational planning and control; 7.4 - Communication; 8.2 - Change management (as applicable)	5.19 - Supplier relationships; 5.22 - Supplier service delivery; 5.23 - Use of cloud services; 5.24-5.27 - Incident management	Risk identification and treatment for external dependencies; documentation, monitoring, and continual improvement	Dimension 2 - Clear allocation of responsibilities within and outside the organization; Dimension 1 - Bound risks upfront for agent tools	Provider obligations for quality management and oversight; transparency and post-market monitoring (as applicable)

Key Deliverables

Third-Party AI and Agent Tool Inventory (Template: Appendix L.1)
Supplier Due Diligence Checklist and Evidence Pack (Appendix L.2)
AI Supplier Contract Clause Library (Appendix L.3)
Shared Responsibility Matrix (Appendix L.4)
Exit and Transition Plan (Appendix L.5)

3.8 GOVERN 1.8: AI Risk Appetite, Risk Criteria, and Risk Acceptance

NEW IN AI RMF 2026: This subcategory formalizes organization-wide AI risk appetite and defines risk acceptance and exception processes for residual risk.

Objective: Define and operationalize AI risk appetite and risk criteria, link them to decision authority, and ensure residual risk acceptance and exceptions are formally approved, time-bounded, and reviewable.

Implementation Procedures

Procedure 1.8.1: Define AI Risk Appetite and Risk Criteria

Develop an AI Risk Appetite Statement covering key dimensions: safety impact severity, security exposure, privacy sensitivity, fairness/DEIA sensitivity, autonomy level, reversibility, financial impact, and regulatory classification (Template: Appendix N.1).

Define risk criteria and scoring scales aligned to the organization's AI risk taxonomy and agent risk scoring (Appendix I) to ensure consistent quantification.

Specify prohibited use cases and non-negotiable controls (e.g., no autonomous financial transfers above a defined threshold without human approval).

Approve the risk appetite through the AI Governance Committee and communicate to all AI development and procurement teams.

Procedure 1.8.2: Establish Residual Risk Acceptance and Exception Process

Require a Residual Risk Acceptance Form (Appendix N.3) for any deployment where residual risk remains above the target appetite but is justified by business need and mitigations.

Define who may accept residual risk by tier and impact, aligned to the Decision Authority Matrix (Appendix C).

Time-bound risk acceptances (e.g., 90/180/365 days) and require re-approval upon expiry or material change.

Maintain an AI Policy Exception Log (Appendix N.4) for deviations from mandatory controls and require compensating controls.

Procedure 1.8.3: Maintain Risk Acceptance Register and Review Cadence

Maintain a centralized register of accepted residual risks and exceptions, including: system/agent identifier, risk summary, acceptance rationale, sign-offs, expiry date, and monitoring conditions.

Review the register at least quarterly in the AI Governance Committee and escalate repeated exceptions or systemic issues.

Trigger immediate review upon: significant incidents, regulatory changes, model/tool changes, or drift in performance metrics.

Use register analytics to identify risk concentration, recurring control gaps, and portfolio-level exposure.

Procedure 1.8.4: Link Risk Appetite to Portfolio Governance and Decision Authority

Integrate risk appetite thresholds into project gating (design, pre-deployment review, and release) and procurement approvals.

Ensure the AI System and Agent Registry (Appendix M) records the risk appetite alignment status, residual risk acceptance decisions, and next review dates.

Report risk appetite alignment metrics (e.g., % of portfolio within appetite, number of exceptions by tier, average time to close exceptions) in quarterly oversight reports.

Procedure 1.8.5 Enforce GPAIS Go/No-Go Criteria Based on Risk Appetite

- Require that all major lifecycle transitions for GPAIS and foundation models (e.g., large-scale training start, external pilot, public release, model-weight release) include a documented Go/No-Go decision referencing:
 - current risk classification and impact assessment;
 - applicable GPAIS risk-tolerance thresholds and unacceptable-risk criteria;
 - results of capability evaluations, red-team and adversarial testing, and incident history.
 - Prohibit a Go decision wherever GPAIS risks exceed defined unacceptable-risk thresholds or where severe or catastrophic scenarios are identified without effective, implemented mitigations.
 - Require that any decision to accept high residual risk for a GPAIS:
 - is explicitly justified in the Residual Risk Acceptance Form Appendix N.3;
 - is approved by the AI Governance Committee (or Board-level body) at the authority level defined in Appendix C;
 - is time-bounded with clear review and sunset dates;
 - is reflected in the AI System and Agent Registry Appendix M and quarterly oversight reporting.
 - Align this procedure with Berkeley GPAIS expectations for Manage 1.1 (explicit Go/No-Go criteria) and Map 1.5 (risk-tolerance thresholds).

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Define AI risk appetite and criteria	Chief Risk Officer	Chief AI Officer	Legal Counsel; CISO; Data Protection Officer; Business Leaders	Board
Approve risk appetite and prohibited uses	AI Governance Committee	Chief AI Officer	Chief Risk Officer; Legal Counsel	Executive Team
Approve residual risk acceptance	System Owners/Agent Owners	Chief AI Officer	AI Risk Team; Legal Counsel; CISO	AI Governance Committee
Maintain risk acceptance register	AI Risk Team	Chief Risk Officer	Chief AI Officer; Internal Audit	Executive Team
Review risk acceptances and exceptions	AI Governance Committee	Chief AI Officer	Chief Risk Officer; Internal Audit	Board

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
6.1 - Actions to address risks and opportunities; 9.3 - Management review; 10 - Improvement	6.1 - Actions to address risks; 5.4 - Management responsibilities; continual improvement	Risk evaluation criteria, acceptance, treatment selection, and review; dynamic risk management	Dimension 1 - Assess and bound risks upfront; Dimension 2 - Clear accountability for risk acceptance	Risk management, quality management system, and documentation requirements; human oversight where applicable

Key Deliverables

- AI Risk Appetite Statement (Template: Appendix N.1)
- AI Risk Criteria Matrix (Template: Appendix N.2)
- Residual Risk Acceptance Form (Template: Appendix N.3)
- AI Policy Exception Log (Template: Appendix N.4)
- Quarterly Risk Acceptance Register Review Report

4. GOVERN 2: Accountability and Responsibility

Expected Outcome: Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible, and trained for mapping, measuring, and managing AI risks.

4.1 GOVERN 2.1: Accountability Structures

Objective: Establish clear accountability for AI system decisions, outcomes, and impacts throughout the lifecycle.

Implementation Procedures

Procedure 2.1.1: Define Decision Authority Framework

- Map AI lifecycle stages: Define all stages from concept through decommissioning (concept, design, development, testing, deployment, operation, monitoring, decommissioning)
- Identify decision points: Document key decisions at each stage (use case approval, data selection, model architecture, deployment approval, performance thresholds, incident response)
- Assign decision authority: Specify who has authority to make each type of decision based on risk level
- Create Decision Authority Matrix: Document decision type, authority level, required approvals, escalation path (Template: Appendix C)
- Communicate decision framework: Ensure all stakeholders understand decision authority and approval requirements
- Review and update annually: Adjust decision authority as organization and AI maturity evolve

Procedure 2.1.2 Implement Three Lines of Defense for GPAIS and Foundation Models

- Establish a formal three-lines-of-defense model for GPAIS and foundation model development and deployment:
 - First Line: Research and engineering teams responsible for model design, training, fine-tuning, evaluation, and documentation of safety and risk controls.
 - Second Line: Independent AI risk, compliance, legal, and technical safety functions responsible for challenging design and release decisions, setting risk-tolerance thresholds, and reviewing GPAIS risk assessments and Go/No-Go proposals.
 - Third Line: Internal audit (and, where applicable, external audit) responsible for independent assessment of GPAIS risk-management effectiveness and escalation to the Board or its committees.

- Define reporting lines so that:
 - the second line has direct access to the CEO and Board-level risk or AI committee and is structurally independent from product performance incentives;
 - the third line reports to the Board or the Board Audit Committee, not to first-line management.
 - Document the three-lines-of-defense structure in the AI Governance Organizational Chart and Role Assignment Matrix Appendix B and review annually for effectiveness.

Procedure 2.1.3: Establish Approval Requirements for High-Risk AI

- Define high-risk AI criteria: Systems impacting fundamental rights, safety, financial decisions, legal determinations, employment, critical infrastructure
- Specify approval levels: Board approval for highest risk, C-suite for high risk, department head for moderate risk
- Document approval workflows: Step-by-step process from initial proposal through final deployment approval
- Require multi-stakeholder review: Technical, legal, ethics, security, privacy reviews before approval
- Implement approval tracking: Document all approvals with timestamp, approver identity, rationale
- Establish re-approval triggers: Material changes to system, data, or deployment context require re-approval

Procedure 2.1.4: Create Escalation Paths

- Define escalation triggers: Situations requiring escalation (performance degradation, security incidents, bias detection, regulatory inquiry, stakeholder complaints)
- Establish escalation levels: Tier 1 (Team Lead), Tier 2 (Department Head), Tier 3 (C-Suite), Tier 4 (Board)
- Document escalation procedures: Who to notify, within what timeframe, what information to include
- Create escalation contact list: Maintain current contact information for all escalation points
- Test escalation procedures: Conduct annual tabletop exercises to verify effectiveness
- Track escalations: Log all escalations, resolutions, and time to resolution for continuous improvement

Procedure 2.1.5: Implement Audit Trails

- Identify auditable events: All key AI decisions, configuration changes, data access, model updates, deployment changes

- Implement logging infrastructure: Automated capture of auditable events with timestamp, user identity, action taken
- Ensure tamper-proof storage: Audit logs stored in immutable, append-only systems
- Define retention periods: Minimum 7 years for high-risk AI systems, aligned with regulatory requirements
- Enable audit trail review: Searchable, filterable interface for audit log analysis
- Conduct regular audits: Quarterly review of audit trails by internal audit or risk management

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Define decision authority framework	Chief AI Officer	CEO	Legal, Risk, Business Units	All Stakeholders
Establish approval requirements	Chief Risk Officer	Chief AI Officer	Legal, Ethics, Security	Business Units
Create escalation paths	Chief Risk Officer	Chief AI Officer	IT, Security, Communications	All Employees
Implement audit trails	CISO	Chief AI Officer	IT Infrastructure, Legal	Audit Committee

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
5.3 - Organizational roles and authorities; 8.1 - Operational planning and control	5.3 - Organizational roles	5.3 - Assign roles and responsibilities (clear decision authority)	Dimension 2 - Make humans meaningfully accountable (clear allocation of responsibilities)	Art. 16 - Obligations of providers (must establish accountability structures)

Key Deliverables

- Decision Authority Matrix (Appendix C - updated with GOVERN 2.1 content)
- High-Risk AI Approval Workflow Documentation
- Escalation Procedures and Contact List
- Audit Trail Implementation Specification

5. GOVERN 3: Workforce Diversity and Team Composition

Expected Outcome: AI design, development, and deployment teams are diverse and include subject matter experts from across disciplines to reduce bias and improve outcomes.

5.1 GOVERN 3.1: Diverse and Multidisciplinary AI Teams

Objective: Build multidisciplinary teams with diverse backgrounds, expertise, and perspectives to develop more robust, fair, and effective AI systems.

Team Composition Requirements

Effective AI teams must include diverse representation across multiple dimensions:

Discipline	Contribution to AI Team
Technical Experts	AI/ML engineers, data scientists, software developers - implement models, algorithms, infrastructure
Domain Experts	Subject matter specialists for application area - ensure AI solves real problems correctly
Social Scientists	Ethicists, sociologists, psychologists - identify social impacts, fairness concerns, ethical issues
Legal and Compliance	Attorneys, regulatory specialists - ensure legal compliance, manage regulatory risks
User Representatives	End users and affected stakeholders - provide real-world perspective, identify usability issues
Security Professionals	Cybersecurity experts - identify vulnerabilities, implement security controls
Privacy Experts	Data protection specialists - ensure privacy compliance, minimize data exposure

Implementation Procedures

Procedure 3.1.1: Establish Multidisciplinary Team Requirements

- Define minimum team composition: Specify required roles and expertise for AI projects based on risk level and complexity
- Document expertise requirements: Technical skills (ML, software engineering, data science), domain knowledge, social science, legal, ethics, user research
- Mandate cross-functional collaboration: Require representatives from multiple functions on all AI project teams
- Create team formation guidelines: Step-by-step process for assembling AI project teams
- Establish team chartering process: Document team composition, roles, decision-making authority
- Monitor compliance: Track whether AI teams meet composition requirements

Procedure 3.1.2: Promote Diverse Perspectives

- Recruit for diversity: Proactively recruit team members from underrepresented groups in AI
- Ensure inclusive team culture: Training on inclusive collaboration, psychological safety, respectful communication
- Rotate team membership: Bring fresh perspectives by rotating team members periodically
- Include affected stakeholders: Involve representatives of populations affected by AI system in design and review
- Seek external perspectives: Engage external advisors, consultants, or review boards for critical systems
- Document how diversity influenced decisions: Capture how diverse perspectives shaped design choices

Procedure 3.1.3: Foster Cross-Functional Collaboration

- Establish collaboration norms: Define how team members from different disciplines work together effectively
- Create shared understanding: Ensure technical and non-technical team members develop mutual understanding of AI system
- Implement collaborative design reviews: Structured sessions where all disciplines contribute to design decisions
- Use collaborative tools: Shared documentation, version control, communication platforms accessible to all team members
- Resolve conflicts constructively: Process for addressing disagreements between disciplines
- Celebrate collaborative successes: Recognize teams that exemplify effective cross-functional collaboration

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Establish team composition requirements	Chief AI Officer	CEO	HR, Department Heads	All Teams
Recruit diverse team members	HR	Department Heads	DEIA Officer, Hiring Managers	Chief AI Officer
Foster cross-functional collaboration	Project Managers	Project Sponsors	Team Leads	All Team Members
Monitor team composition compliance	HR Analytics	Chief AI Officer	DEIA Officer	Executive Team

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
7.2 - Competence (diverse expertise requirements)	7.2 - Competence	6.1 - AI system context (multidisciplinary teams reduce risks from narrow perspectives)	Dimension 1 - Assess and bound risks (diverse teams identify broader range of risks)	Art. 10 - Data and data governance (diverse teams ensure representative data)

Key Deliverables

- Multidisciplinary Team Composition Guidelines
- Team Formation Checklist
- Cross-Functional Collaboration Best Practices
- Team Diversity Metrics and Reporting

6. GOVERN 4: Organizational Culture and AI Risk Management

Expected Outcome: Organizational culture supports addressing AI risks effectively through comprehensive training, awareness, and cultural reinforcement.

6.1 GOVERN 4.1: Training and Awareness Programs

Objective: Ensure all AI stakeholders have appropriate training and awareness of AI risks, responsible AI principles, and their specific responsibilities in AI risk management.

Training Program Framework

Comprehensive training program addressing multiple audiences and competency levels:

Module	Audience	Duration	Key Topics
AI Fundamentals	All Employees	2 hours	What is AI, types of AI, AI in our organization, basic terminology
Responsible AI Principles	All Employees	2 hours	Trustworthy AI characteristics, fairness, transparency, accountability, privacy
AI Risk Management	AI Practitioners	8 hours	Risk identification, assessment, mitigation, monitoring; GOVERN, MAP, MEASURE, MANAGE functions
Bias and Fairness	Data Scientists, ML Engineers	4 hours	Types of bias, detection methods, mitigation techniques, fairness metrics, testing
AI Security	Engineers, Security Team	4 hours	Adversarial attacks, prompt injection, data poisoning, model extraction, defenses
AI Privacy	Data Scientists, Privacy Team	4 hours	Privacy regulations (GDPR, CCPA), data minimization, de-identification, privacy-enhancing technologies
AI Governance	Executives, Managers	3 hours	Governance structures, policies, oversight, accountability, compliance
Agentic AI Governance	AI Teams, Managers	4 hours	Agent characteristics, risks, boundaries, identity management, oversight, automation bias
Environmental Sustainability	AI Teams, Sustainability Team	2 hours	AI carbon footprint, energy monitoring, resource optimization, sustainability reporting

Implementation Procedures

Procedure 4.1.1: Develop Training Curriculum

- Conduct training needs assessment: Identify competency gaps across different roles and levels
- Define learning objectives: Specific, measurable outcomes for each training module
- Develop training content: Create or procure training materials, exercises, case studies, assessments
- Design role-specific pathways: Customize training sequences for different roles (executives, engineers, business users)
- Establish delivery methods: Combine online self-paced, instructor-led, hands-on workshops, certification programs
- Create training calendar: Schedule regular training sessions, mandatory completion deadlines
- Document training curriculum: Complete description of all training modules, prerequisites, duration (Appendix D - enhanced)

Procedure 4.1.2: Implement Training Programs

- Launch foundational training: All employees complete AI fundamentals and responsible AI principles training
- Deploy role-specific training: Technical teams complete advanced modules on bias mitigation, security, testing
- Conduct executive briefings: Board and C-suite receive strategic briefings on AI governance, risks, opportunities
- Provide just-in-time training: Contextual training when employees begin working with new AI systems
- Offer continuous learning: Regular updates on emerging AI risks, techniques, regulations
- Track training completion: Monitor completion rates, time to completion, assessment scores
- Enforce mandatory training: Link training completion to system access, project participation, performance reviews

Procedure 4.1.3: Assess Training Effectiveness

- Conduct knowledge assessments: Pre- and post-training tests to measure learning outcomes
- Gather participant feedback: Surveys and focus groups to evaluate training quality and relevance
- Measure behavioral change: Observe whether training leads to improved AI risk management practices
- Track incident metrics: Monitor whether training correlates with reduced AI incidents
- Benchmark against industry: Compare training program maturity to industry best practices

- Conduct annual training review: Comprehensive evaluation of training program effectiveness
- Continuously improve: Update training content, delivery methods, and curriculum based on assessment results

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Develop training curriculum	Learning & Development	Chief AI Officer	Subject Matter Experts, HR	All Employees
Implement training programs	Learning & Development	Chief AI Officer	Department Heads	All Employees
Assess training effectiveness	Learning & Development	Chief AI Officer	Internal Audit, HR Analytics	Executive Team
Complete required training	Individual Employees	Manager	Learning & Development	HR
Ensure multi-agent teams have diverse expertise across vendors/domains (SINGAPORE)	HR	Chief AI Officer	DEIA Officer, Department Heads	Executive Team

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
7.2 - Competence; 7.3 - Awareness	7.2 - Competence; A.6.3 - Awareness training	7.2 - Competence requirements (comprehensive training on AI risks)	Dimension 4 - Enable end-user responsibility (training equips users for effective oversight)	Art. 4 - Definitions (users of high-risk AI must have appropriate training)

Key Deliverables

- AI Risk Management Training Curriculum (Appendix D - enhanced with GOVERN 4.1)
- Training Needs Assessment Report
- Training Calendar and Schedule
- Training Effectiveness Assessment Reports

7. GOVERN 5: Oversight and Monitoring

Expected Outcome: Organizational oversight and monitoring processes ensure AI systems are developed and deployed in accordance with policies and risk management practices.

7.1 GOVERN 5.1: Governance Bodies and Oversight Mechanisms

Objective: Establish oversight bodies with appropriate authority, expertise, and independence to ensure AI governance effectiveness.

Key Oversight Mechanisms

Effective AI governance requires multiple oversight mechanisms operating at different organizational levels:

Oversight Mechanism	Purpose and Scope
AI Governance Committee	Cross-functional executive oversight body. Approves policies, reviews high-risk systems, allocates resources, monitors governance effectiveness.
AI Ethics Review Board	Independent ethical assessment of high-risk and socially impactful AI systems. Reviews alignment with values, fairness, potential for harm.
Internal Audit	Independent assessment of AI management system effectiveness, policy compliance, control adequacy. Reports to Audit Committee.
Management Review	Quarterly executive evaluation of AI governance performance. Reviews metrics, incidents, audit findings, makes strategic decisions.
External Audit	Third-party certification audit for ISO 42001, ISO 27001, or industry standards. Provides independent validation of compliance.
Agentic AI Committee	Specialized oversight for autonomous AI systems. Reviews agent risks, approves high-risk agent deployments, monitors agent incidents (GOVERN 1.5).
Environmental Sustainability Committee	Oversight of AI environmental impact. Sets sustainability targets, monitors carbon footprint, reviews energy efficiency (GOVERN 1.6).
Governor and Auditor Agent Patterns (Singapore Framework)	Defines governor agents and auditor agents, implementation requirements

Human Oversight Patterns for Agentic and Multi-Agent Systems

For agentic AI systems and multi-agent ecosystems, GOVERN MUST define and document explicit human oversight patterns that are proportionate to the agent's autonomy, authority and environment complexity. At minimum, the organization SHOULD standardize the following patterns and criteria for their use:

- Human-in-the-loop (HITL): Agents may propose actions, but execution requires explicit human review and approval. This pattern is REQUIRED for high-stakes decisions (e.g. affecting safety, fundamental rights, or high-value transactions) and for agentic systems classified in the highest risk tiers.
- Human-on-the-loop (HOTL): Agents may execute actions autonomously within defined guardrails, while humans monitor behavior and can intervene or override. This pattern MAY be used for lower-tier agentic systems in well-defined environments, provided effective monitoring, alerts and emergency stop mechanisms are in place.
- Human-in-command: Humans set objectives, constraints and risk appetite for one or more agents or multi-agent ecosystems and retain authority to pause, reconfigure or decommission systems. This pattern SHOULD be applied to all multi-agent ecosystems and to any governor/auditor agents with broad visibility or control.

Oversight patterns, and the conditions under which they apply, SHALL be documented in governance policies, Agent Profile Cards and System-Level Agentic Profiles. Changes to oversight patterns for high-tier agents or multi-agent ecosystems require AI Governance Committee review and approval.

GPAIS and Open-Weights Decommissioning Constraints

- Recognize that GPAIS and foundation models whose parameter weights are released as downloadable, fully open, or open-source artefacts cannot be fully decommissioned once widely copied or integrated by third parties.
- For any GPAIS where open-weights release is contemplated, shift the primary risk-management burden to pre-release stages by requiring:
 - staged release (e.g., internal use → restricted access → API-only → potential open-weights), with intensified monitoring and red-teaming between stages;
 - explicit AI Governance Committee approval before any expansion in access level, referencing GPAIS risk-tolerance thresholds and unacceptable-risk criteria;
 - a documented rationale that explains why residual risks are deemed acceptable given the irreversibility of open-weights release.
 - Require that decommissioning plans for GPAIS describe what can and cannot be achieved for open-weights models (e.g., revoking first-party hosting and API

access, but not copies already held by third parties) and define compensating measures (e.g., disclosure to regulators, updating documentation, support for downstream mitigations).

- Apply these constraints explicitly in vendor and third-party risk management for external GPAIS used by the organization (see Manage 3.2 alignment in Berkeley GPAIS Profile).

Implementation Procedures

Procedure 5.1.1: Establish AI Governance Committee

- Define committee charter: Purpose, scope, authority, membership, meeting frequency, reporting relationships
- Appoint committee members: Cross-functional representation including executives, technical leaders, risk management, legal, ethics, business units
- Establish meeting schedule: Monthly meetings minimum, special sessions for urgent matters
- Define decision-making authority: Approval requirements for policies, high-risk AI systems, resource allocation, strategic initiatives
- Create reporting mechanism: Regular reports to Board of Directors and CEO on AI governance status
- Document committee procedures: Meeting protocols, decision processes, voting requirements, conflict resolution
- Charter template available in Appendix E

Procedure 5.1.2: Establish AI Ethics Review Board

- Define ethics review scope: High-risk AI systems, systems with significant social impact, novel AI applications
- Appoint ethics board members: Ethicists, social scientists, community representatives, legal experts, technical advisors
- Establish review process: Application submission, review criteria, deliberation procedures, approval or rejection with rationale
- Create ethics review criteria: Alignment with organizational values, societal benefit, fairness, transparency, potential for harm
- Document review decisions: Maintain record of all ethics reviews including recommendations and conditions
- Provide appeal process: Mechanism for project teams to appeal ethics board decisions
- Conduct annual board effectiveness review: Assess whether ethics reviews are improving AI outcomes

Procedure 5.1.3: Implement Internal Audit Program

- Develop audit plan: Risk-based approach identifying AI systems and processes for audit
- Define audit scope: AI management system effectiveness, compliance with policies, adherence to procedures, control effectiveness
- Assign audit team: Independent auditors with AI, risk, and technology expertise
- Conduct audits: Document review, interviews, system testing, evidence collection
- Report audit findings: Written reports to executive management and AI Governance Committee detailing findings, non-conformities, recommendations
- Track remediation: Monitor corrective actions for audit findings through to closure
- Audit frequency: Annual comprehensive audit minimum, quarterly audits of high-risk systems

Procedure 5.1.4: Conduct Management Reviews

- Schedule management reviews: Quarterly reviews by executive leadership minimum
- Prepare review inputs: AI governance performance metrics, audit results, incident reports, stakeholder feedback, regulatory changes, resource needs
- Conduct review meeting: Executive team evaluates AI governance effectiveness, identifies improvement opportunities, makes strategic decisions
- Document review outputs: Decisions on policy updates, resource allocation, strategic direction, improvement initiatives
- Communicate decisions: Disseminate management review decisions to relevant stakeholders
- Track action items: Monitor implementation of management review decisions
- Maintain review records: Permanent record of all management reviews for compliance and historical reference

Procedure 5.1.5: Engage External Auditors

- Select certification body: Choose accredited third-party auditor for ISO 42001, ISO 27001, or other relevant standards
- Prepare for external audit: Internal readiness assessment, documentation review, evidence compilation
- Coordinate audit execution: Provide access to systems, personnel, documentation during external audit
- Address audit findings: Develop corrective action plans for any non-conformities identified
- Maintain certification: Annual surveillance audits, recertification every three years
- Leverage audit insights: Use external audit recommendations to improve AI governance
- Consider multiple certifications: ISO 42001 (AI management), ISO 27001 (information security), industry-specific certifications

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Establish AI Governance Committee	Chief AI Officer	CEO	Board, Executives	All Stakeholders
Establish AI Ethics Review Board	Chief Ethics Officer	Chief AI Officer	Board, Legal, Executives	Organization
Conduct internal audits	Internal Audit Team	Chief Audit Executive	Audit Committee, Management	Executives
Conduct management reviews	Chief AI Officer	CEO	Executive Team	Board, Organization
Engage external auditors	Chief AI Officer	CEO	Finance, Legal	Board, Stakeholders

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
9.2 - Internal audit; 9.3 - Management review	9.2 - Internal audit; 9.3 - Management review	5.1 - Establish framework (oversight bodies are part of governance framework)	Dimension 2 - Make humans meaningfully accountable (oversight ensures accountability)	Art. 17 - Quality management system (requires oversight and monitoring)

Key Deliverables

- AI Governance Committee Charter (Appendix E)
- AI Ethics Review Board Charter
- Internal Audit Plan and Schedule
- Management Review Records
- External Audit Reports and Certificates

7.2 GOVERN 5.4: AI System & Agent Registry and Portfolio Monitoring

Objective: Maintain an authoritative portfolio-level registry of AI systems and agents, enable traceability of ownership and approvals, and support continuous monitoring and governance reporting.

Implementation Procedures

Procedure 5.4.1: Establish and Maintain the AI System & Agent Registry

Create and maintain the AI System & Agent Registry using the template in Appendix M. Assign a unique Registry ID for each AI system and each agent (including agent versions) and link to documentation (model card, impact assessment, risk assessment, approvals, monitoring plan).

Record key governance attributes: owner, purpose, deployment environment, autonomy level, tool permissions/action-space, data categories, supplier dependencies, and risk tier.

Update the registry upon material changes (new model version, new tools, new data sources, new user populations, supplier changes, or scope expansions).

Procedure 5.4.2: Portfolio Monitoring and Governance Reporting

Define portfolio KPIs and KRIs, including at minimum: incident rate, exception count, drift alerts, fairness metric deviations, privacy/security events, and supplier change events.

Implement a governance reporting cadence: monthly operational reporting and quarterly executive oversight reporting.

Use registry analytics to identify high-concentration risks (e.g., many high-autonomy agents in one business unit) and prioritize audit and assurance activities.

Escalate portfolio-level risks to the AI Governance Committee and Board as defined in the Decision Authority Matrix.

Procedure 5.4.3: Trigger-Based Reviews

Initiate an out-of-cycle review when triggers occur: material performance degradation, new high-severity incident, significant model/tool update, new regulatory requirement, or new third-party dependency.

Re-run relevant assessments (Integrated AI Impact Assessment, security testing, bias testing) and refresh approvals if triggers materially alter risk posture.

Document trigger outcomes and update registry fields for review date, decision, and follow-up actions.

Procedure 5.4.4: Decommissioning and Record Retention

Record decommission decisions and dates in the registry, including data retention/deletion steps and credential revocation for agent tools.

Retain governance records (approvals, assessments, logs) per the retention policy and regulatory requirements.

Perform a decommission post-mortem for high-risk systems to capture lessons learned.

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Maintain AI system and agent registry	AI Risk Team	Chief AI Officer	System Owners; Agent Owners; CISO	AI Governance Committee
Update registry for material changes	System Owners/Agent Owners	Chief AI Officer	AI Risk Team; Procurement Lead	AI Governance Committee
Generate portfolio governance reports	AI Risk Team	Chief Risk Officer	Chief AI Officer; Internal Audit	Executive Team
Perform trigger-based reviews	AI Risk Team	Chief AI Officer	System Owners; Legal Counsel; CISO	AI Governance Committee
Approve decommission and retention actions	System Owners	Chief AI Officer	Legal Counsel; CISO	AI Risk Team

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
7.5 - Documented information; 8.1 - Operational control; 9.1 - Monitoring, measurement, analysis and evaluation	5.1-5.4 - Governance; 8.12 - Data leakage prevention (as applicable); incident management controls	Monitoring and review; documentation and traceability across lifecycle	Dimension 3 - Adapt and monitor; operationalize traceability for agentic systems	Post-market monitoring and documentation requirements (as applicable)

Key Deliverables

- AI System & Agent Registry (Template: Appendix M)
- Portfolio Governance Dashboard and KPI/KRI Definitions
- Quarterly Portfolio Oversight Report
- Trigger-Based Review Log

7.3 GOVERN 5.5: End-User Enablement, Feedback, Complaints, and Redress

Objective: Equip end users and overseers with clear information, operating guidance, and reporting channels for AI systems and agents, and provide a defined complaints and redress process.

Implementation Procedures

Procedure 5.5.1: End-User Disclosure and Usage Guidance

- Provide standardized end-user disclosures for AI systems and agents using Appendix P.1, including: intended use, known limitations, confidence/uncertainty cues (where applicable), prohibited uses, and escalation pathways.
- Publish an Operator Quick Reference for overseers and frontline staff (Appendix P.2) describing how to supervise outputs, when to override/stop actions, and how to report issues.
- For agentic systems, disclose action authority boundaries: what the agent can access, what it can change, and what approvals are required.

Procedure 5.5.2: Feedback and Complaint Intake Channels

- Establish feedback and complaint channels appropriate to the user population (e.g., in-product reporting, help desk category, hotline, email intake).
- Use the Complaint/Feedback Intake Form (Appendix P.3) to capture minimum required information and ensure traceability to the system/agent registry entry.
- Define SLAs for acknowledgement and response based on severity and impact.

Procedure 5.5.3: Triage, Investigation, and Resolution

- Triage incoming reports within defined SLA and classify severity (Low/Moderate/High/Critical) based on safety, security, privacy, and financial impact.
- Investigate using relevant logs and evidence; for agentic actions, include tool-call traces and approval logs.
- Determine corrective and preventive actions (CAPA) and record outcomes in the Redress Log (Appendix P.4).
- Escalate High/Critical cases to the AI Governance Committee and invoke incident response where applicable.

Procedure 5.5.4: Redress, Remediation, and Communication

- Provide redress appropriate to harm type: correction of records, reversal of decisions where possible, compensating actions, and communication of resolution steps.
- Where outcomes cannot be reversed, provide an explanation of the decision path and mitigation steps to prevent recurrence.

- Document communications and approvals for sensitive cases; coordinate with Legal Counsel and Privacy/Compliance roles as needed.

Procedure 5.5.5: Metrics and Continual Improvement

- Track metrics: complaint volume, time-to-acknowledge, time-to-resolve, recurrence rate, and systemic issue categories.
- Review trends at least quarterly and update training, controls, and disclosures accordingly.
- Feed lessons learned into management review and improvement planning.

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Publish end-user disclosures and operator guidance	System Owners/Agent Owners	Chief AI Officer	Legal Counsel; Communications; Data Protection Officer	AI Governance Committee
Manage complaint intake and triage	Customer Support/Help Desk Lead	Chief Risk Officer	AI Risk Team; System Owners	Chief AI Officer
Investigate and resolve complaints	AI Risk Team	System Owners	CISO; Legal Counsel; Data Protection Officer	Chief Risk Officer
Approve redress actions for High/Critical cases	AI Governance Committee	Chief AI Officer	Legal Counsel; Chief Risk Officer	Board
Monitor metrics and drive improvements	AI Risk Team	Chief Risk Officer	Training Lead; System Owners	Executive Team

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
7.4 - Communication; 8.1 - Operational control; 9.1 - Monitoring and evaluation; 10 - Improvement	5.24-5.27 - Incident management; 5.20 - Information security in supplier relationships (as applicable)	Risk communication and stakeholder engagement; monitoring, review, and improvement	Dimension 4 - Enable end-user responsibility; Dimension 3 - Monitor and adapt	Transparency and information to users; human oversight; post-market monitoring (as applicable)

Key Deliverables

- End-User Disclosure Template (Appendix P.1)
- Operator Quick Reference (Appendix P.2)
- Complaint/Feedback Intake Form (Appendix P.3)
- Redress Log and CAPA Tracking (Appendix P.4)
- Complaint Triage Workflow (Appendix P.5)

8. GOVERN 6: Trustworthy AI Characteristics

Expected Outcome: Policies and procedures are in place to address trustworthy AI characteristics including validity, reliability, safety, security, resilience, accountability, transparency, explainability, interpretability, privacy enhancement, and fairness with harmful bias managed.

8.1 GOVERN 6.1: Trustworthy AI Policy

Objective: Establish comprehensive policies addressing all trustworthiness characteristics to ensure AI systems are developed and deployed responsibly.

Seven Trustworthiness Characteristics

The AI RMF 2026 defines seven characteristics that trustworthy AI systems must embody:

Characteristic	Description and Requirements
1. Valid and Reliable	Systems perform consistently and accurately for intended purposes. Testing and validation procedures ensure fitness for use. Performance metrics meet defined thresholds.
2. Safe	Systems do not pose unacceptable risk of harm to people, property, or environment. Safety mechanisms prevent dangerous failures. Fail-safe behaviors protect against worst-case scenarios.
3. Secure and Resilient	Protected against unauthorized access, adversarial attacks, and malicious manipulation. Can recover from failures and continue operating. Implements defense-in-depth security controls.
4. Accountable and Transparent	Clear responsibility for system decisions and outcomes. Appropriate information about capabilities, limitations, and operation available to users, regulators, and affected parties. Audit trails enable traceability.
5. Explainable and Interpretable	Meaningful explanations of system behavior and decisions available. Users and stakeholders can understand how outputs were generated. Level of explanation appropriate to audience and use case.
6. Privacy-Enhanced	Protects individual privacy and data rights. Implements privacy-by-design principles. Uses privacy-enhancing technologies. Complies with data protection regulations (GDPR, CCPA).
7. Fair - with Harmful Bias Managed	Does not systematically disadvantage individuals or groups based on protected characteristics. Bias identification, assessment, and mitigation throughout lifecycle. Ongoing fairness monitoring in production.

- For GPAIS, foundation models, and other highly capable general-purpose systems, transparency artefacts (e.g., model cards, system cards, system-level Agentic Profiles) must at minimum document:
 - intended purposes, out-of-scope uses, and explicitly disallowed use cases;
 - reasonably foreseeable misuses and abuses (e.g., disinformation, targeted manipulation, assistance to cyber/CBRN threats) and associated impact scenarios;
 - known dangerous capabilities, failure modes, and knowledge limits, including how they were identified (e.g., red-team campaigns, adversarial testing);
 - applicable risk-tolerance thresholds and unacceptable-risk criteria, and whether they are currently met;
 - major Go/No-Go and risk-acceptance decisions taken for the GPAIS, including high-level rationale and compensating controls.

Where public disclosure of specific technical or security details would materially increase misuse risk, sensitive details may be restricted to auditors and competent authorities, but high-level risk information must still be communicated to external stakeholders

Implementation Procedures

Procedure 6.1.1: Develop Trustworthy AI Policy

- Define policy scope: All AI systems across organization, all lifecycle stages, all trustworthiness characteristics
- Address each characteristic: Specific policy statements and requirements for validity, reliability, safety, security, resilience, accountability, transparency, explainability, interpretability, privacy, fairness
- Establish minimum standards: Baseline requirements that all AI systems must meet
- Define risk-based requirements: Additional requirements for high-risk AI systems
- Include agentic AI provisions: Specific requirements for autonomous AI systems per GOVERN 1.5
- Include sustainability provisions: Environmental requirements per GOVERN 1.6
- Document policy exceptions process: How to request and approve deviations from policy
- Obtain executive approval: CEO and Board approval of Trustworthy AI Policy
- Policy template structure provided in Appendix E

Procedure 6.1.2: Implement Validity and Reliability Requirements

- Define performance requirements: Accuracy, precision, recall, F1 score, or other metrics appropriate to system purpose
- Establish testing protocols: Validation testing with representative data, test-retest reliability, cross-validation
- Set performance thresholds: Minimum acceptable performance levels for deployment
- Require performance monitoring: Ongoing tracking of system performance in production
- Define performance degradation triggers: Conditions that require system review or deactivation
- Document validation evidence: Maintain records of validation testing and results
- Implement continuous validation: Regular revalidation to ensure continued fitness for purpose

Procedure 6.1.3: Implement Safety Requirements

- Conduct safety risk assessment: Identify potential harms from system failures or misuse
- Implement safety controls: Technical and procedural controls to prevent harm
- Define fail-safe mechanisms: How system should behave when errors occur
- Establish testing requirements: Safety testing including edge cases and failure modes
- Create incident response procedures: How to respond when safety issues arise
- Document safety analysis: Safety risk assessment and mitigation documentation
- Implement ongoing safety monitoring: Continuous monitoring for safety-related incidents

Procedure 6.1.4: Implement Security and Resilience Requirements

- Apply security-by-design: Security considerations throughout system development lifecycle
- Implement security controls: Access controls, encryption, secure communication, input validation, output encoding
- Address AI-specific threats: Adversarial attacks, data poisoning, model extraction, prompt injection
- Implement resilience mechanisms: Redundancy, graceful degradation, recovery procedures
- Conduct security testing: Penetration testing, adversarial testing, security code review
- Establish incident response: Security incident detection, containment, remediation, recovery
- Maintain security documentation: Security architecture, threat model, security testing results

Procedure 6.1.5: Implement Accountability and Transparency Requirements

- Establish clear accountability: Document who is responsible for each AI system per GOVERN 2.1
- Implement audit logging: Comprehensive logging of system decisions and actions
- Create system documentation: Purpose, capabilities, limitations, intended use, data sources, performance characteristics
- Provide transparency information: Make appropriate information available to users and stakeholders
- Implement model cards: Standardized documentation of model characteristics and performance
- Define transparency levels: Different information disclosure for different audiences (users, regulators, researchers)
- Maintain decision records: Ability to trace and explain specific system outputs

Procedure 6.1.6: Implement Explainability and Interpretability Requirements

- Define explainability requirements: Level of explanation needed based on system risk and impact
- Select interpretable techniques: Choose inherently interpretable models when appropriate (decision trees, linear models, rule-based systems)
- Implement explanation methods: Post-hoc explanation techniques (LIME, SHAP, attention visualization) when complex models required
- Provide user-appropriate explanations: Tailor explanations to audience (technical users, end users, regulators)
- Test explanation quality: Verify explanations are accurate, useful, understandable
- Document model architecture: Technical documentation of model structure and decision logic
- Train users on interpretation: Help users understand system outputs and limitations

Procedure 6.1.7: Implement Privacy Enhancement Requirements

- Apply privacy-by-design: Privacy considerations throughout development lifecycle
- Implement data minimization: Collect and use only necessary data
- Apply de-identification: Anonymization, pseudonymization where appropriate
- Implement privacy-enhancing technologies: Differential privacy, federated learning, homomorphic encryption, secure multi-party computation
- Establish data retention limits: Delete data when no longer needed
- Conduct privacy impact assessments: Evaluate privacy risks for new AI systems
- Ensure regulatory compliance: GDPR, CCPA, sector-specific privacy regulations

Procedure 6.1.8: Implement Fairness Requirements and Bias Management

- Define fairness requirements: Appropriate fairness criteria for system purpose (demographic parity, equalized odds, equal opportunity, individual fairness)

- Assess training data for bias: Evaluate data representativeness, identify potential sources of bias
- Test for bias: Measure fairness metrics across protected groups and relevant demographics
- Implement bias mitigation: Pre-processing (data balancing, reweighting), in-processing (fairness constraints), post-processing (threshold optimization)
- Conduct disparate impact analysis: Evaluate whether system has disproportionate impact on protected groups
- Establish bias monitoring: Ongoing tracking of fairness metrics in production
- Create bias incident response: Procedures when bias is detected in deployed systems
- Document fairness analysis: Record bias assessment, mitigation approaches, residual limitations

Roles & Responsibilities (RACI Matrix)

Activity	Responsible	Accountable	Consulted	Informed
Develop Trustworthy AI Policy	Chief AI Officer	CEO	Legal, Ethics, Risk, Technical Leads	Board, Organization
Implement validity and reliability	AI Engineering	System Owner	QA, Data Science	Chief AI Officer
Implement safety requirements	AI Engineering, Safety Team	System Owner	Risk Management, Legal	Chief AI Officer
Implement security and resilience	CISO, AI Engineering	System Owner	Security Team, Risk	Chief AI Officer
Implement accountability and transparency	System Owner	Chief AI Officer	Legal, Communications	Stakeholders
Implement explainability	Data Science, ML Engineering	System Owner	UX, Product	Users
Implement privacy enhancement	Data Protection Officer	System Owner	Legal, Security, Engineering	Chief Privacy Officer
Implement fairness and bias management	AI Fairness Team	System Owner	Ethics, Legal, Domain Experts	Chief AI Officer

Control Mappings

ISO 42001	ISO 27001	ISO 23894	Singapore MGF	EU AI Act
5.2 - AI management system policy (must address trustworthiness)	5.2 - Policy for information security	5.2 - Define risk management policy (trustworthiness characteristics reduce risks)	All Dimensions - Trustworthiness is foundational to all four Singapore MGF dimensions	Art. 9 - Risk management system; Art. 13 - Transparency; Art. 14 - Human oversight

Key Deliverables

- Trustworthy AI Policy (comprehensive) - Appendix E
- Validity and Reliability Testing Procedures
- Safety Risk Assessment and Controls Documentation
- Security Architecture and Testing Reports
- Transparency and Accountability Documentation (Model Cards)
- Explainability Implementation Guidelines
- Privacy Impact Assessments and PETs Implementation
- Fairness Testing Reports and Bias Mitigation Documentation

Appendices

This section provides templates and reference materials to support implementation of the GOVERN function procedures.

- Appendix A: AI Compliance Register Template
- Appendix B: RACI Matrix Template - Complete GOVERN Function
- Appendix C: Decision Authority Matrix
- Appendix D: AI Risk Management Training Curriculum
- Appendix E: Governance Committee Charters and Policy Templates
- Appendices F-K: Agentic AI and Environmental Sustainability
- Appendix G: Environmental Sustainability Committee Charter Template
- Appendix H: Agent Identity Management Framework Template
- Appendix I: Agent Risk Assessment Template
- Appendix J: Carbon Footprint Assessment Template
- Appendix K: Sustainability Metrics Tracking Template
- Appendix L: Third-Party & Tool Ecosystem Governance Toolkit
- Appendix M: AI System & Agent Registry Template
- Appendix N: AI Risk Appetite and Risk Acceptance Templates
- Appendix O: Integrated AI Impact Assessment (AIIA) Template
- Appendix P: End-User Enablement, Complaint Intake, and Redress Toolkit
- Appendix Q: Framework Compliance Matrix
- Appendix R: Implementation Roadmap
- Appendix S: Industry-Specific Adaptations
- Appendix T: AI Governance KPI Dashboard
- Appendix U: AI Governance Case Studies
- Appendix V: Agent Identity and Credential Management Protocol

Appendix A: AI Compliance Register Template

Purpose: Track all applicable legal and regulatory requirements for AI systems. Maintain as living document with quarterly reviews.

Instructions:

- Add new row for each identified legal or regulatory requirement
- Assign owner responsible for ensuring compliance with each requirement
- Update status regularly (Compliant, In Progress, Non-Compliant, Not Applicable)
- Document evidence of compliance in Evidence/Notes column
- Review quarterly and when regulations change

Template:

Requirement ID	Regulation / Law	Description	Owner	Status	Evidence / Notes
REQ-001	EU AI Act Art. 9	Risk management system	Chief Risk Officer	Compliant	Risk management framework documented, reviewed Q1 2026
REQ-002	EU AI Act Art. 10	Data governance	Data Protection Officer	In Progress	Data governance policy draft under review
REQ-003	ISO/IEC 42001:2023	AI management system	Chief AI Officer	Compliant	Certified December 2025, surveillance audit Q4 2026
[Add rows as needed]					

Appendix B: RACI Matrix Template - Complete GOVERN Function

Purpose: Define roles and responsibilities for all GOVERN function activities across the organization.

RACI Definitions:

- R - Responsible: Person(s) who perform the work to complete the task
- A - Accountable: Person who is ultimately answerable for completion (only one A per activity)
- C - Consulted: People who provide input and must be consulted before decision or action
- I - Informed: People who must be informed after decision or action is taken

Complete GOVERN Function RACI Matrix:

GOVERN Activity	Responsible	Accountable	Consulted	Informed
GOVERN 1.1: Maintain compliance register	CRO	Legal	CAIO	Board
GOVERN 1.2: Define roles and responsibilities	CAIO	CEO	HR, Legal	All
GOVERN 1.3: Establish DEIA goals	CAIO	CEO	HR, DEIA	All
GOVERN 1.4: Communicate AI risk priorities	CEO	Board	CAIO	All
GOVERN 1.5: Establish Agentic AI Committee	CAIO	CEO	Board, CRO	All
GOVERN 1.6: Establish Environmental Committee	CSO	CEO	CAIO, CFO	Board
GOVERN 2.1: Define decision authority	CAIO	CEO	Legal, Risk	All
GOVERN 3.1: Recruit diverse team members	HR	Dept Heads	DEIA	CAIO
GOVERN 4.1: Develop training curriculum	L&D	CAIO	SMEs, HR	All
GOVERN 5.1: Conduct internal audits	Audit Team	CAE	Mgmt	Exec
GOVERN 6.1: Develop Trustworthy AI Policy	CAIO	CEO	Legal, Ethics	Board

Appendix C: Decision Authority Matrix

Purpose: Define who has authority to make specific types of AI-related decisions based on risk level and impact.

Instructions:

- Use this matrix to determine required approvals for AI decisions
- Decision authority increases with risk level of AI system
- Multiple approvals may be required for high-risk decisions
- All decisions must be documented with rationale
- Escalate to next level when decision is outside authority

Template:

Decision Type	Low Risk	Moderate Risk	High Risk	Critical Risk
AI Use Case Approval	Team Lead	Department Head	C-Suite	Board
Model Architecture Selection	ML Engineer	Technical Lead	Chief AI Officer	AI Governance Committee
Data Source Selection	Data Scientist	Data Lead + DPO	Chief AI Officer + DPO	Board
Deployment Authorization	System Owner	Department Head	AI Governance Committee	Board
Agentic AI Deployment	Not Allowed	Agentic AI Committee	Agentic AI Committee + C-Suite	Board
Policy Exception Request	Not Allowed	Chief AI Officer	AI Governance Committee	Board
System Decommissioning	System Owner	Department Head	Chief AI Officer	AI Governance Committee

Appendix D: AI Risk Management Training Curriculum

Purpose: Comprehensive training program for all stakeholders in AI risk management, including agentic AI and environmental sustainability.

Training Program Structure:

- Foundational training required for all employees
- Role-specific training for AI practitioners and specialists
- Executive briefings for leadership and Board
- Continuous learning for emerging topics (agentic AI, sustainability)
- Annual refresher training for all participants

Complete Training Curriculum:

Module	Audience	Duration	Prerequisites	Frequency
AI Fundamentals	All Employees	2 hours	None	Once (onboarding)
Responsible AI Principles	All Employees	2 hours	AI Fundamentals	Annual
AI Risk Management	AI Practitioners	8 hours	AI Fundamentals	Annual
Bias and Fairness	Data Scientists, ML Engineers	4 hours	AI Risk Mgmt	Annual
AI Security	Engineers, Security	4 hours	AI Fundamentals	Annual
AI Privacy	Data Scientists, Privacy	4 hours	AI Fundamentals	Annual
AI Governance	Executives, Managers	3 hours	Responsible AI	Annual
Agentic AI Governance	AI Teams, Managers	4 hours	AI Risk Mgmt	Annual
Environmental Sustainability	AI Teams, Sustainability	2 hours	AI Fundamentals	Annual

Appendix E: Governance Committee Charters and Policy Templates

Purpose: Templates for establishing key governance bodies and policies.

E.1: AI Governance Committee Charter Template

- Committee Name: AI Governance Committee
- Purpose: Provide executive-level oversight of AI risk management across the organization
- Scope: All AI systems, policies, procedures, and risk management activities
- Authority: Approve AI policies, review high-risk AI systems, allocate AI resources, escalate to Board as needed
- Membership: CEO (Chair), Chief AI Officer, Chief Risk Officer, CISO, CFO, CLO, Business Unit Heads
- Meeting Frequency: Monthly regular meetings, ad-hoc for urgent matters
- Quorum: 50% of members must be present for decisions
- Decision Making: Consensus preferred, majority vote if needed
- Reporting: Quarterly reports to Board of Directors
- Documentation: Meeting minutes, decisions, rationales maintained permanently

E.2: AI Ethics Review Board Charter Template

- Committee Name: AI Ethics Review Board
- Purpose: Provide independent ethical assessment of high-risk and socially impactful AI systems
- Scope: High-risk AI per organizational criteria, novel AI applications, systems with significant social impact
- Authority: Recommend approval, conditional approval, or rejection of AI systems based on ethical assessment
- Membership: External ethicists (50%), internal representatives from legal, social science, technical teams, community representatives
- Meeting Frequency: Monthly, with additional sessions as needed for reviews
- Review Process: Project team submits application, board reviews documentation, deliberates, issues decision with rationale
- Decision Criteria: Alignment with values, fairness, transparency, potential harms, societal benefit
- Appeal Process: Project teams may appeal decisions to AI Governance Committee
- Reporting: Annual report to Board summarizing reviews and recommendations

E.3: Trustworthy AI Policy Template Outline

Policy Structure:

- 1) Purpose and Scope: Define policy objectives and coverage
- 2) Principles: Organizational commitment to trustworthy AI
- 3) Trustworthiness Requirements:
 - Valid and Reliable: Performance standards, testing requirements
 - Safe: Risk assessment, safety controls, fail-safe mechanisms
 - Secure and Resilient: Security-by-design, threat mitigation, recovery procedures
 - Accountable and Transparent: Clear responsibility, documentation, transparency
 - Explainable and Interpretable: Explanation requirements by system risk level
 - Privacy-Enhanced: Privacy-by-design, PETs, data minimization
 - Fair with Bias Managed: Fairness criteria, bias testing, mitigation
- 4) Agentic AI Requirements: Agent boundaries, identity management, oversight (reference GOVERN 1.5)
- 5) Environmental Sustainability: Carbon tracking, energy monitoring, optimization (reference GOVERN 1.6)
- 6) Roles and Responsibilities: Who is responsible for ensuring compliance
- 7) Compliance and Enforcement: Consequences for policy violations
- 8) Policy Review: Annual review and update process
- 9) Exceptions: Process for requesting policy exceptions
- 10) References: Related policies, standards, regulations

E.4: Complete Control Mapping Summary

Comprehensive mapping of all GOVERN subcategories to ISO 42001, ISO 27001, ISO 23894, Singapore MGF, and EU AI Act requirements.

GOVERN	ISO 42001	ISO 27001	ISO 23894	Singapore	EU AI Act
1.1	4.2, 6.3	4.2, A.5.1	4.2	Dim 1	Art. 16, 71
1.2	5.3	5.3	5.3	Dim 2	Art. 16
1.3	6.1.3, 7.2	7.2	6.1	Dim 1	Art. 10
1.4	5.1, 7.3	5.1, 7.3	5.1	Dim 2	Art. 9
1.5	5.3, 8.1, 6.1.3, 9.1	5.3, 8.1, A.5.1, 9.1	5.1-5.5	All Dims	Art. 9, 14, 16
1.6	4.1, 5.1, 5.3, 6.1.3, 9.1	4.1, 5.1, 5.3, 8.1, 9.1	4.1, 5.1-5.5	All Dims	Art. 9, Recital 60
2.1	5.3, 8.1	5.3	5.3	Dim 2	Art. 16
3.1	7.2	7.2	6.1	Dim 1	Art. 10
4.1	7.2, 7.3	7.2, A.6.3	7.2	Dim 4	Art. 4
5.1	9.2, 9.3	9.2, 9.3	5.1	Dim 2	Art. 17
6.1	5.2	5.2	5.2	All Dims	Art. 9, 13, 14

Appendices F: Agentic AI and Environmental Sustainability

NEW IN AI RMF 2026: These appendices provide templates and frameworks specifically for governing agentic AI systems and managing environmental sustainability of AI operations.

Appendix F: Agentic AI Committee Charter Template

Purpose: Establish specialized governance committee for autonomous AI systems with planning and action-taking capabilities.

Charter Template:

1. Committee Name and Authority

- Official Name: Agentic AI Committee
- Reporting Structure: Reports to AI Governance Board and CEO
- Authority Level: Approval required for all high-risk agentic AI deployments, agent identity frameworks, agentic incident responses
- Escalation Path: Critical agentic risks escalate to Board of Directors

2. Scope of Oversight

- All AI systems with autonomous planning and action-taking capabilities
- Multi-agent systems (sequential, supervisor, swarm architectures)
- AI agents with tool access (APIs, databases, external systems)
- AI agents using Model Context Protocol (MCP) or Agent-to-Agent (A2A) communication
- AI agents with computer use capabilities

3. Committee Membership

Role	Voting Member	Responsibilities
Committee Chair	Yes	Chief AI Officer or designated senior technical leader with agentic AI expertise
AI Engineering Lead	Yes	Technical expertise on agent architectures, capabilities, limitations
Security Representative	Yes	CISO or delegate, security threat modeling, attack surface analysis
Risk Management Representative	Yes	Chief Risk Officer or delegate, enterprise risk assessment expertise
Legal/Compliance Representative	Yes	Legal counsel, regulatory compliance, liability considerations
Ethics Representative	Yes	AI Ethics Board liaison, ethical implications of autonomous systems
Domain Expert (Rotating)	Yes	Business unit representative for domain where agent will operate
Agent Owners (As Needed)	No	Present specific agents for review, answer technical questions

4. Meeting Frequency and Procedures

- Regular Meetings: Monthly minimum
- Special Meetings: Ad-hoc for critical agentic AI deployments or incidents
- Quorum: 50% of voting members must be present
- Decision Making: Consensus preferred, majority vote (>50%) if consensus not reached
- Minutes: Detailed minutes including decisions, rationales, dissenting opinions
- Attendance: Members may designate alternates with committee approval

5. Key Responsibilities

- Review and approve all high-risk agentic AI deployments before production release
- Establish and maintain agent identity management framework and policies
- Define agent boundaries, permissions, and autonomy levels for different risk categories
- Review agentic AI risk assessments and approve mitigation strategies
- Oversee implementation of agent monitoring and oversight mechanisms
- Approve automation bias mitigation programs for human overseers
- Review agentic AI incidents and ensure appropriate corrective actions
- Update agentic AI governance policies based on emerging risks and learnings
- Coordinate with AI Ethics Review Board on ethical implications of agentic systems
- Report quarterly to AI Governance Board on agentic AI governance status

6. Decision Criteria for Agentic AI Approval

- Risk Assessment: Comprehensive agentic risk assessment completed using Appendix I template
- Boundaries Defined: Clear action-space, autonomy, and environmental boundaries documented
- Identity Management: Agent identity and permission inheritance properly implemented
- Human Oversight: Appropriate level of human oversight for risk level
- Monitoring: Real-time monitoring and alerting capabilities in place
- Incident Response: Agentic incident response procedures prepared
- Testing: Extensive testing in sandboxed environments completed
- Rollout Plan: Gradual deployment strategy with rollback capability

Appendix G: Environmental Sustainability Committee Charter Template

Purpose: Establish governance committee to track, manage, and minimize environmental impact of AI systems.

1. Committee Name and Authority

- Official Name: Environmental Sustainability Committee (AI Systems)
- Reporting Structure: Reports to Chief Sustainability Officer and AI Governance Board
- Authority Level: Sets sustainability targets, approves high-energy AI deployments, mandates environmental reporting
- Escalation Path: Significant environmental impacts escalate to Board ESG Committee

2. Scope of Oversight

- Energy consumption of all AI model training and inference operations
- Carbon footprint (Scope 2 and Scope 3 emissions) from AI systems
- Computational resource utilization and optimization
- Hardware lifecycle environmental impact (manufacturing, use, disposal)
- Data center environmental context and renewable energy usage
- Sustainability reporting and stakeholder communications

3. Committee Membership

Role	Voting Member	Responsibilities
Committee Chair	Yes	Chief Sustainability Officer or Environmental Sustainability Officer
Chief AI Officer	Yes	AI strategy alignment, technical feasibility of sustainability initiatives
Energy Management Lead	Yes	Data center energy expertise, renewable energy procurement
AI Engineering Representative	Yes	Technical optimization opportunities, model efficiency
Finance Representative	Yes	Energy cost tracking, sustainability investment ROI
Legal/Compliance Representative	Yes	Environmental regulations (SEC climate rules, EU CSRD), reporting requirements
External Advisor (As Needed)	No	Climate scientists, sustainability consultants for expert guidance

4. Meeting Frequency and Procedures

- Regular Meetings: Quarterly minimum
- Special Meetings: For significant AI deployments exceeding energy thresholds
- Quorum: 50% of voting members must be present
- Decision Making: Consensus preferred, majority vote if needed
- Minutes: Document sustainability metrics, decisions, improvement initiatives
- External Expertise: May invite climate scientists, sustainability consultants as advisors

5. Key Responsibilities

- Define sustainability metrics and targets for AI systems (energy, carbon, efficiency)
- Review carbon footprint assessments for new AI systems before deployment
- Monitor energy consumption and carbon emissions of deployed AI systems
- Approve or reject high-energy AI deployments based on sustainability criteria
- Oversee implementation of resource optimization initiatives
- Review energy monitoring dashboards and sustainability metrics
- Ensure compliance with environmental regulations and corporate sustainability commitments
- Approve annual AI environmental impact reports
- Coordinate with corporate sustainability team on AI-related environmental initiatives
- Report quarterly to AI Governance Board and annually to Board ESG Committee

6. Sustainability Approval Criteria

- Carbon Assessment: Complete carbon footprint assessment using Appendix J template
- Energy Budget: Project fits within allocated energy budget or justification for exceeding
- Optimization: Evidence of resource optimization efforts (model compression, efficient hardware)
- Renewable Energy: Deployment in data centers with high renewable energy percentage when possible
- Monitoring: Energy and carbon monitoring instrumented and operational
- Reporting: Sustainability metrics will be tracked and reported
- Alternatives Considered: Less energy-intensive approaches evaluated
- Business Justification: Environmental impact proportionate to business value

Appendix H: Agent Identity Management Framework Template

Purpose: Define how AI agents are identified, authenticated, authorized, and tracked within organizational systems.

1. Agent Identity Principles

Principle	Description
Unique Identification	Each agent has unique identity distinguishable from human users. No shared credentials across agents.
Hierarchical Ownership	Agent identity linked to supervising agent, human user, or organizational unit. Ownership chain clearly documented.
Capacity Recording	Document whether agent acts independently or on behalf of specific human. Record in agent identity metadata.
Permission Inheritance	Agents cannot receive permissions exceeding those of authorizing human or supervising entity. Permissions ceiling enforced.
Audit Trail	All agent identity delegations, permission grants, and actions logged. Immutable audit records retained per policy.

2. Agent Identity Schema

Required Agent Identity Attributes:

- Agent ID: Unique identifier (UUID or similar) globally unique across organization
- Agent Name: Human-readable name describing agent purpose
- Agent Type: Classification (autonomous, semi-autonomous, tool-augmented, etc.)
- Owner: Individual or team responsible for agent behavior
- Supervising Entity: Parent agent, human user, or organizational unit agent acts on behalf of
- Creation Timestamp: When agent identity was created
- Status: Active, suspended, deactivated
- Risk Classification: Low, moderate, high, critical based on agent risk assessment
- Permissions Ceiling: Maximum permissions agent can ever receive based on supervising entity
- Audit Requirements: Logging level and retention period for agent actions

3. Agent Authentication Methods

- API Keys: Long-lived credentials for programmatic agent access
- OAuth 2.0 Tokens: Short-lived tokens with refresh capability
- Service Account Certificates: X.509 certificates for agent-to-system authentication
- Agent-to-Agent Tokens: Specialized tokens for inter-agent communication (A2A protocol)
- Mutual TLS: Certificate-based mutual authentication for high-security agents
- Rotation Policy: Credentials rotated minimum every 90 days, more frequently for high-risk agents

4. Agent Authorization Framework

Authorization Type	Mechanism	Use Cases
Role-Based (RBAC)	Agent assigned predefined role	Agents with standard, repeatable functions (customer service, reporting)
Attribute-Based (ABAC)	Permissions based on agent attributes	Dynamic permission needs based on context (time, location, data sensitivity)
Dynamic Permission Grants	Temporary elevated permissions for specific task	One-time operations requiring elevated access, with automatic expiration
Delegated Authorization	Agent receives subset of human user permissions	Personal assistants acting on behalf of specific users
Principle of Least Privilege	Minimum permissions necessary for function	All agents, reviewed quarterly to remove unused permissions

5. Agent Identity Lifecycle

- Creation: Agent identity created with approval from Agent Owner, registered in identity management system
- Activation: Agent identity activated, authentication credentials issued, permissions granted
- Modification: Permission changes, ownership changes, risk reclassification documented with approval
- Suspension: Temporary deactivation for investigation or maintenance, credentials revoked
- Reactivation: Restoration after suspension, requires re-approval
- Deactivation: Permanent removal, credentials permanently revoked, audit logs archived

- Audit: All lifecycle events logged with timestamp, actor, reason

6. Agent Identity Monitoring

- Access Monitoring: Track all agent authentication attempts (successful and failed)
- Permission Usage: Monitor which permissions agents actually use vs. granted
- Anomaly Detection: Alert on unusual agent behavior (access patterns, resource usage, error rates)
- Orphaned Agents: Identify agents without active owners or supervisors
- Overprivileged Agents: Flag agents with permissions they never use
- Compliance Review: Quarterly review of all agent identities and permissions

Appendix I: Agent Risk Assessment Template

Purpose: Systematically assess risks posed by agentic AI systems to determine appropriate governance controls.

Instructions:

- Complete this assessment for every agentic AI system before deployment
- Use scoring system: 1 (Very Low) to 5 (Critical) for each risk dimension
- Overall Risk = Average of all dimension scores
- Risk Classification: 1.0-1.9 Low, 2.0-2.9 Moderate, 3.0-3.9 High, 4.0-5.0 Critical
- Submit completed assessment to Agentic AI Committee for approval

Agent Risk Assessment Template:

Risk Dimension	Assessment Questions	Score (1-5)	Mitigation
Domain & Use Case	How tolerant is domain to errors? (Low tolerance = 5, High tolerance = 1)	[1-5]	[Notes]
Data Access	What is sensitivity of data agent accesses? (PII/Confidential = 5, Public = 1)	[1-5]	[Notes]
System Access	Does agent access external systems? (External/production = 5, Sandboxed = 1)	[1-5]	[Notes]
Action Scope	Range of actions agent can take? (Many tools, write access = 5, Read-only, few tools = 1)	[1-5]	[Notes]
Reversibility	Can agent actions be undone? (Irreversible = 5, Easily reversible = 1)	[1-5]	[Notes]
Autonomy Level	How much independent judgment? (Independent planning = 5, Follows strict SOP = 1)	[1-5]	[Notes]
Task Complexity	Number of steps and analysis required? (Many steps, deep analysis = 5, Simple task = 1)	[1-5]	[Notes]
Threat Surface	Attack vectors (memory poisoning, tool misuse, privilege escalation)? (High exposure = 5, Low = 1)	[1-5]	[Notes]
Cascading Impact	Can agent errors propagate to other systems? (Wide propagation = 5, Isolated = 1)	[1-5]	[Notes]
OVERALL RISK SCORE	Average of all scores above	[Avg]	Classification: L/M/H/C

Appendix J: Carbon Footprint Assessment Template

Purpose: Calculate and document carbon emissions associated with AI system throughout its lifecycle.

Instructions:

- Complete for all new AI systems before deployment
- Update annually for deployed systems
- Calculate Scope 2 (purchased electricity) and Scope 3 (embodied hardware carbon) emissions
- Use regional carbon intensity factors for electricity consumption
- Submit to Environmental Sustainability Committee if carbon footprint exceeds thresholds

Carbon Footprint Assessment Template:

Lifecycle Phase	Calculation Method	CO2e (tonnes)	Notes
Model Training	$(\text{Training kWh}) \times (\text{Regional carbon intensity gCO2e/kWh}) \div 1,000,000$	[Value]	GPU hours, location, carbon intensity
Annual Inference (Scope 2)	$(\text{Estimated annual inference kWh}) \times (\text{Carbon intensity}) \div 1,000,000$	[Value]	Based on projected usage
Development/Experimentation	$(\text{Dev environment kWh}) \times (\text{Carbon intensity}) \div 1,000,000$	[Value]	Prototype iterations
Infrastructure Overhead	$(\text{Data center overhead kWh}) \times (\text{Carbon intensity}) \div 1,000,000$	[Value]	Load balancers, monitoring
Hardware Embodied Carbon (Scope 3)	$(\text{GPU embodied CO2e per unit}) \times (\text{Number of GPUs}) \div (\text{Expected lifetime years})$	[Value]	Amortized over hardware life
TOTAL ANNUAL CARBON FOOTPRINT	Sum of all phases above	[Total]	Review annually

Appendix K: Sustainability Metrics Tracking Template

Purpose: Track ongoing environmental sustainability metrics for deployed AI systems.

Instructions:

- Update monthly for all production AI systems
- Aggregate metrics quarterly for committee review
- Track trends over time to measure improvement
- Highlight systems exceeding targets in red
- Use for quarterly reporting to Environmental Sustainability Committee

Monthly Sustainability Metrics Template:

AI System	Energy (MWh)	Carbon (tonnes)	GPU Util %	Target Met?	Actions
Customer Service Bot	12.5	4.8	78%	Yes	On track
Fraud Detection Model	45.2	18.1	92%	Yes	Excellent efficiency
Content Recommendation	156.8	62.7	65%	No - Over 10%	Optimization needed
Research LLM Training	892.3	356.9	88%	Review Required	High carbon - justify value
TOTALS (Month)	1,106.8	442.5	Avg: 81%	3 of 4 on target	Overall trending positive

Appendix L: Third-Party & Tool Ecosystem Governance Toolkit

Purpose: Provide templates and guidance for governing third-party AI suppliers, model providers, agent toolchains, and cloud services.

Instructions:

- Use this appendix for all new third-party AI engagements and for annual reassessment of existing suppliers.
- Tailor depth of due diligence to supplier risk tier (Low/Moderate/High). High risk suppliers require documented exit planning before production.
- Attach evidence for each due diligence item (reports, attestations, policies, technical documentation).

Guidance: *If a supplier refuses to provide sufficient evidence for High risk use, treat as High risk and escalate to the AI Governance Committee before proceeding.*

L.1 Third-Party AI and Agent Tool Inventory Template:

Supplier/Service	Component Type	Business Purpose	System/Agent Link	Data Categories	Autonomy/Action Authority	Environments	Risk Tier	Contract Reference	Owner
<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>	<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>	<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>	<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>	<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>	<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>	<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>	<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>	<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>	<i>[Guidance: Enter the required portfolio metadata. Use concise, auditable language.]</i>

L.2 AI Supplier Due Diligence Checklist (summary):

Domain	Minimum Questions	Evidence Examples	Pass/Fail/Notes
Security	Does the supplier have documented security controls and independent assurance?	SOC 2/ISO 27001 cert; pen test summary; security policy	<i>[Guidance: Record gaps and compensating controls.]</i>
Privacy & Data Use	How is customer data used, retained, and deleted? Is training on customer data allowed?	DPA; data retention policy; opt-out terms	<i>[Guidance: Flag any secondary use of data.]</i>
Model/Service Changes	How are model updates announced and controlled? Is version pinning supported?	Change policy; release notes; versioning controls	<i>[Guidance: Require notification for material changes.]</i>
Logging & Traceability	Can logs support investigation (request IDs, model version, tool calls)?	Logging docs; retention/availability commitments	<i>[Guidance: Ensure evidence supports forensic needs.]</i>
Reliability & SLAs	What are SLAs and support response times? What are degradation modes?	SLA contract; support policy; status page	
Incident Management	How are incidents detected and notified? What are notification SLAs?	IR policy; notification SLA; post-incident reports	
Business Continuity	How does the supplier handle outages and disaster recovery?	BCP/DR summaries; redundancy architecture	
Compliance	Which regulatory obligations are supported (as applicable)?	Compliance attestations; audit reports	

L.3 AI Supplier Contract Clause Library (select and tailor):

Guidance: *Include clauses proportionate to risk tier. For High risk and agentic toolchains, include audit rights, incident SLAs, change notification, and data processing restrictions at minimum.*

- **Audit and assurance:** Supplier provides independent assurance (SOC/ISO) and allows customer audit or third-party audit reports upon request.
- **Incident notification:** Supplier notifies customer within defined SLA for confirmed incidents impacting confidentiality, integrity, availability, or safety.
- **Model/service change notification:** Supplier provides advance notice for material changes, supports version pinning where feasible, and documents changes.
- **Data processing restrictions:** Customer data is used only to provide the service; no training on customer data unless explicitly approved in writing.
- **Logging and investigation support:** Supplier provides access to relevant logs/metadata to support investigations and regulatory inquiries.
- **Subprocessor transparency:** Supplier discloses subprocessors and provides notice and objection mechanism for new subprocessors.
- **Exit support:** Supplier supports data export, deletion certification, and reasonable transition assistance upon termination.

L.4 Shared Responsibility Matrix Template:

Control Area	Supplier Responsibilities	Customer Responsibilities	Notes/Agreed Artifacts
Identity and Access Management	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>
Logging and Monitoring	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>
Vulnerability and Patch Management	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>
Data Protection and Privacy	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>
Incident Response and Notification	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>	<i>[Guidance: Define who does what and reference the contract or internal procedure.]</i>

L.5 Exit and Transition Plan Checklist (for High risk suppliers):

- Data export plan (formats, timing, verification).
- Credential and key rotation plan (including agent tool credentials).
- Service replacement strategy and timeline.
- Data deletion/return certification requirements.
- Investigation and record retention requirements post-exit.

Guidance: *Attach the exit plan to the AI System & Agent Registry entry and validate at least annually for High risk suppliers.*

Appendix M: AI System & Agent Registry Template

Purpose: Maintain an authoritative inventory of AI systems and agents to support traceability, approvals, monitoring, and auditability.

Instructions:

- Create a new registry entry for each AI system and for each agent (including agent versions) deployed or piloted in the organization.
- Update the registry whenever a material change occurs (new model/tool, new data category, new autonomy, new supplier, new use case).
- Link registry entries to assessments and approvals (impact assessment, risk assessment, security review, and residual risk acceptance if applicable).

Guidance: *Treat the registry as a controlled record. It should be reviewable and exportable for audits without manual rework.*

Registry Template:

Registry ID	System/Agent Name	Type (System/Agent)	Owner	Business Unit	Purpose/Use Case	Deployment (Dev/Test/Prod)	Model/Version	Supplier(s)	Autonomy Level	Tool Permissions/Action-Space	Data Categories	Risk Tier	Last/Next Review Date
<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>	<i>[Guidance: Use controlled terms where possible (e.g., Autonomy = Low/Moderate/High).]</i>

Appendix N: AI Risk Appetite and Risk Acceptance Templates

Purpose: Provide standardized templates to define AI risk appetite, risk criteria, and residual risk acceptance and exceptions.

Instructions:

- Complete N.1 and N.2 at the organizational level and update at least annually or upon major strategy/regulatory changes.
- Use N.3 for any deployment where residual risk exceeds the target appetite and requires explicit acceptance.
- Use N.4 to document exceptions to mandatory controls and track compensating controls and closure dates.

N.1 AI Risk Appetite Statement Template:

Guidance: Write appetite statements in measurable terms where feasible (e.g., thresholds, prohibitions, required approvals). Avoid vague language such as 'low risk' without criteria.

Dimension	Risk Appetite Statement
Safety/Harm Severity	<i>[Guidance: State tolerance and required controls/approvals. Example: 'No autonomous external transactions above \$X without dual human approval.']</i>
Security Exposure	<i>[Guidance: State tolerance and required controls/approvals. Example: 'No autonomous external transactions above \$X without dual human approval.']</i>
Privacy Sensitivity	<i>[Guidance: State tolerance and required controls/approvals. Example: 'No autonomous external transactions above \$X without dual human approval.']</i>
Fairness/DEIA Sensitivity	<i>[Guidance: State tolerance and required controls/approvals. Example: 'No autonomous external transactions above \$X without dual human approval.']</i>
Autonomy Level and Action Authority	<i>[Guidance: State tolerance and required controls/approvals. Example: 'No autonomous external transactions above \$X without dual human approval.']</i>
Reversibility and Human Override	<i>[Guidance: State tolerance and required controls/approvals. Example: 'No autonomous external transactions above \$X without dual human approval.']</i>
Financial Impact	<i>[Guidance: State tolerance and required controls/approvals. Example: 'No autonomous external transactions above \$X without dual human approval.']</i>
Regulatory Classification	<i>[Guidance: State tolerance and required controls/approvals. Example: 'No autonomous external transactions above \$X without dual human approval.']</i>

N.2 AI Risk Criteria Matrix Template:

Guidance: Define scoring scales and thresholds consistently with your risk taxonomy. Include 'Critical' thresholds that trigger executive review.

Risk Domain	Low (1)	Moderate (2)	Elevated (3)	High (4)	Critical (5)
Safety	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>
Security	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>
Privacy	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>
Fairness/DEIA	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>	<i>[Guidance: Describe criteria for this level.]</i>

N.3 Residual Risk Acceptance Form Template:

Guidance: *Residual risk acceptance must be time-bounded and include monitoring conditions. Acceptance is not approval to ignore mandatory controls.*

Field	Entry
System/Agent Registry ID	<i>[Guidance: Provide specific, auditable detail. Reference evidence artifacts and link to impact assessment.]</i>
Residual Risk Summary (what remains and why)	<i>[Guidance: Provide specific, auditable detail. Reference evidence artifacts and link to impact assessment.]</i>
Residual Risk Rating (domain + overall)	<i>[Guidance: Provide specific, auditable detail. Reference evidence artifacts and link to impact assessment.]</i>
Mitigations in Place (controls and monitoring)	<i>[Guidance: Provide specific, auditable detail. Reference evidence artifacts and link to impact assessment.]</i>
Justification for Acceptance (business need)	<i>[Guidance: Provide specific, auditable detail. Reference evidence artifacts and link to impact assessment.]</i>
Acceptance Duration and Expiry Date	<i>[Guidance: Provide specific, auditable detail. Reference evidence artifacts and link to impact assessment.]</i>
Monitoring Conditions/Triggers for Re-Review	<i>[Guidance: Provide specific, auditable detail. Reference evidence artifacts and link to impact assessment.]</i>
Approving Authority (per Decision Authority Matrix)	<i>[Guidance: Provide specific, auditable detail. Reference evidence artifacts and link to impact assessment.]</i>
Signatures and Date	<i>[Guidance: Provide specific, auditable detail. Reference evidence artifacts and link to impact assessment.]</i>

N.4 AI Policy Exception Log Template:

Guidance: Every exception must list compensating controls and a closure date. Exceptions should be rare and reviewed at least quarterly.

Exception ID	Date	System/Agent ID	Policy/Control	Exception Description	Compensating Controls	Approver	Closure Date/Status
<i>[Guidance: Keep entries concise; link evidence separately.]</i>	<i>[Guidance: Keep entries concise; link evidence separately.]</i>	<i>[Guidance: Keep entries concise; link evidence separately.]</i>	<i>[Guidance: Keep entries concise; link evidence separately.]</i>	<i>[Guidance: Keep entries concise; link evidence separately.]</i>	<i>[Guidance: Keep entries concise; link evidence separately.]</i>	<i>[Guidance: Keep entries concise; link evidence separately.]</i>	<i>[Guidance: Keep entries concise; link evidence separately.]</i>

Appendix O: Integrated AI Impact Assessment (AIIA) Template

Purpose: Provide an integrated impact assessment template covering stakeholder impact, safety, security, privacy, fairness, transparency, agentic autonomy, environmental impact, and supplier/tool dependencies.

Instructions:

- Complete this assessment before production deployment and update upon material changes (new model, new tools, new data, new user population).
- Attach supporting evidence (testing reports, evaluations, model cards, security reviews, privacy assessments).
- Use clear, measurable language and document residual limitations and mitigations.

Guidance: *If the system or agent is High risk, submit this assessment to the AI Governance Committee for review and record the decision in the registry.*

O.1 System Overview:

Field	Response
System/Agent Registry ID	<i>[Guidance: Describe the operational decision and its consequences. Avoid marketing language.]</i>
System/Agent Name and Version	<i>[Guidance: Describe the operational decision and its consequences. Avoid marketing language.]</i>
Owner and Business Unit	<i>[Guidance: Describe the operational decision and its consequences. Avoid marketing language.]</i>
Intended Use and Decision Context	<i>[Guidance: Describe the operational decision and its consequences. Avoid marketing language.]</i>
User Population (who will use/oversee it)	<i>[Guidance: Describe the operational decision and its consequences. Avoid marketing language.]</i>
Deployment Context (Dev/Test/Prod; geographies)	<i>[Guidance: Describe the operational decision and its consequences. Avoid marketing language.]</i>
Model/Provider and Key Dependencies	<i>[Guidance: Describe the operational decision and its consequences. Avoid marketing language.]</i>

O.2 Stakeholder and Impact Analysis:

Topic	Assessment
Affected stakeholders (direct/indirect)	<i>[Guidance: Include worst-case plausible harm scenarios and mitigations.]</i>
Potential benefits and who receives them	<i>[Guidance: Include worst-case plausible harm scenarios and mitigations.]</i>
Potential harms and who bears them	<i>[Guidance: Include worst-case plausible harm scenarios and mitigations.]</i>
Human rights / dignity considerations (if applicable)	<i>[Guidance: Include worst-case plausible harm scenarios and mitigations.]</i>
Redress and escalation mechanisms (link to Appendix P)	<i>[Guidance: Include worst-case plausible harm scenarios and mitigations.]</i>

O.3 Data, Privacy, and Security:

Topic	Assessment
Data categories processed (sensitive? regulated?)	<i>[Guidance: Reference concrete controls and evidence artifacts.]</i>
Data minimization and retention approach	<i>[Guidance: Reference concrete controls and evidence artifacts.]</i>
Privacy safeguards (PIA completed? PETs used?)	<i>[Guidance: Reference concrete controls and evidence artifacts.]</i>
Security threat model summary and key controls	<i>[Guidance: Reference concrete controls and evidence artifacts.]</i>
Logging/traceability and retention	<i>[Guidance: Reference concrete controls and evidence artifacts.]</i>

O.4 Fairness/DEIA and Transparency:

Topic	Assessment
Fairness risks and protected group considerations	<i>[Guidance: State metrics, thresholds, and monitoring triggers.]</i>
Testing approach and metrics (pre/post deployment)	<i>[Guidance: State metrics, thresholds, and monitoring triggers.]</i>
Bias mitigation approach and residual limitations	<i>[Guidance: State metrics, thresholds, and monitoring triggers.]</i>
Explainability/transparency approach (model cards, user notices)	<i>[Guidance: State metrics, thresholds, and monitoring triggers.]</i>
Accessibility and usability considerations	<i>[Guidance: State metrics, thresholds, and monitoring triggers.]</i>

O.5 Agentic Autonomy and Action Authority (if applicable):

Topic	Assessment
Autonomy level (Low/Moderate/High) and justification	<i>[Guidance: Describe concrete controls and how they are enforced technically.]</i>
Action-space and tool permissions (what actions are possible)	<i>[Guidance: Describe concrete controls and how they are enforced technically.]</i>
Human approval requirements and override mechanisms	<i>[Guidance: Describe concrete controls and how they are enforced technically.]</i>
Guardrails (rate limits, transaction limits, sandboxing)	<i>[Guidance: Describe concrete controls and how they are enforced technically.]</i>
Failure modes and containment (rollback, kill switch)	<i>[Guidance: Describe concrete controls and how they are enforced technically.]</i>
Monitoring signals for unsafe/unauthorized actions	<i>[Guidance: Describe concrete controls and how they are enforced technically.]</i>

O.6 Environmental and Supplier Impact:

Topic	Assessment
Energy/carbon footprint estimate (training vs inference) and thresholds	<i>[Guidance: Use best available data; document assumptions.]</i>
Optimization measures (efficiency, caching, scheduling)	<i>[Guidance: Use best available data; document assumptions.]</i>
Third-party dependencies and risk tier (link to Appendix L)	<i>[Guidance: Use best available data; document assumptions.]</i>
Exit plan required? (High risk suppliers)	<i>[Guidance: Use best available data; document assumptions.]</i>
Sustainability metrics tracking plan (Appendix J/K)	<i>[Guidance: Use best available data; document assumptions.]</i>

O.7 Risk Summary, Decision, and Sign-Off:

Field	Entry
Overall Risk Tier and Rationale	<i>[Guidance: If approval is conditional, list conditions as verifiable deliverables.]</i>
Key Residual Risks	<i>[Guidance: If approval is conditional, list conditions as verifiable deliverables.]</i>
Required Mitigations and Owners	<i>[Guidance: If approval is conditional, list conditions as verifiable deliverables.]</i>
Recommendation (Approve/Approve with Conditions/Reject)	<i>[Guidance: If approval is conditional, list conditions as verifiable deliverables.]</i>
Residual Risk Acceptance Required? (Appendix N.3)	<i>[Guidance: If approval is conditional, list conditions as verifiable deliverables.]</i>
Approvals and Dates (System Owner, AI Risk, CISO/Privacy as applicable)	<i>[Guidance: If approval is conditional, list conditions as verifiable deliverables.]</i>

Appendix P: End-User Enablement, Complaint Intake, and Redress Toolkit

Purpose: Provide standardized disclosures, operating guidance, intake forms, and redress tracking for AI systems and agents.

Instructions:

- Use P.1 and P.2 for all deployments; tailor based on risk tier and user population.
- Use P.3 and P.4 to ensure complaints and redress actions are traceable and auditable.
- Use P.5 workflow to define SLAs, escalation, and linkage to incident response where appropriate.

Guidance: *When the AI system influences high-impact outcomes, ensure disclosures and escalation routes are prominently available and accessible.*

P.1 End-User Disclosure Template:

Disclosure Element	Content
What the system/agent is (capabilities)	<i>[Guidance: Use plain language. Include contact points and SLAs for response.]</i>
Intended use and prohibited uses	<i>[Guidance: Use plain language. Include contact points and SLAs for response.]</i>
Known limitations and uncertainty cues	<i>[Guidance: Use plain language. Include contact points and SLAs for response.]</i>
Human oversight and how to escalate	<i>[Guidance: Use plain language. Include contact points and SLAs for response.]</i>
Data handling notice (high-level)	<i>[Guidance: Use plain language. Include contact points and SLAs for response.]</i>
How to report issues/complaints	<i>[Guidance: Use plain language. Include contact points and SLAs for response.]</i>

P.2 Operator Quick Reference Template:

- Do: Verify critical outputs; confirm approvals for agent actions; report anomalies immediately.
- Do not: Treat the system as a final authority; bypass approval gates; use outside intended context.
- Stop/Override: Use the designated kill switch or manual override procedure when unsafe or unauthorized behavior is suspected.

Escalation: List roles, contact methods, and expected response times.

Guidance: *Keep this one page where possible and publish where operators will actually see it (in-product, runbook, or help center).*

P.3 Complaint/Feedback Intake Form Template:

Field	Entry
Date/Time Received	<i>[Guidance: Capture enough detail to reproduce and investigate.]</i>
Reporter (optional) and contact info	<i>[Guidance: Capture enough detail to reproduce and investigate.]</i>
System/Agent Registry ID	<i>[Guidance: Capture enough detail to reproduce and investigate.]</i>
Issue type (accuracy/bias/privacy/security/other)	<i>[Guidance: Capture enough detail to reproduce and investigate.]</i>
Description of issue and impact	<i>[Guidance: Capture enough detail to reproduce and investigate.]</i>
Evidence (screenshots, logs, transcripts)	<i>[Guidance: Capture enough detail to reproduce and investigate.]</i>
Severity (Low/Moderate/High/Critical)	<i>[Guidance: Capture enough detail to reproduce and investigate.]</i>
Immediate containment actions taken	<i>[Guidance: Capture enough detail to reproduce and investigate.]</i>

P.4 Redress Log Template:

Case ID	Date	System/Agent ID	Issue Summary	Severity	Root Cause	Remediation/Redress	Owner	Status/Close Date
<i>[Guidance: Update as actions occur; link artifacts separately.]</i>	<i>[Guidance: Update as actions occur; link artifacts separately.]</i>	<i>[Guidance: Update as actions occur; link artifacts separately.]</i>	<i>[Guidance: Update as actions occur; link artifacts separately.]</i>	<i>[Guidance: Update as actions occur; link artifacts separately.]</i>	<i>[Guidance: Update as actions occur; link artifacts separately.]</i>	<i>[Guidance: Update as actions occur; link artifacts separately.]</i>	<i>[Guidance: Update as actions occur; link artifacts separately.]</i>	<i>[Guidance: Update as actions occur; link artifacts separately.]</i>

P.5 Complaint Triage Workflow (text):

- 1) Intake and acknowledgement within SLA.
- 2) Severity classification and immediate containment if needed.
- 3) Assign investigator and gather evidence (logs, traces, approvals).
- 4) Determine root cause and corrective/preventive actions (CAPA).
- 5) Provide redress and communicate resolution.
- 6) Close case with verification and update training/controls if systemic.

Guidance: *If the complaint indicates a security/privacy incident or safety-critical harm, invoke incident response immediately and treat complaint handling as part of incident management.*

Appendix Q: Framework Compliance Matrix

Purpose: Provide comprehensive cross-reference showing how each GOVERN procedure satisfies specific requirements from ISO 42001, ISO 27001, ISO 23894, Singapore MGF, and EU AI Act. Use this matrix for certification audits, regulatory compliance demonstrations, and gap analyses.

Instructions:

- Use this matrix during ISO 42001/27001 certification audits to show evidence of compliance
- Reference specific procedures when responding to EU AI Act compliance inquiries
- Conduct gap analysis by reviewing unchecked framework requirements
- Update matrix when procedures change or new framework requirements emerge
- Share relevant sections with external auditors and regulators

Q.1: ISO/IEC 42001:2023 Compliance Matrix

Complete mapping of ISO 42001 clauses to GOVERN procedures:

ISO 42001 Clause	Requirement	GOVERN Procedure(s)	Evidence Location
4.1	Understanding organization context	GOVERN 1.1.1, 1.6.1	Appendix A (Compliance Register), Appendix G
4.2	Stakeholder needs	GOVERN 1.1.1, 1.3.1	Stakeholder analysis in procedures
5.1	Leadership commitment	GOVERN 1.4.1, 5.1.1	Committee charters (Appendices E, F, G)
5.2	Policy	GOVERN 6.1.1	Trustworthy AI Policy (Appendix E.3)
5.3	Roles and responsibilities	GOVERN 1.2.1, 1.2.2	RACI Matrix (Appendix B), Roles table
6.1.3	AI impact assessment	GOVERN 1.3.1	AIIA Template (Appendix O)
6.3	Compliance obligations	GOVERN 1.1.1	Compliance Register (Appendix A)
7.2	Competence	GOVERN 3.1.1, 4.1.1	Training Curriculum (Appendix D)
7.3	Awareness	GOVERN 4.1.2	Training records, communications
8.1	Operational planning	GOVERN 2.1.1, 2.1.2	Decision Authority Matrix (Appendix C)
9.1	Monitoring and measurement	GOVERN 1.5.5, 1.6.3	Monitoring dashboards, metrics
9.2	Internal audit	GOVERN 5.1.3	Audit plan and reports
9.3	Management review	GOVERN 5.1.4	Management review records

Q.2: EU AI Act Compliance Matrix

Mapping of EU AI Act articles to GOVERN procedures:

AI Act Article	Requirement	GOVERN Procedure(s)	Evidence Location
Art. 4	Definitions (users training)	GOVERN 4.1.1, 4.1.2	Training Curriculum (Appendix D)
Art. 9	Risk management system	GOVERN 1.4.1, 1.5.4, 1.6.2	Risk assessment templates (Appendices I, N)
Art. 10	Data governance	GOVERN 1.1.1, 1.3.1	Data governance in procedures
Art. 13	Transparency obligations	GOVERN 6.1.5, 6.1.6	Disclosure templates (Appendix P.1)
Art. 14	Human oversight	GOVERN 1.5.5, 1.5.6	Oversight procedures, automation bias
Art. 16	Provider obligations	GOVERN 1.1.1, 1.2.1, 2.1.1	Accountability structures
Art. 17	Quality management	GOVERN 5.1.1, 5.1.3, 5.1.4	QMS procedures, audits
Art. 71	Regulatory compliance	GOVERN 1.1.1	Compliance Register (Appendix A)
Art. 72	Post-market monitoring	GOVERN 1.5.5, 1.6.3	Monitoring procedures
Recital 60	Environmental sustainability	GOVERN 1.6 (all)	Environmental procedures and templates

Q.3: ISO 23894 Risk Management Extensions

How GOVERN procedures extend ISO 23894 with agentic AI and environmental sustainability:

ISO 23894 Clause	Base Requirement	AI RMF 2026 Extension	Implementation
5.1 Framework	Establish risk management framework	Add Agentic AI Committee, Environmental Sustainability Committee	GOVERN 1.5.1, 1.6.1; Appendices F, G
5.2 Policy	Define risk management policy	Mandate agentic risk assessment, sustainability metrics	GOVERN 1.5.4, 1.6.2; Appendices I, K
5.3 Roles	Assign roles and responsibilities	Agent Owner, Environmental Sustainability Officer	GOVERN 1.2.1; Section 3.2 roles table
5.4 Resources	Allocate resources	Agent monitoring tools, energy measurement systems	GOVERN 1.5.5, 1.6.3 procedures
5.5 Communication	Establish channels	Real-time agent incident reporting, sustainability dashboards	GOVERN 1.5.5, 1.6.6 procedures
4.1 Context	Understand context	Multi-agent ecosystem mapping, data center environmental context	GOVERN 1.5.1, 1.6.1 context analysis
4.2 Stakeholders	Understand stakeholders	Agent-affected parties, environmental impact communities	GOVERN 1.5.1, 1.6.1 stakeholder analysis

Q.4: Singapore MGF Four Dimensions Detailed Mapping

Detailed mapping of Singapore Model AI Governance Framework dimensions to specific procedure steps:

Dimension	Sub-Practice	GOVERN Procedure Step	Evidence
1. Assess & Bound Risks	Define action-space boundaries	GOVERN 1.5.3 Step 1-5	Agent boundaries documentation
1. Assess & Bound Risks	Conduct agent risk assessment	GOVERN 1.5.4 Step 1-7	Agent Risk Assessment (Appendix I)
2. Make Humans Accountable	Establish Agentic AI Committee	GOVERN 1.5.1 Step 1-7	Committee Charter (Appendix F)
2. Make Humans Accountable	Define Agent Owner role	GOVERN 1.2.1 Step 3	RACI Matrix, Roles table
2. Make Humans Accountable	Create accountability structures	GOVERN 2.1.1, 2.1.2	Decision Authority Matrix (Appendix C)
3. Implement Technical Controls	Agent identity management	GOVERN 1.5.2 Step 1-7	Agent Identity Framework (Appendix H)
3. Implement Technical Controls	Real-time agent monitoring	GOVERN 1.5.5 Step 1-6	Monitoring dashboards, alerts
4. Enable End-User Responsibility	Training on agent oversight	GOVERN 4.1.1, 4.1.2	Agentic AI training module (Appendix D)
4. Enable End-User Responsibility	Automation bias mitigation	GOVERN 1.5.6 Step 1-6	Training, procedures, checklists
4. Enable End-User Responsibility	Operator guidance	Appendix P.2	Operator Quick Reference template

Appendix R: Implementation Roadmap

Purpose: Provide structured, phased approach to implementing the AI RMF 2026 GOVERN function, with realistic timelines, resource estimates, and practical guidance for organizations at different maturity levels.

This roadmap helps organizations prioritize governance implementation efforts, allocate resources effectively, and build sustainable AI governance capabilities over time.

R.1: Quick Start Guide (First 90 Days)

For Organizations New to AI Governance

Focus on establishing foundational governance elements that provide immediate risk reduction and enable future expansion:

Week	Priority Actions	GOVERN Element	Deliverable
1-2	Stakeholder engagement, current state assessment	GOVERN 1.1	Initial compliance register
3-4	Define roles, establish AI Governance Board	GOVERN 1.2, 5.1	Charter, RACI matrix
5-6	Draft Trustworthy AI Policy	GOVERN 6.1	Policy document
7-8	Establish accountability structures	GOVERN 2.1	Decision authority matrix
9-10	Launch foundational training	GOVERN 4.1	Training materials
11-12	Pilot with one AI system, document lessons	All GOVERN	Pilot report, refinements

R.2: Phased Implementation Roadmap

Comprehensive 18-month implementation plan organized by phase:

Phase 1: Foundation (Months 1-3)

Objective: Establish core governance structures and critical processes

GOVERN Element	Key Activities	Completion Criteria
1.1 Legal/Regulatory	Complete compliance register	All applicable regulations identified and owned
1.2 Roles	Define all governance roles, create RACI	All roles documented and assigned
1.4 Risk Culture	Leadership commitment, risk priorities	Executive endorsement obtained
2.1 Accountability	Decision authority, escalation paths	Decision matrix approved
5.1 Oversight	Establish AI Governance Board	Board charter approved, first meeting held
6.1 Trustworthy AI	Draft policy covering 7 characteristics	Policy approved by CEO/Board

Resource Estimate: 2-3 FTEs (Chief AI Officer, Governance Lead, Legal/Compliance), External consultant support (20-40 hours)

Phase 2: Expansion (Months 4-6)

Objective: Build out comprehensive governance capabilities and operational procedures

GOVERN Element	Key Activities	Completion Criteria
1.3 DEIA	DEIA goals, team diversity initiatives	Goals set, initiatives launched
1.5 Agentic AI	Establish committee, agent identity framework	Committee operational, framework documented
1.6 Environmental	Establish committee, sustainability metrics	Committee operational, metrics defined
3.1 Diverse Teams	Team composition requirements, recruitment	Requirements documented, hiring adjusted
4.1 Training	Complete curriculum, launch programs	All modules available, 50% completion
All Templates	Customize Appendices A-P for organization	All templates tailored and in use

Resource Estimate: 4-5 FTEs (add Agent Owner, Environmental Sustainability Officer, Training Coordinator)

Phase 3: Maturity (Months 7-12)

Objective: Achieve operational excellence and prepare for certification

Capability	Key Activities	Completion Criteria
Internal Audit	Conduct first comprehensive audit	Audit complete, findings addressed
Management Review	Quarterly reviews established	2 reviews completed with action items
Metrics & KPIs	Implement governance dashboard	Dashboard operational, reporting to Board
Certification Prep	Gap analysis for ISO 42001	Gaps identified, remediation plan approved
Supplier Governance	Implement third-party assessments	All high-risk suppliers assessed
System Registry	Complete inventory of all AI systems/agents	Registry complete and current

Resource Estimate: 5-6 FTEs (add Internal Auditor, additional governance support)

Phase 4: Optimization (Months 13-18)

Objective: Continuous improvement and advanced capabilities

Capability	Key Activities	Completion Criteria
ISO Certification	Complete ISO 42001 certification audit	Certificate obtained
Continuous Improvement	Process optimization based on 12mo experience	Improvements documented and implemented
Advanced Monitoring	AI-powered governance monitoring tools	Automated monitoring operational
Benchmark & Share	Industry benchmarking, thought leadership	Benchmark report, conference presentation
Innovation	Pilot emerging governance capabilities	At least one pilot launched

Resource Estimate: Maintain 5-6 FTEs, add specialized expertise as needed

R.3: AI Governance Maturity Model

Assess your organization's current maturity level and chart progression:

Level	Governance	Processes	Tools	Culture	Typical Timeline
1: Ad Hoc	No formal governance	Informal, inconsistent	Basic tools only	Limited awareness	Starting point
2: Defined	Basic structures	Documented procedures	Standard templates	Growing awareness	0-6 months
3: Managed	Active committees	Consistent execution	Governance platform	Embedded practices	6-12 months
4: Measured	Data-driven oversight	Continuous monitoring	Integrated systems	Proactive culture	12-18 months
5: Optimized	Industry leadership	Continuous improvement	AI-powered gov	Innovation mindset	18+ months

R.4: Resource Planning Guide

R.4.1: Staffing Requirements by GOVERN Subcategory

GOVERN Subcategory	Primary Role	FTE Estimate	Support Roles
1.1 Legal/Regulatory	Compliance Lead	0.5 FTE	Legal, Risk team
1.2 Roles	Governance Lead	0.3 FTE	HR, Department heads
1.3 DEIA	DEIA Officer	0.5 FTE	HR, Recruitment
1.5 Agentic AI	Agent Owner(s)	1.0 FTE	Security, Engineering
1.6 Environmental	Sustainability Officer	0.5 FTE	Facilities, Finance
4.1 Training	Training Coordinator	0.5 FTE	L&D team
5.1 Oversight	Internal Auditor	0.5 FTE	Audit team
Overall Coordination	Chief AI Officer	1.0 FTE	Executive team
TOTAL	Core Governance Team	5-6 FTEs	Plus support roles

R.4.2: Technology and Tools Budget

Estimated technology investments for comprehensive AI governance:

Tool Category	Examples	Annual Cost Estimate
Governance Platform	OneTrust, ServiceNow, custom platform	\$50K - \$200K
Model Monitoring	Arize, Fiddler, WhyLabs	\$30K - \$100K
Energy Monitoring	CodeCarbon, ML CO2 Impact, custom	\$10K - \$50K
Bias Testing Tools	Fairlearn, AI Fairness 360, Aequitas	\$20K - \$75K
Document Management	SharePoint, Confluence, Notion	\$5K - \$20K
Training Platform	LMS (Docebo, Cornerstone)	\$10K - \$40K
TOTAL ESTIMATE	Varies by organization size and needs	\$125K - \$485K

R.4.3: Training Budget

Training Component	Scope	Cost Estimate
Content Development	Create all 9 training modules	\$40K - \$80K (one-time)
External Training	Certifications, workshops for governance team	\$15K - \$30K annually
Internal Delivery	Instructor time, facilities	\$20K - \$50K annually
Platform & Tools	LMS, assessment tools	Included in tech budget
TOTAL (Year 1)	Development + first year delivery	\$75K - \$160K
TOTAL (Ongoing)	Annual maintenance and delivery	\$35K - \$80K

R.5: Common Implementation Challenges and Solutions

Practical guidance for overcoming typical obstacles:

Challenge	Why It Happens	Solution
Executive buy-in lacking	Governance seen as bureaucratic overhead	Frame as risk mitigation and enabler. Show ROI: faster deployment, reduced incidents, certification benefits
Resource constraints	Competing priorities, budget limits	Start small (Phase 1 only), demonstrate value, build incrementally. Leverage existing roles initially
Resistance from AI teams	Fear of slowing innovation	Involve AI teams early, emphasize risk reduction not gatekeeping. Streamline approval processes for low-risk systems
Process complexity	Too many procedures, unclear workflows	Simplify for pilot phase. Create quick reference guides. Automate where possible with governance platform
Lack of expertise	New domain, specialized knowledge needed	Invest in training (Appendix D). Hire external consultants initially. Join industry working groups for peer learning
Tool integration issues	Multiple disparate systems	Start with manual processes, document integration requirements, implement governance platform in Phase 3
Keeping pace with AI evolution	Rapid technology changes	Build flexibility into governance (risk-based approach). Regular policy reviews. Monitor emerging risks quarterly

R.6: Implementation Success Metrics

Track progress with these key performance indicators:

Metric Category	Leading Indicator	Lagging Indicator	Target
Governance Structures	Committees established	Meeting attendance rates	100% by Month 3; >80% attendance
Policy & Procedures	Procedures documented	Procedure compliance rate	100% documented by M6; >90% compliance
Training	Training completion rate	Knowledge assessment scores	>85% completion; >80% pass rate
Risk Management	Risk assessments completed	Incident reduction	100% high-risk systems; 50% reduction
Oversight	Audits conducted	Audit findings closed	Annual audit; >90% closed in 90 days
Certification	Gap assessment complete	ISO 42001 certified	By Month 9; By Month 18
Culture	Employee surveys	AI risk awareness score	>75% positive sentiment

APPENDIX S: INDUSTRY-SPECIFIC ADAPTATIONS

S.1 PURPOSE AND STRUCTURE

This appendix provides tailored guidance for implementing the AI RMF 2026 GOVERN function across five key industry sectors. Organizations in these sectors face unique regulatory requirements, operational challenges, and AI use cases that necessitate sector-specific adaptations of the governance framework.

Purpose of This Appendix

The AI RMF 2026 GOVERN function establishes enterprise-wide governance structures for responsible AI development and deployment. While the core GOVERN procedures (1.1 through 6.1) provide comprehensive governance guidance applicable across all industries, practical implementation requires adaptation to sector-specific contexts including:

- Regulatory and legal frameworks unique to each sector
- Industry-specific AI use cases and deployment patterns
- Sector-particular risks and ethical considerations
- Specialized stakeholder requirements and accountability structures
- Domain-specific technical standards and best practices

How to Use This Appendix

Organizations should:

- Review the relevant sector section (S.2 through S.6) corresponding to their primary industry
- Identify sector-specific regulations and map them to existing GOVERN procedures
- Assess their AI systems against the high-risk use cases identified for their sector
- Apply the sector-specific GOVERN procedure adaptations when implementing the framework
- Customize templates from Appendices A-P using the modifications specified
- Prioritize implementation activities based on the sector-specific roadmap provided

Structure Overview

This appendix covers five key sectors:

- S.2: Healthcare and Life Sciences
- S.3: Financial Services
- S.4: Government and Public Sector
- S.5: Manufacturing and Industrial
- S.6: Retail and E-Commerce

Each sector section includes:

- Sector Overview - AI adoption context and unique governance challenges
- Sector-Specific Regulations Table - Comprehensive regulatory mapping
- High-Risk AI Use Cases Table - Risk assessment and control requirements
- GOVERN Procedure Adaptations - Sector-specific implementation guidance
- Template Modifications - Customizations for sector-specific needs
- Implementation Priorities - 90-day roadmap for getting started

S.2 HEALTHCARE AND LIFE SCIENCES

S.2.1 Sector Overview

The healthcare and life sciences sector has emerged as one of the most active areas for AI innovation, with applications ranging from clinical decision support and medical imaging analysis to drug discovery and personalized treatment planning. AI systems in healthcare directly impact patient safety, health outcomes, and fundamental rights to health and privacy.

Key AI Use Cases

- Clinical Decision Support Systems (CDSS) - Diagnosis and treatment planning assistance
- Medical Imaging Analysis - Radiology, pathology, and diagnostic image interpretation
- Drug Discovery and Development - Molecular analysis and clinical trial optimization
- Patient Monitoring and Predictive Analytics - Sepsis prediction, readmission forecasting
- Precision Medicine - Genomic analysis for personalized treatment
- Administrative AI - Scheduling, resource allocation, documentation
- Remote Monitoring - Telehealth and chronic disease management

Unique Governance Challenges

- Patient Safety - Direct harm potential requires rigorous validation
- Regulatory Complexity - FDA approval pathways for medical devices
- Clinical Validation - Evidence requirements for clinical settings
- Explainability - Physicians need interpretable outputs
- Privacy Requirements - HIPAA and state medical privacy laws
- Health Equity - Risk of exacerbating healthcare disparities

S.2.2 Sector-Specific Regulations

Regulation	Jurisdiction	Key Requirements	GOVERN Procedures
HIPAA	US Federal	Secure PHI, Business Associate Agreements, privacy assessments	GOVERN 1.1.1, 6.1.2, Appendix L
FDA SaMD Guidance	US Federal	FDA clearance for medical devices, change control plans	GOVERN 1.1.1, 2.1.2, 5.1
21 CFR Part 11	US Federal	Trustworthy electronic records, audit trails, validation	GOVERN 2.1, 5.1.4, 6.1.3
EU MDR 2017/745	European Union	CE marking, clinical evaluation, post-market surveillance	GOVERN 1.1.1, 5.1.5, Appendix O
State Privacy Laws	US States	Enhanced privacy protections, patient consent requirements	GOVERN 1.1.1, 6.1.2
CLIA	US Federal	Lab-developed test certification, quality control	GOVERN 4.1.2, 5.1, 6.1.5
GDPR Article 22	European Union	Right to human review, explanation requirements	GOVERN 5.1.1, 6.1.6, Appendix O
ISO 13485	International	Quality management for medical device AI	GOVERN 5.1, 6.1

S.2.3 High-Risk AI Use Cases

Use Case	Risk Level	Primary Risks	Sector-Specific Controls
Clinical Decision Support - Cancer Treatment	Critical	Patient harm, liability, diagnostic bias	FDA clearance, physician review, fairness testing, audit logs
Medical Imaging Diagnosis	High	False negatives/positives, automation bias	FDA clearance, double-reading protocols, performance monitoring
Patient Deterioration Prediction	Critical	Missed life-threatening conditions, alert fatigue	Clinical Safety Committee review, human acknowledgment required
AI-Enabled Robotic Surgery	Critical	Surgical complications, system malfunctions	Surgeon training/credentialing, real-time oversight, manual override
Drug Discovery AI	High	Invalid predictions, unsafe candidates progressing	Validation against historical trials, expert scientific review
Genomic Treatment Recommendations	High	Privacy risks, ancestry-based disparities	Enhanced privacy controls, fairness testing, genetic counselor involvement
AI Triage Systems	High	Incorrect triage, access barriers	Validation against nurse triage, escalation protocols, equity monitoring
Readmission Prediction Models	High	Discriminatory resource allocation	Disparate impact analysis, case manager review, appeals process

S.2.4 GOVERN Procedure Adaptations

GOVERN 1.1.1: Legal and Regulatory Compliance Register

Healthcare-Specific Considerations: Healthcare presents one of the most complex regulatory landscapes. Organizations must navigate FDA, HIPAA, CLIA, state laws, and international regulations.

Modified Procedure Steps:

- Establish multidisciplinary compliance working group (regulatory affairs, legal, clinical, IT, AI teams)
- Create tiered compliance register: Tier 1 (FDA medical devices), Tier 2 (HIPAA systems), Tier 3 (clinical support), Tier 4 (administrative)
- Document medical device determination, HIPAA applicability, state requirements for each system
- Quarterly reviews tracking FDA guidance changes, new legislation, practice guideline updates
- Maintain evidence files: regulatory determinations, validation records, clinical evidence, performance monitoring

GOVERN 2.1.2: AI System Approval Requirements

Healthcare-Specific Considerations: Approvals must account for clinical risk and patient impact. Clinical representation required in approval authority.

Modified Procedure Steps:

- Tiered approval: Level 1 (Critical) requires AI Governance + Clinical Safety Committees; Level 2 (High) requires Clinical Leadership + AI Governance
- For Level 1-2: Complete AIIA, clinical validation study, patient safety assessment, health equity analysis, privacy assessment
- Establish Clinical Safety Committee with CMO, specialty reps, patient safety officer, quality director, informaticist
- For FDA devices: Integrate regulatory approval process, obtain clearance before operational deployment
- Require physician champion for all clinical AI systems
- Mandatory post-deployment monitoring: 30-day safety review (Critical), 90-day validation, ongoing monitoring

GOVERN 5.1: AI Oversight Mechanisms

Healthcare-Specific Considerations: Oversight must emphasize clinical outcomes, patient safety, and health equity. Real-time alerting for performance degradation required.

Modified Procedure Steps:

- Clinical AI Performance Dashboard: accuracy metrics, patient safety indicators, equity metrics, workflow metrics, physician satisfaction, outcomes
- Automated alerting: accuracy drops >5%, fairness disparities >10%, adverse event patterns, override rates >30%, system unavailability
- Quarterly Clinical AI Performance Reviews: metric trends, AI vs non-AI outcomes, equity dashboard, physician feedback, safety events
- Patient Safety Event Investigation: immediate investigation when AI implicated, root cause analysis, FDA reporting assessment
- Risk-appropriate human oversight: Critical (mandatory physician approval), High (physician review with override), Moderate (notification)
- Annual comprehensive reviews: external expert review, guidelines comparison, approval criteria validation

S.2.5 Template Modifications

AI System Inventory (Appendix M) - Additional Fields:

- FDA Medical Device Classification (Class I/II/III/Not a Device)
- FDA Clearance/Approval Status (510(k), PMA, De Novo number)
- Clinical Specialty (primary department)
- PHI Processing (Yes/No)
- HIPAA Business Associate Agreement Status
- Clinical Validation Status (study type)
- Patient-Facing (Yes/No)
- EHR Integration Level

AI Impact Assessment (Appendix O) - Additional Sections:

- Section 4.5: Patient Safety Assessment - risks, severity, likelihood, safeguards, investigation process
- Section 4.6: Clinical Validation - evidence, validation status, outcomes metrics, standard of care comparison
- Section 4.7: Health Equity Analysis - training data demographics, performance by population, disparities, mitigation
- Section 4.8: Regulatory Compliance - FDA determination, clearance status, HIPAA requirements, quality standards

Supplier Governance Toolkit (Appendix L) - Healthcare Requirements:

- FDA Registration and Listing status
- ISO 13485 certification
- Clinical validation evidence (publications, studies)
- HIPAA Business Associate Agreement
- Training data diversity documentation
- Intended use statement (FDA-compliant)
- Contraindications and limitations
- Clinical support availability
- Adverse event reporting processes
- Model updating and revalidation approach

S.2.6 Implementation Priorities: First 90 Days

Days 1-30: Foundation and Regulatory Mapping

- Establish Clinical AI Governance Structure: Form committees, assign CMO champion, secure executive sponsorship
- Complete AI System Inventory: Identify all systems, classify by FDA status, document PHI processing, prioritize by risk
- Conduct Regulatory Gap Analysis: Review FDA compliance, identify systems needing clearance, assess HIPAA compliance, map state laws
- Establish Patient Safety Integration: Integrate AI tracking into safety reporting, train safety staff, create escalation procedures

Days 31-60: Risk Assessment and Policy Development

- Conduct High-Risk AI Assessments: Complete AIAs for critical systems, perform clinical validation reviews, conduct equity audits
- Develop Healthcare AI Policies: Adapt Trustworthy AI Policy, create clinical approval policy, establish equity policy, develop explainability standards
- Initiate Clinical Stakeholder Engagement: Physician focus groups, nursing leadership involvement, patient advocacy engagement

Days 61-90: Monitoring and Training

- Implement Clinical AI Performance Monitoring: Deploy dashboard, establish automated alerting, create equity reports, schedule quarterly reviews
- Launch Healthcare AI Training: Train governance committees, provide clinical leadership training, physician education, patient safety staff training
- Address Priority Compliance: Remediate regulatory gaps, execute BAAs with vendors, enhance monitoring for fairness concerns, develop FDA submission strategy

S.2.7 Healthcare Case Study: Clinical AI Deployment

A large academic medical center deployed an AI-powered sepsis prediction system following the governance framework outlined in this section.

- **Initial Assessment:** The hospital completed a comprehensive AI Impact Assessment with healthcare sections from S.2.5. The Patient Safety Assessment identified sepsis as critical-risk. Clinical Validation showed 92% sensitivity and 88% specificity. Health Equity Analysis revealed 7% performance disparity between racial groups requiring mitigation.
- **Approval Process:** Following Level 1 (Critical Risk) pathway from S.2.4, the system underwent review by AI Governance and Clinical Safety Committees. Key concerns included alert fatigue risk, automation bias potential, and equity gaps.
- **Mitigation Measures:** Implemented (1) Alert threshold calibrated to 20% PPV; (2) Mandatory nurse acknowledgment; (3) Prospective pilot study; (4) Algorithm retraining with fairness constraints reducing disparity to 3%; (5) Detailed explainability.
- **Post-Deployment Monitoring:** Established monitoring per S.2.4 including monthly accuracy metrics, quarterly equity reviews, patient safety tracking. After six months, demonstrated 15% reduction in sepsis mortality with sustained performance.

S.3 FINANCIAL SERVICES

S.3.1 Sector Overview

Financial services has been an early and aggressive adopter of AI for credit underwriting, fraud detection, algorithmic trading, customer service, risk management, and anti-money laundering. AI systems directly impact access to credit, employment, housing, and economic opportunity, making fairness paramount. The sector faces intense regulatory scrutiny spanning consumer protection, fair lending, model risk management, anti-discrimination, privacy, and market integrity.

Key AI Use Cases

- Credit Underwriting - Credit scoring, loan approval, pricing decisions
- Fraud Detection - Transaction monitoring, account takeover prevention
- Algorithmic Trading - Automated trading, portfolio optimization, robo-advisors
- Customer Service - Chatbots, product recommendations, personalization
- AML/KYC - Transaction monitoring, suspicious activity detection
- Risk Management - Credit risk, market risk, stress testing, compliance monitoring
- Process Automation - Back-office RPA, document processing

Unique Governance Challenges

- Fair Lending - ECOA and Fair Housing Act compliance requirements
- Explainability - Adverse action notices require specific reasons
- Model Risk Management - SR 11-7 validation expectations
- Systemic Risk - Correlated strategies could amplify market instability
- Cybersecurity - Adversarial attacks on fraud detection
- Consumer Protection - UDAAP concerns
- Data Privacy - GLBA, FCRA, GDPR, state privacy laws
- Regulatory Fragmentation - Multiple regulators with different expectations

S.3.2 Sector-Specific Regulations

Regulation	Jurisdiction	Key Requirements	GOVERN Procedures
ECOA / Regulation B	US Federal	Prohibits credit discrimination, requires adverse action notices, disparate impact testing	GOVERN 1.1.1, 1.3, 6.1.6, 6.1.8
Fair Credit Reporting Act	US Federal	Regulates consumer data use, accuracy requirements, factor disclosure	GOVERN 1.1.1, 6.1.2, 6.1.6
GLBA	US Federal	Protects customer financial information, safeguards requirements	GOVERN 1.1.1, 6.1.2, Appendix L
SR 11-7 / OCC 2011-12	US Federal	Model risk management framework, independent validation required	GOVERN 2.1, 5.1, 6.1.1, 6.1.5
Dodd-Frank Act	US Federal	Algorithmic trading controls, qualified mortgage standards	GOVERN 1.1.1, 5.1
GDPR Article 22	European Union	Right not to be subject to automated decisions, explanation rights	GOVERN 1.1.1, 5.1.1, 6.1.6
Regulation Best Interest	US Federal	Broker-dealers must act in customer best interest	GOVERN 1.1.1, 2.1, 6.1.1
State Fair Lending Laws	US States	State-specific protections, additional protected classes	GOVERN 1.1.1, 1.3, 6.1.8

S.3.3 High-Risk AI Use Cases

Use Case	Risk Level	Primary Risks	Sector-Specific Controls
Credit Scoring and Loan Approval	Critical	Discrimination, fair lending violations, explainability	Disparate impact testing, fairness constraints, independent validation, ECOA-compliant reasons
Fraud Detection Systems	High	False positives, bias, adversarial attacks	Risk-based thresholds, demographic monitoring, appeals process, adversarial testing
Algorithmic Trading	Critical	Market manipulation, systemic risk, losses	Pre-trade controls, kill switch, oversight, backtesting, regulatory registration
Robo-Advisors	High	Unsuitable advice, fiduciary duty violations	Suitability assessments, Reg BI compliance, disclosures, human advisor access
AML Transaction Monitoring	High	False positives, false negatives, bias, penalties	Tuning process, human investigation, bias assessment, SAR filing compliance
Dynamic Pricing	High	Discrimination, unfairness, UDAAP violations	Fairness testing, prohibited variables, transparency, appeals, disparate impact monitoring
Customer Service Chatbots	High	Inaccurate advice, unauthorized transactions, bias	Clear AI disclosure, limited authority, monitoring, human escalation, demographic analysis
Employment Screening	High	Discrimination, FCRA violations, adverse impact	Adverse impact analysis, EEOC validation, adverse action notices, appeals

S.3.4 GOVERN Procedure Adaptations

GOVERN 1.3: Diversity, Equity, Inclusion, and Accessibility

Financial Services Considerations: AI systems in lending, employment, and housing trigger fair lending laws with legal obligations beyond DEIA commitments. Even models without protected class variables may discriminate through proxies.

Modified Procedure Steps:

- Establish Fair Lending Compliance Program: Designate Fair Lending Officer, create Fair Lending Committee, integrate AI fairness
- Pre-Deployment Fair Lending Testing: Obtain demographic data, calculate comparative statistics, apply 80% rule, document justifications
- Fairness-Aware Development: Remove protected variables, identify proxies, apply fairness constraints, compare models
- Ongoing Fairness Monitoring: Quarterly demographic reporting, identify disparities, monitor for proxy discrimination, annual examination
- Human Review for Borderline Decisions: Route near-threshold applications to underwriters
- Transparency and Contestability: ECOA-compliant adverse actions, customer inquiry process, reconsideration mechanism
- Engage Diverse Stakeholders: Include diversity in development, consult community groups, consumer testing, board reporting

GOVERN 5.1.2: Independent Review and Validation

Financial Services Considerations: SR 11-7 and OCC 2011-12 establish comprehensive model risk management expectations. Independent validation by qualified validators required.

Modified Procedure Steps:

- Establish Model Risk Management Framework: Create policy, establish independent validation team, form Model Risk Committee
- Create AI Model Inventory: Identify all models, classify by risk tier, document purpose, assign ownership
- Require Validation Before Deployment: Conceptual soundness (technique appropriateness, data quality, discrimination assessment); Ongoing monitoring (metrics, frequency, triggers); Outcomes analysis (performance testing, backtesting, subpopulation evaluation)
- Document Validation Findings: Prepare comprehensive reports, identify limitations, classify findings, require management response

- Establish Model Approval Process: Route to Model Risk Committee, require approval, document conditions, escalate high-risk models
- Implement Ongoing Monitoring: Monitor per recommendations, conduct performance reviews, trigger revalidation when needed, annual revalidation for high-risk
- Address AI-Specific Challenges: Validate using inputs/outputs for black-box, use explainability techniques, conduct adversarial testing, review hyperparameter tuning

GOVERN 6.1.6: Explainability and Interpretability

Financial Services Considerations: ECOA requires "specific reasons" for credit denials. FCRA requires credit score factor disclosure. Model risk management requires understanding limitations.

Modified Procedure Steps:

- Establish Use Case Requirements: Credit (generate ECOA-compliant reasons), Fraud (internal explanations, customer communication), Investment (clear factors, suitability support), Validation (conceptual soundness support)
- Implement Adverse Action Reason Generation: Deploy explainability techniques (SHAP, LIME), map to ECOA reason codes, validate accuracy, test across profiles
- Establish Explanation Validation: Validate faithfulness to model, test stability, assess quality with reviewers, verify no protected class disclosure
- Provide Tiered Explanations: Customer-facing (plain language, actionable factors, context), Employee (feature importance, confidence, comparisons), Regulator (comprehensive documentation, validation evidence)
- Address Complex Models: Use post-hoc techniques for deep learning, consider interpretable models for high-stakes, develop proxy models, document trade-offs
- Train Staff: Customer service on explaining reasons, underwriters on interpreting outputs, compliance on ECOA assessment
- Document for Regulators: Document methodology in validation, maintain adverse action evidence, prepare for examinations

S.3.5 Template Modifications

AI Impact Assessment (Appendix O) - Additional Sections:

- Section 4.5: Fair Lending Assessment - credit decision influence, protected characteristics in data, disparate impact analysis, business justification, proxy variables
- Section 4.6: Model Risk Assessment - risk classification, validation status, limitations, assumptions, degradation scenarios, compensating controls
- Section 4.7: Regulatory Compliance - applicable regulators, specific regulations, adverse action notices, disclosure requirements, examination readiness
- Section 4.8: Explainability Assessment - reason generation capability, adverse action validation, appeal process, regulator explanations, techniques used

AI Approval Authority Matrix (GOVERN 2.1.2) - Modified Tiers:

- Tier 1 (High Risk): Material credit/financial impact - Model Risk Committee, CRO, Board Risk Committee approval; Full independent validation, annual revalidation
- Tier 2 (Moderate Risk): Moderate impact - Model Risk Committee, Business Unit approval; Independent validation, revalidation every 2-3 years
- Tier 3 (Low Risk): Limited impact - Model Risk Manager, Business Unit approval; Targeted validation, periodic reviews

AI Governance Committee Charter - Financial Services Additions:

- Membership: Chief Risk Officer (required), Fair Lending Officer (required), Model Validation Head
- Responsibilities: Oversee fair lending compliance, review model validation results, escalate risks to Board, review examination findings, approve model risk policies, monitor against risk appetite
- Reporting: Quarterly to Board Risk Committee, notification to regulators for material issues, annual effectiveness attestation

S.3.6 Implementation Priorities: First 90 Days

Days 1-30: Foundation and Regulatory Assessment

- Establish AI Governance Integrated with Model Risk: Form AI Governance Committee with CRO/Fair Lending Officer, clarify relationship with Model Risk Committee
- Complete AI Model Inventory: Identify all models in credit/fraud/trading/AML, classify by risk tier, identify validation gaps
- Conduct Fair Lending Gap Analysis: Identify credit decision models, assess disparate impact testing, review adverse action notices, identify proxy variables
- Assess Model Validation Status: Identify validated models, review validation reports, identify production models without validation, assess team capability

Days 31-60: Risk Assessment and Policy Development

- Conduct High-Risk AI Assessments: Complete AIAs for credit/fraud/trading models, perform fairness testing, assess explainability, identify compliance gaps
- Develop Financial Services AI Policies: Adapt Trustworthy AI Policy with model risk requirements, create Fair Lending Policy, establish Explainability Standards, develop Model Risk Management Policy for AI/ML, create AI Vendor Management Policy
- Initiate Regulatory Engagement Preparation: Prepare examination documentation, develop model inventory for regulators, identify concern areas, schedule regulator meetings

Days 61-90: Monitoring and Validation

- Implement AI Model Performance Monitoring: Deploy dashboards, establish alerts for degradation, create fair lending monitoring, schedule Model Risk Committee reviews
- Launch Financial Services AI Training: Train governance and model risk committees, provide validation team AI/ML training, fair lending training for data scientists, customer service training on AI decisions
- Initiate Independent Validation: Begin validation of highest-risk models, engage external validators if needed, prioritize credit/fraud/trading, establish validation work plan for 12-18 months

S.3.7 Financial Services Case Study: Fair Lending Analysis

A regional bank implemented AI credit decisioning for auto loans with comprehensive fair lending analysis per GOVERN 1.3.

- **Pre-Deployment Testing:** Fair Lending Officer analysis revealed disparities - African American applicants 15 points lower approval (72% vs 87%); Hispanic applicants 8 points lower (79% vs 87%).
- **Root Cause Analysis:** Model heavily weighted stability variables (residence/employment length) correlating with race; ZIP code income estimates as race proxies; Education level disadvantaging first-generation students.
- **Mitigation:** (1) Removed granular ZIP codes and education; (2) Engineered alternative stability measures; (3) Applied fairness constraints (equalized odds); (4) Developed challenger models. Fairness model reduced disparities to 4 points with 88% accuracy vs 91% unconstrained.
- **Ongoing Monitoring:** Quarterly fair lending reports to Fair Lending Committee. First year showed sustained improvements with <5 point disparities. Human review implemented for borderline decisions.

S.4 GOVERNMENT AND PUBLIC SECTOR

S.4.1 Sector Overview

Government and public sector organizations increasingly deploy AI for benefits eligibility, law enforcement, regulatory compliance, public service delivery, infrastructure management, and fraud detection. AI systems affect fundamental rights including liberty, due process, equal protection, privacy, and access to government services.

- **Key Use Cases:** Benefits eligibility determination, predictive policing, regulatory enforcement, public service chatbots, traffic management, program fraud detection.
- **Unique Challenges:** Constitutional rights protection, transparency obligations, administrative procedure requirements, disparate impact on marginalized communities, public trust, procurement constraints, vendor transparency issues.

S.4.2 Sector-Specific Regulations

Regulation	Jurisdiction	Key Requirements	GOVERN Procedures
Administrative Procedure Act	US Federal	Notice and comment for rules, rational basis for decisions, judicial review	GOVERN 1.1.1, 6.1.6, 5.1.1
Freedom of Information Act	US Federal	Public access to records, algorithm transparency requirements	GOVERN 1.1.1, 5.1.4, 6.1.6
Constitutional Due Process	US Federal/State	Notice, hearing, impartial decision-maker for rights-affecting decisions	GOVERN 5.1.1, 6.1.6, Appendix O
Constitutional Equal Protection	US Federal/State	No discrimination, disparate impact analysis, strict scrutiny for race	GOVERN 1.3, 6.1.8, Appendix O
Canadian Directive on ADM	Canada Federal	Algorithmic Impact Assessments, accountability, transparency, human review	GOVERN 2.1.2, Appendix O, 6.1.6
EU AI Act (Public Sector)	European Union	Most government AI is high-risk, biometric restrictions, transparency registers	GOVERN 1.1.1, 5.1, 6.1.1, Appendix O
State/Local AI Laws	US State/Local	Impact assessments, public notice, accountability, bias audits	GOVERN 1.1.1, 6.1.8, Appendix O
OMB Memo on AI	US Federal	Minimum practices for federal agencies, impact assessments, monitoring	GOVERN 1.1.1, 5.1, Appendix O

S.4.3 High-Risk Use Cases

- **Benefits Eligibility:** Civil rights, errors causing hardship, bias - Require human review of denials, plain language explanations, robust appeals, disparate impact analysis
- **Predictive Policing:** Over-policing minorities, constitutional concerns - Civil rights assessment, prohibit race as input, community oversight, transparency reports
- **Bail/Sentencing Risk:** Liberty impacts, racial bias, due process - Full disclosure to defendants, contestability, independent validation, judicial training
- **Facial Recognition:** Privacy, bias, surveillance, false positives - Legislative authorization, strict use limits, no mass surveillance, accuracy thresholds, demographic testing
- **Fraud Detection:** False accusations, bias, due process - High confidence thresholds, human investigation, clear notification, appeals process
- **Employment Screening:** Discrimination, equal opportunity - Adverse impact analysis, validation, appeals, transparency, oversight

S.4.4 Key Procedure Adaptations

- **GOVERN 1.1.2:** Conduct constitutional rights analysis, assess APA applicability, identify transparency obligations, evaluate procurement law compliance.
- **GOVERN 5.1.1:** Establish human oversight based on rights impact - Level 1 (liberty/constitutional rights): human decision, AI advisory only; Level 2 (significant interests): human review before automated decision; Level 3 (admin processes): exception-based review; Level 4 (minimal risk): human appeal available.
- **GOVERN 6.1.6:** Multi-layered explanations for individuals (plain language), administrative review (comprehensive), and public transparency (high-level).
- **GOVERN 6.1.8:** Conduct civil rights impact assessment, establish robust fairness testing, engage affected communities, implement independent oversight.

S.4.5 Government Template Modifications

AI Impact Assessment - Government Additions: Constitutional Rights Analysis, Transparency and Public Records requirements, Community Impact Assessment.

S.4.6 Implementation Priorities: First 90 Days

- **Days 1-30:** Establish governance with constitutional/transparency focus. Complete inventory. Conduct constitutional rights gap analysis. Assess FOIA compliance.
- **Days 31-60:** High-risk assessments. Develop policies for government context. Stakeholder engagement with civil rights groups.
- **Days 61-90:** Public accountability monitoring. Training on constitutional requirements. Address compliance gaps. Establish public AI register.

S.5 MANUFACTURING AND INDUSTRIAL

S.5.1 Sector Overview

Manufacturing and industrial sectors deploy AI for predictive maintenance, quality control, supply chain optimization, autonomous robotics, and production scheduling. AI systems directly affect workplace safety, product quality, supply chain resilience, and operational efficiency.

Key Use Cases: Predictive maintenance, automated quality inspection, supply chain optimization, autonomous robotics, production scheduling, energy optimization.

Unique Challenges: Workplace safety risks, product liability, real-time decision requirements, operational technology security, human-robot collaboration, equipment damage potential.

S.5.2 Sector-Specific Regulations

Regulation	Jurisdiction	Key Requirements	GOVERN Procedures
OSHA Safety Standards	US Federal	Workplace safety for AI-controlled machinery, hazard communication	GOVERN 1.1.1, 6.1.1, 5.1
ISO 9001	International	Quality management system, process control, continuous improvement	GOVERN 5.1, 6.1
ISO 45001	International	Occupational health and safety management	GOVERN 1.1.1, 5.1
Machinery Safety Directives (EU 2006/42/EC)	European Union	Essential health and safety requirements for machinery	GOVERN 1.1.1, 6.1.1
Product Liability Laws	US Federal/State	Liability for defective products, warning requirements	GOVERN 1.1.1, 6.1.1
ISO 10218 (Robotics Safety)	International	Safety requirements for industrial robots, collaborative robots	GOVERN 1.1.1, 5.1
IEC 62443 (Industrial Cybersecurity)	International	Security for industrial automation and control systems	GOVERN 6.1.2, 6.1.1

S.5.3 High-Risk Use Cases

- **Predictive Maintenance:** Equipment failure, safety hazards - Safety margins in predictions, human verification, backup systems
- **Autonomous Robotics:** Worker injury, collision risks - Safety zones, human override, redundant sensors, emergency stops, training
- **Quality Control AI:** Defective products, recalls - Statistical validation, human verification for critical defects, root cause analysis
- **Supply Chain Optimization:** Production disruptions, shortages - Scenario testing, human oversight for critical decisions, contingency planning
- **Energy Management AI:** Production disruptions, safety risks - Operating bounds, human monitoring, override capability

S.5.4 Key Procedure Adaptations

- **GOVERN 6.1.1:** Incorporate safety-first principles, conduct hazard analysis, establish safety integrity levels, implement functional safety standards.
- **GOVERN 5.1:** Deploy real-time monitoring dashboards, establish safety alert systems, conduct regular safety audits, maintain incident investigation protocols.
- **GOVERN 4.1:** Provide safety-specific training, human-robot collaboration training, emergency procedures, hazard recognition.

S.5.5 Template Modifications

AI Impact Assessment - Manufacturing: Safety Risk Assessment (hazards, SIL, functional safety standards), Operational Continuity (failure impacts, backups, emergency procedures).

S.5.6 Implementation Priorities: First 90 Days

- **Days 1-30:** Safety-focused governance. Inventory with safety classifications. Hazard analysis. OSHA compliance.
- **Days 31-60:** Safety risk assessments. Safety-first policies with functional safety standards. Safety monitoring dashboards.
- **Days 61-90:** Real-time safety monitoring. Safety-specific training. Safety audits. Incident investigation protocols.

S.6 RETAIL AND E-COMMERCE

S.6.1 Sector Overview

Retail and e-commerce sectors leverage AI for personalized recommendations, dynamic pricing, inventory optimization, customer service chatbots, and demand forecasting. AI systems influence purchasing decisions, pricing fairness, data privacy, and consumer trust.

Key Use Cases: Product recommendations, dynamic pricing, inventory management, customer service chatbots, demand forecasting, fraud detection, visual search.

Unique Challenges: Consumer privacy, discriminatory pricing concerns, manipulative design, data breaches, advertising accuracy, third-party data sharing.

S.6.2 Sector-Specific Regulations

Regulation	Jurisdiction	Key Requirements	GOVERN Procedures
GDPR	European Union	Consent requirements, data minimization, purpose limitation, right to explanation	GOVERN 1.1.1, 6.1.2, 6.1.6
CCPA/CPRA	California	Consumer data rights, opt-out of sale, deletion rights, data use disclosures	GOVERN 1.1.1, 6.1.2
FTC Act Section 5	US Federal	Prohibits unfair or deceptive practices, algorithmic transparency expectations	GOVERN 1.1.1, 6.1.1
Consumer Protection Laws	US Federal/State	False advertising, pricing accuracy, refund policies	GOVERN 1.1.1, 6.1.1
Advertising Standards	US Federal/State	Truth in advertising, substantiation requirements, endorsement disclosures	GOVERN 1.1.1, 6.1.1
PCI DSS	International	Payment card data security, secure AI access to payment information	GOVERN 6.1.2, 6.1.1

Accessibility Laws (ADA)	US Federal	Website accessibility, AI chat accessibility for disabilities	GOVERN 1.3, 6.1.1
--------------------------	------------	---	-------------------

S.6.3 High-Risk Use Cases

- **Personalized Recommendations:** Privacy violations, filter bubbles - Consent management, data minimization, transparency about logic, user controls
- **Dynamic Pricing:** Discriminatory pricing, unfairness perceptions - Fairness testing, prohibited variables, transparency, business justification, complaint monitoring
- **Customer Service Chatbots:** Inaccurate information, accessibility barriers - Clear AI disclosure, human escalation, accuracy monitoring, accessibility testing
- **Inventory/Demand Forecasting:** Stockouts, excess inventory - Scenario testing, human oversight, performance monitoring
- **Fraud Detection:** False positives, customer friction - Balanced thresholds, clear explanation, quick resolution, appeals

S.6.4 Key Procedure Adaptations

- **GOVERN 6.1.2:** Implement consent management, data minimization practices, retention policies, third-party data governance.
- **GOVERN 6.1.6:** Provide clear explanations of recommendations, pricing factors, personalization logic; enable user controls.
- **GOVERN 6.1.8:** Test for pricing discrimination, monitor customer complaints for bias patterns, provide fair treatment regardless of demographics.
- **GOVERN 1.3:** Ensure accessibility of AI-powered customer interfaces, provide alternative access methods, test with assistive technologies.

S.6.5 Template Modifications

AI Impact Assessment: Add consumer protection assessment, pricing fairness analysis, advertising accuracy review, accessibility evaluation.

Data Governance Policy: Add consent management, cookie management, third-party data sharing controls, retention schedules.

S.6.6 Implementation Priorities

- **Days 1-30:** Inventory AI systems, assess GDPR/CCPA compliance, conduct privacy gap analysis.
- **Days 31-60:** Develop consumer-facing AI policies, implement consent management, conduct pricing fairness assessments.
- **Days 61-90:** Deploy privacy monitoring, launch consumer transparency initiatives, establish complaint tracking.

S.6.7 Retail Case Study: Privacy-Preserving Personalization

A major e-commerce platform implemented AI recommendations while maintaining GDPR/CCPA compliance through privacy-by-design.

Privacy Assessment:

- GDPR Article 35 DPIA identified: 50M+ user profiling, cross-device tracking, third-party data sharing, potential discriminatory recommendations.
- Privacy-Preserving Design:
 - (1) Data minimization;
 - (2) Granular consent management;
 - (3) Federated learning;
 - (4) Differential privacy;
 - (5) 12-month data retention.
- **Transparency:** Clear recommendation explanations, user controls, privacy dashboard, human review rights, transparency reports.
- **Compliance Outcomes:** GDPR compliant with 78% user opt-in. Regulators cited as positive example. Demonstrated commercial AI aligns with strong privacy.

CONCLUSION

This appendix provides sector-specific adaptations of the AI RMF 2026 GOVERN function for five key industries.

Organizations should:

- Begin with their primary industry section to understand unique regulatory and governance requirements
- Apply sector-specific modifications to GOVERN procedures and templates
- Follow the 90-day implementation roadmap to establish foundational governance
- Adapt guidance to their organization's specific context, size, and AI maturity
- Regularly review and update sector-specific practices as regulations and best practices evolve
- Consider engaging sector-specific experts (legal counsel, compliance professionals, industry consultants) for complex implementations

For organizations operating across multiple sectors, implement the most stringent requirements and maintain separate documentation for different business units and regulatory contexts.

This appendix should be used in conjunction with:

- Core GOVERN procedures (1.1 through 6.1) in the main Procedural Manual
- Appendices A-P providing templates, frameworks, and tools
- Appendix T for KPIs and metrics specific to each sector
- Appendix U for case studies demonstrating sector-specific governance in practice

Appendix T: AI Governance KPI Dashboard

Comprehensive Metrics Framework for Measuring AI Governance Effectiveness

T.1 Purpose and Structure

This appendix provides a comprehensive framework for measuring AI governance effectiveness through key performance indicators (KPIs). The KPI Dashboard enables organizations to monitor compliance, track outcomes, provide data-driven insights, benchmark performance, enable continuous improvement, and support regulatory reporting.

Structure Overview

- T.2: KPI Framework Overview
- T.3: GOVERN Function KPIs (all 11 subcategories)
- T.4: Executive Dashboard Template
- T.5: Board Reporting Template
- T.6: Governance Maturity Scoring
- T.7: Benchmark Ranges

T.2 KPI Framework Overview

T.2.1 Leading vs. Lagging Indicators

Leading indicators measure activities and processes that drive outcomes (e.g., training completion rates, risk assessments completed). Lagging indicators measure results and impacts (e.g., incidents, violations, audit findings).

T.2.2 Measurement Frequency

- **Monthly:** Operational activities requiring regular monitoring
- **Quarterly:** Governance outcomes and strategic activities
- **Annual:** Strategic effectiveness and long-term trends

T.2.3 Data Collection and Governance

Effective KPI measurement requires robust data collection processes and clear data governance. Organizations should establish comprehensive data sources, quality standards, and clear roles.

Data Sources and Systems

- AI inventory systems tracking all AI systems and their governance status
- Learning management systems (LMS) for training completion data
- Incident management systems recording AI-related issues
- Document management systems tracking governance documentation
- Committee management tools capturing meeting attendance and decisions
- Vendor management systems for third-party AI governance data
- Environmental monitoring tools for carbon and resource consumption

Data Quality Standards

- Accuracy: Data must accurately reflect actual governance activities and outcomes
- Completeness: All required data points must be captured without gaps
- Timeliness: Data must be collected and reported within defined timeframes
- Consistency: Data collection methods must remain consistent over time
- Auditability: Data sources and collection processes must be documented and verifiable

Roles and Responsibilities

- KPI Owners: Senior leaders responsible for specific GOVERN functions who set targets and drive improvement
- Data Stewards: Individuals responsible for collecting, validating, and maintaining KPI data
- Governance Office: Central team responsible for aggregating, analyzing, and reporting KPIs
- Internal Audit: Independent validation of KPI data accuracy and completeness
- Executive Leadership: Reviews KPIs, approves targets, and allocates resources for improvement

T.3 GOVERN Function KPIs by Subcategory

T.3.1 GOVERN 1.1: Legal/Regulatory Compliance KPIs

These KPIs measure legal/regulatory compliance kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
Compliance Register Completeness	Leading	Percentage of applicable regulations documented	$(\text{Documented regulations} / \text{Total applicable}) \times 100$	Compliance Register (Appendix A)	Quarterly	100%
Regulatory Assessment Currency	Leading	Regulations assessed within required timeframe	$(\text{Assessed within 30 days} / \text{Total}) \times 100$	Compliance tracking system	Monthly	$\geq 95\%$
Regulatory Violations	Lagging	Number of AI regulation violations	Count of confirmed violations	Legal incident reports	Quarterly	0
Gap Remediation Rate	Leading	Compliance gaps remediated within SLA	$(\text{Closed within deadline} / \text{Total gaps}) \times 100$	Gap tracking system	Monthly	$\geq 90\%$
Regulatory Monitoring Coverage	Leading	Jurisdictions with active monitoring	$(\text{Monitored jurisdictions} / \text{Total relevant}) \times 100$	Monitoring subscriptions	Quarterly	100%
Assessment Timeliness	Leading	Days from identification to assessment	$\text{Avg}(\text{Assessment date} - \text{ID date})$	Compliance register timestamps	Quarterly	≤ 21 days
Reporting Timeliness	Lagging	Reports submitted on time	$(\text{On-time reports} / \text{Total required}) \times 100$	Regulatory correspondence log	Quarterly	100%

T.3.2 GOVERN 1.2: Roles and Responsibilities KPIs

These KPIs measure roles and responsibilities kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
Role Definition Completeness	Leading	Governance roles with documented responsibilities	$(\text{Documented roles} / \text{Required roles}) \times 100$	Role charters (Appendix B)	Quarterly	100%
Role Vacancy Rate	Lagging	Critical governance roles unfilled	$(\text{Vacant roles} / \text{Total critical}) \times 100$	HR systems	Monthly	≤5%
Decision Authority Currency	Leading	Decision types with assigned authority	$(\text{Assigned} / \text{Total decision types}) \times 100$	Decision Matrix (Appendix C)	Quarterly	100%
Escalation Process Usage	Leading	Complex decisions properly escalated	$(\text{Escalated} / \text{Complex decisions}) \times 100$	Escalation logs	Quarterly	≥80%
Accountability Clarity Score	Lagging	Staff rating of governance accountability clarity	Mean survey score (1-5 scale)	Annual governance survey	Annual	≥4.0/5.0

T.3.3 GOVERN 1.3: DEIA KPIs

These KPIs measure deia kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
AI Team Diversity Index	Lagging	Demographic diversity across protected characteristics	Weighted avg representation	HR diversity dashboard	Quarterly	≥0.70
DEIA Training Completion	Leading	AI practitioners completing DEIA training	(Completed / Total AI staff) × 100	Learning management system	Quarterly	≥95%
Fairness Assessment Coverage	Leading	High-risk systems with fairness assessments	(Assessed / High-risk systems) × 100	Fairness testing database	Monthly	100%
Detected Fairness Issues	Lagging	Systems with fairness concerns	Count failing fairness thresholds	Fairness testing reports	Quarterly	Trending down
Accessibility Compliance	Leading	User-facing systems meeting WCAG 2.1 AA	(Accessible systems / Total user-facing) × 100	Accessibility testing results	Quarterly	100%
Stakeholder Diversity	Leading	Diverse groups in governance processes	Count of distinct demographic groups	Engagement logs (GOVERN 6.1.4)	Quarterly	≥8 groups
DEIA Impact Assessments	Leading	New systems with DEIA assessments	(Systems with assessments / New deployments) × 100	DEIA assessment repository	Monthly	100%
Remediation Success Rate	Lagging	Fairness issues successfully resolved	(Resolved / Total identified) × 100	Issue tracking system	Quarterly	≥85%

T.3.4 GOVERN 1.4: Risk Culture KPIs

These KPIs measure risk culture kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
Risk Culture Survey Score	Lagging	Employee AI risk awareness rating	Mean annual survey score (1-5)	Risk culture survey	Annual	≥4.0/5.0
Risk Reporting Rate	Leading	AI risks reported through channels	Count of risk reports submitted	Risk reporting system	Quarterly	Trending up then stable
Awareness Campaign Reach	Leading	Employees reached by governance communications	$(\text{Reached} / \text{Total employees}) \times 100$	Communication platform analytics	Quarterly	≥90%
Risk Champion Network	Leading	Designated risk champions across units	Count of active trained champions	Risk champion registry	Quarterly	≥1 per 50 practitioners
Leadership Communication Frequency	Leading	Executive messages on AI governance	Count of leadership communications	Communications archive	Quarterly	≥4 per quarter
Speak-Up Culture Index	Lagging	Confidence in raising concerns without retaliation	% rating speak-up as good/excellent	Annual survey	Annual	≥80%

T.3.5 GOVERN 1.5: Agentic AI Governance KPIs

These KPIs measure agentic ai governance kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
Agentic System Inventory Coverage	Leading	Agentic AI systems documented	$(\text{In inventory} / \text{Total agentic}) \times 100$	Agentic inventory (Appendix F)	Monthly	100%
Autonomy Classification Accuracy	Leading	Systems with validated classifications	$(\text{Validated} / \text{Total agentic}) \times 100$	Classification records (GOVERN 1.5.2)	Quarterly	≥95%
Agent Identity Compliance	Leading	Systems with agent identity controls	$(\text{With controls} / \text{Requiring controls}) \times 100$	Agent identity registry (Appendix H)	Quarterly	100%
Oversight Breach Rate	Lagging	Incidents exceeding oversight boundaries	Count of boundary violations	Incident system, monitoring logs	Monthly	0
Action Monitoring Coverage	Leading	Agent actions logged and monitored	$(\text{Monitored} / \text{Total actions}) \times 100$	Agent monitoring platform	Monthly	≥99%
Risk Assessment Currency	Leading	Agentic systems with current assessments	$(\text{Current assessments} / \text{Total agentic}) \times 100$	Risk repository (Appendix I)	Monthly	≥90%
Testing Coverage	Leading	Systems with boundary/failure testing	$(\text{Tested} / \text{Production agentic}) \times 100$	Testing records (GOVERN 1.5.4)	Quarterly	100%
Agent Failure Rate	Lagging	Agent operations resulting in failures	$(\text{Failed actions} / \text{Total actions}) \times 100$	Monitoring platform, error logs	Monthly	≤0.1%

T.3.6 GOVERN 1.6: Environmental Sustainability KPIs

These KPIs measure environmental sustainability kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
AI Carbon Footprint	Lagging	Total emissions from AI operations (training + inference)	Sum(Compute hours × Carbon intensity) in metric tons CO2e	Carbon tracking (Appendix J)	Monthly	Trending down
Carbon Assessment Coverage	Leading	Systems with carbon footprint assessments	(With assessments / Total systems) × 100	Carbon assessment repository	Quarterly	≥95%
Renewable Energy Usage	Lagging	AI compute powered by renewable energy	(Renewable hours / Total hours) × 100	Cloud provider energy reports	Monthly	≥80%
Energy Efficiency	Lagging	Energy per AI inference operation	Total energy / Total inferences (kWh per 1000)	System monitoring, energy metering	Monthly	Trending down
Green Optimization Rate	Leading	Systems with energy optimization	(Optimized / Total systems) × 100	Optimization project records	Quarterly	≥70%
Environmental Reporting Compliance	Leading	Required disclosures completed on time	(Completed reports / Required) × 100	Sustainability reporting system	Quarterly	100%
Training Carbon Efficiency	Lagging	Emissions per model trained	Sum training emissions / Models trained (kg CO2e)	Training logs (Appendix K)	Monthly	≥5% YoY reduction
Water Usage Efficiency	Lagging	Water for infrastructure cooling per compute	Water consumed / Compute hours (liters per hour)	Data center metering	Monthly	≤Target

T.3.7 GOVERN 2.1: Accountability and Transparency KPIs

These KPIs measure accountability and transparency kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
System Registration Rate	Leading	Deployed AI systems in official inventory	$(\text{Registered} / \text{Total deployed}) \times 100$	AI inventory (GOVERN 2.1.1)	Monthly	100%
Approval Process Adherence	Leading	Systems following required approvals	$(\text{With approval records} / \text{Requiring approval}) \times 100$	Approval workflow (GOVERN 2.1.2)	Quarterly	100%
Unauthorized System Incidents	Lagging	Systems deployed without authorization	Count discovered operating without approval	Internal audit reports	Quarterly	0
Escalation Usage Rate	Leading	Complex decisions properly escalated	$(\text{Using escalation} / \text{Complex decisions}) \times 100$	Escalation logs (GOVERN 2.1.3)	Quarterly	≥90%
Transparency Documentation	Leading	Public-facing systems with documentation	$(\text{With public docs} / \text{Public-facing systems}) \times 100$	Transparency register (GOVERN 2.1.4)	Quarterly	100%
Avg Approval Cycle Time	Lagging	Days from submission to approval	Avg(Approval date - Submission date)	Approval workflow timestamps	Monthly	≤30 days

T.3.8 GOVERN 3.1: Diverse and Skilled Teams KPIs

These KPIs measure diverse and skilled teams kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
Team Demographic Diversity	Lagging	Underrepresented groups in AI teams	% from underrepresented demographics	HR diversity analytics	Quarterly	≥40%
Cognitive Diversity Score	Lagging	Diversity of educational backgrounds	Distinct disciplines / Team size	Employee education records	Quarterly	≥0.30
Inclusive Hiring Adherence	Leading	AI hires using inclusive practices	(Inclusive hires / Total AI hires) × 100	Recruitment system (GOVERN 3.1.2)	Quarterly	100%
Diverse Hiring Panel Usage	Leading	Decisions made by diverse panels	(Diverse panels / Total decisions) × 100	Interview panel records	Quarterly	≥95%
Skills Assessment Coverage	Leading	Team members with skills assessments	(With assessments / Total members) × 100	Skills assessment system	Annual	100%
Critical Skill Coverage	Lagging	Critical AI skills adequately represented	(Covered skills / Total critical) × 100	Skills inventory	Quarterly	≥85%
Team Retention Rate	Lagging	Diverse AI talent retained over 12 months	(Retained / At start of period) × 100	HR attrition tracking	Annual	≥90%
Inclusive Culture Score	Lagging	Team member rating of inclusiveness	Mean inclusion survey score (1-5)	Engagement survey	Annual	≥4.2/5.0

T.3.9 GOVERN 4.1: Awareness and Training KPIs

These KPIs measure awareness and training kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
Foundational Training Completion	Leading	Personnel completing foundational training	$(\text{Completed} / \text{Requiring training}) \times 100$	LMS (GOVERN 4.1.1)	Monthly	≥95%
Role-Specific Training Coverage	Leading	Practitioners with role-specific training	$(\text{With training} / \text{Total practitioners}) \times 100$	LMS role-based curriculum	Quarterly	≥90%
Training Effectiveness Score	Lagging	Assessment scores showing knowledge retention	Mean post-training assessment score	LMS assessment results	Quarterly	≥85%
Training Currency Rate	Leading	Personnel with current (not expired) certifications	$(\text{Current} / \text{Total trained}) \times 100$	LMS expiration tracking	Monthly	≥98%
Executive Training Participation	Leading	Executives completing governance training	$(\text{Completed} / \text{Total executives}) \times 100$	LMS executive records	Annual	100%
Training Update Frequency	Leading	Months since materials last updated	Current date - Last update for each course	Content management system	Quarterly	≤6 months
Behavioral Application Score	Lagging	Manager rating of training application	Mean manager rating (1-5)	Manager evaluations	Annual	≥4.0/5.0

T.3.10 GOVERN 5.1: Oversight Functions KPIs

These KPIs measure oversight functions kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
Committee Meeting Frequency	Leading	Governance committee meetings held vs planned	Actual meetings / Planned meetings	Committee calendar	Quarterly	100%
Member Attendance Rate	Leading	Average attendance across meetings	(Members attending / Invited) × 100	Meeting attendance records	Quarterly	≥85%
Ethics Review Completion	Leading	High-risk systems receiving ethics review	(Reviewed / Requiring review) × 100	Ethics review logs (GOVERN 5.1.2)	Quarterly	100%
Internal Audit Coverage	Leading	GOVERN procedures audited annually	(Audited / Total procedures) × 100	Audit plan (GOVERN 5.1.3)	Annual	≥80%
Audit Remediation Rate	Lagging	Findings closed within timeframe	(Remediated within SLA / Total findings) × 100	Audit management system	Quarterly	≥90%
High-Risk Finding Rate	Lagging	High-severity audit findings	Count of high-severity findings	Audit reports	Quarterly	≤2 per quarter
External Audit Readiness	Lagging	Organizational readiness for external audits	% score from readiness assessment	External audit prep review	Annual	≥90%

T.3.11 GOVERN 6.1: Trustworthy AI Characteristics KPIs

These KPIs measure trustworthy ai characteristics kpis effectiveness and compliance.

KPI Name	Type	Description	Calculation Method	Data Source	Frequency	Target
AI Incident Rate	Lagging	Incidents causing harm or violations	Count meeting severity threshold	Incident management (GOVERN 2.1.3)	Monthly	≤5 per month
Critical Response Time	Lagging	Hours from detection to response	Avg(Response time - Detection time)	Incident response logs	Monthly	≤4 hours
Performance Monitoring Coverage	Leading	Production models with active monitoring	(Monitored / Total production models) × 100	Model monitoring platform	Monthly	100%
Model Drift Detection	Lagging	Models exhibiting performance drift	Count flagged exceeding thresholds	Drift analysis reports	Monthly	≤10% of models
Security Testing Coverage	Leading	Systems completing security testing	(Tested / Total systems) × 100	Security testing records	Annual	100%
Privacy Impact Assessments	Leading	Systems processing personal data with PIAs	(With PIAs / Processing personal data) × 100	Privacy assessment repository	Quarterly	100%
Fairness Testing Pass Rate	Lagging	Systems passing fairness evaluation	(Passing / Total tested) × 100	Fairness testing platform	Quarterly	≥95%
Explainability Documentation	Leading	High-impact systems with explainability docs	(With docs / High-impact systems) × 100	Explainability repository	Quarterly	100%

T.4 Executive Dashboard Template

The Executive Dashboard provides a high-level visual summary of AI governance health for senior leadership, consolidating critical KPIs into an at-a-glance format for rapid assessment and decision-making.

T.4.1 Dashboard Components

1. Governance Health Scorecard - Color-coded status across all 11 GOVERN subcategories
2. Critical Risk Indicators - Highlight lagging indicators signaling potential failures
3. Monthly Performance Trends - Six-month trend visualization for key leading indicators
4. Priority Actions - Items requiring executive attention

T.4.2 Status Indicators

- **Green (On Track):** All critical KPIs meeting or exceeding targets
- **Yellow (Monitor):** One or more KPIs within 10% of target threshold
- **Red (Action Required):** One or more KPIs significantly below target or critical issues identified

T.4.3 Sample Dashboard Metrics

- Example Executive Dashboard showing current status across all GOVERN functions:
 - 1.1 Legal/Regulatory: GREEN - 100% compliance register completeness, 0 violations
 - 1.2 Roles/Responsibilities: GREEN - 100% role definitions, 2% vacancy rate
 - 1.3 DEIA: YELLOW - 0.68 diversity index (target 0.70), improving trend
 - 1.4 Risk Culture: GREEN - 4.2/5.0 culture survey score
 - 1.5 Agentic AI: GREEN - 100% inventory coverage, 0 oversight breaches
 - 1.6 Environmental: YELLOW - 78% renewable energy (target 80%)
 - 2.1 Accountability: GREEN - 100% system registration, 0 unauthorized systems
 - 3.1 Diverse Teams: GREEN - 42% underrepresented groups representation
 - 4.1 Training: GREEN - 96% training completion rate
 - 5.1 Oversight: GREEN - 100% meeting frequency, 92% audit remediation
 - 6.1 Trustworthy AI: YELLOW - 6 incidents this month (target ≤5)

T.4.4 Dashboard Refresh and Distribution

- Updated monthly with the latest KPI data from all source systems
- Reviewed in monthly executive leadership meetings with action items
- Distributed to C-suite and AI Governance Committee members
- Available as live digital dashboard for real-time access
- Accompanied by brief narrative summary prepared by AI Governance Office

T.5 Board Reporting Template

The Board Reporting Template provides quarterly governance updates to the Board of Directors, offering strategic oversight of AI risk management and governance effectiveness.

T.5.1 Quarterly Report Structure

- Section 1: Executive Summary - Overall health, achievements, critical risks, recommendations
- Section 2: Governance Maturity Progress - Scoring trends across GOVERN categories
- Section 3: AI Portfolio Overview - System counts, high-risk systems, new deployments
- Section 4: Risk and Incident Summary - Quarterly statistics and trends
- Section 5: Regulatory and Compliance Updates - Key regulatory developments and status
- Section 6: Strategic Initiatives - Major programs and investments
- Section 7: Board Decisions Required - Approvals, investments, policy updates

T.5.2 Sample Board Report Excerpts

Executive Summary Example

Q1 2026 AI Governance Summary: Overall governance health remains strong with 9 of 11 GOVERN subcategories rated Green. Two areas requiring attention: DEIA (diversity index at 0.68, target 0.70) and Environmental (renewable energy at 78%, target 80%). Both show improving trends with remediation plans on track for Q2 completion. Key achievements this quarter: successful deployment of agentic AI monitoring platform covering all 34 autonomous systems, completion of ISO 42001 gap analysis showing 92% readiness, launch of enhanced DEIA recruitment partnerships with 3 universities.

Risk and Incident Summary Example

Total incidents decreased 18% from Q4 2025 (17 incidents) to Q1 2026 (14 incidents). Performance/accuracy issues declined 25%, fairness/bias concerns declined 33%, privacy incidents declined 50%. However, security vulnerabilities increased 25% (from 4 to 5), requiring enhanced pre-deployment security testing. No compliance violations occurred. All incidents were contained within defined SLAs with average response time of 3.2 hours (target ≤4 hours). Root cause analysis identified need for enhanced testing protocols for generative AI applications.

T.6 Governance Maturity Scoring

The Governance Maturity Scoring methodology provides a structured approach to assessing and tracking AI governance capabilities across all GOVERN subcategories.

T.6.1 Maturity Levels

- **Level 1 - Initial (0.0-1.0):** Ad hoc, reactive, limited documentation
- **Level 2 - Developing (1.1-2.0):** Some processes, basic structures, reactive approach
- **Level 3 - Defined (2.1-3.0):** Comprehensive processes, governance established, mix of reactive/proactive
- **Level 4 - Managed (3.1-4.0):** Measured and controlled, integrated, proactive, consistent
- **Level 5 - Optimized (4.1-5.0):** Continuous improvement, business enabler, adaptive, industry-leading

T.6.2 Scoring Methodology

Composite Score = (KPI Performance × 0.60) + (Process Implementation × 0.30) + (Qualitative Assessment × 0.10)

- **KPI Performance (60%):** Percentage of KPIs meeting or exceeding targets
- **Process Implementation (30%):** Completeness and quality of governance processes
- **Qualitative Assessment (10%):** Expert evaluation of culture, innovation, improvement

T.7 Benchmark Ranges and Industry Standards

This section provides benchmark ranges for key governance KPIs based on industry research, regulatory guidance, and leading practices. Organizations can use these to contextualize performance and identify improvements.

T.7.1 Maturity Benchmarks by Organization Type

Organization Type	Early Stage	Developing	Mature	Leading
Large Enterprise (>5000)	2.0-2.8	2.9-3.5	3.6-4.2	4.3-5.0
Mid-Size (500-5000)	1.5-2.3	2.4-3.2	3.3-4.0	4.1-5.0
Small (<500)	1.0-2.0	2.1-2.9	3.0-3.8	3.9-5.0
High-Risk AI-Intensive	2.5-3.2	3.3-3.9	4.0-4.5	4.6-5.0

T.7.2 Functional Area Benchmarks

- **Compliance Register Completeness:** Min 90%, Average 95%, Best Practice 100%
- **Training Completion Rate:** Min 85%, Average 92%, Best Practice ≥95%
- **AI System Registration:** Min 95%, Average 98%, Best Practice 100%
- **Fairness Assessment Coverage:** Min 90%, Average 96%, Best Practice 100%
- **Renewable Energy Usage:** Min 50%, Average 72%, Best Practice ≥80%

T.7.3 Industry-Specific Adjustments

- **Healthcare:** Minimum maturity 3.5/5.0 due to patient safety; 100% fairness coverage for clinical AI
- **Financial Services:** Minimum maturity 3.0/5.0; 100% MRM compliance for lending decisions
- **Government:** Minimum maturity 3.0/5.0; 100% transparency for public-facing systems

Conclusion and Implementation Guidance

The KPI Dashboard provides organizations with a comprehensive framework for measuring and improving AI governance effectiveness. Successful implementation requires executive commitment, phased rollout, technology enablement, continuous improvement, and integration with governance processes.

Phased Implementation Approach

- Phase 1 (Months 1-3): Critical lagging indicators and compliance KPIs
- Phase 2 (Months 4-6): Leading indicators for proactive risk management
- Phase 3 (Months 7-9): Complete framework across all GOVERN subcategories
- Phase 4 (Month 10+): Optimize data collection, automate reporting, refine targets

Effective AI governance is fundamentally about making good decisions informed by accurate data. The KPI Dashboard transforms governance from compliance theater into genuine risk management and business enablement. Organizations that embrace data-driven governance will not only reduce AI-related risks but also unlock the full potential of AI to create value while maintaining stakeholder trust.

The journey toward governance excellence begins with measurement.

Appendix U: AI Governance Case Studies

U.1 Purpose and Structure

This appendix provides real-world case studies demonstrating the application of the AI RMF 2026 GOVERN framework in practice. Each case study illustrates how governance procedures, templates, and decision-making processes work together to address specific AI governance challenges. These examples are based on common scenarios organizations face when implementing comprehensive AI governance, with details adapted to protect confidentiality while maintaining instructive value.

The case studies cover diverse governance scenarios including high-risk system approvals, agentic AI deployment, environmental impact assessments, ethics reviews, supplier assessments, incident response, bias mitigation, and certification journeys. Each follows a standardized structure to facilitate learning and comparison across different governance situations.

U.2 Case Study 1: High-Risk AI System Approval Process

Section 1: Case Overview

Title: Clinical Decision Support System for Oncology Treatment Recommendations

Organization: MedHealth Regional Cancer Center, a mid-sized healthcare provider with 450 beds and specialized oncology services

Timeline: 8 weeks from initial proposal to conditional approval (March 1 - April 26, 2025)

Key GOVERN Procedures Involved:

- GOVERN 2.1.2: AI System Approval Requirements and Decision Authority
- GOVERN 6.1: Trustworthy AI Policy Framework
- GOVERN 5.1.2: Ethics Review Board Procedures
- Appendix O: AI Impact Assessment (AIIA) Template
- Appendix A: Decision Authority Matrix

Section 2: Background and Context

MedHealth's oncology department proposed deploying an AI-powered clinical decision support system (CDSS) to assist oncologists in developing personalized cancer treatment plans. The system, developed by a reputable third-party vendor, uses machine learning to analyze patient medical histories, genomic data, clinical trial outcomes, and treatment efficacy patterns to recommend evidence-based treatment protocols for breast, lung, and colorectal cancers.

The proposal emerged from the oncology department's desire to leverage the latest evidence-based medicine more comprehensively than individual physicians could maintain through manual literature review. Initial discussions with the vendor indicated the system achieved 92% concordance with expert tumor board recommendations in validation studies, potentially improving treatment consistency and patient outcomes while reducing physician cognitive load.

However, the AI Governance Committee recognized this as a high-risk AI system requiring comprehensive review. The system would directly influence life-or-death treatment decisions, process sensitive health information including genetic data, and required careful consideration of FDA regulatory requirements, HIPAA compliance, medical liability implications, and potential algorithmic bias in treatment recommendations.

Section 3: Governance Process Applied

- **Week 1-2: Initial Assessment and AIIA Completion.** The oncology department lead worked with the AI Governance Office to complete a comprehensive AI Impact Assessment using Appendix O. This 28-page document analyzed the system's purpose, technical architecture, data requirements, patient impact, regulatory obligations, and risk profile. The AIIA identified the system as Category 3 (High Risk) under GOVERN 2.1.2.
- **Week 3-4: Detailed Risk Analysis and Control Design.** The cross-functional AI Safety Working Group conducted deep-dive sessions examining potential failure modes. The team designed a comprehensive control framework including mandatory physician oversight, clinical validation with tumor boards, bias testing across demographic groups, and integration with existing clinical workflows.
- **Week 5: Ethics Review Board Evaluation.** Following GOVERN 5.1.2 procedures, the AI Ethics Review Board conducted a structured review focusing on patient autonomy, informed consent, fairness, and physician judgment preservation. The Board required enhanced patient consent processes, quarterly bias audits, and physician education emphasizing critical evaluation of AI recommendations.
- **Week 6-7: Governance Committee Review.** The AI Governance Committee received the complete approval package. During a 3-hour meeting, committee members scrutinized vendor transparency, post-market surveillance plans, incident response protocols, and physician training requirements.
- **Week 8: Conditional Approval Decision.** The Committee granted conditional approval with 12 specific requirements including pilot deployment limited to breast cancer cases only, mandatory physician override documentation, monthly tumor board concordance reviews, quarterly demographic bias analyses, enhanced patient consent process, and 6-month comprehensive evaluation before expansion.

Section 4: Key Documents and Artifacts

Document 1: Excerpt from Completed AI Impact Assessment (Appendix O)

AIIA Section	Content Summary
System Classification	Category 3 - High Risk Clinical Decision Support. Direct impact on patient treatment decisions involving life-critical care (oncology). Processes sensitive PHI including genetic information.
Primary Risk Factors	Patient Safety: Incorrect recommendations could lead to inappropriate treatment. Privacy: Genetic and clinical data processing. Bias: Underrepresentation of minority populations in training data.
Regulatory Requirements	FDA: Software as Medical Device considerations. HIPAA: Business Associate Agreement required, encryption, audit logging. State: Genetic privacy law compliance.
Mitigation Controls	Clinical validation against tumor board gold standard. Physician override required. Bias testing across demographics. Patient consent process. Quarterly performance monitoring.

Document 2: Approval Conditions Summary

#	Approval Condition	Due Date/Frequency
1	Limit initial deployment to breast cancer treatment recommendations only	Ongoing
2	Mandatory physician documentation of override decisions and clinical rationale	Every case
3	Monthly tumor board concordance review (AI vs expert consensus)	Monthly
4	Quarterly demographic bias analysis across race, ethnicity, age, gender	Quarterly
5	Enhanced patient consent process approved by Legal and Privacy	Before May 20
6	Six-month comprehensive evaluation before expanding to other cancer types	October 30

Section 5: Outcome and Results

The AI Governance Committee granted conditional approval for pilot deployment limited to breast cancer treatment recommendations, with 12 specific conditions addressing patient safety, bias monitoring, physician oversight, and patient consent. The phased approach balanced clinical innovation with risk management.

The approval authorized deployment to begin May 20, 2025. The system would initially support approximately 15-20 new breast cancer patients per month, with all recommendations requiring physician review and explicit acceptance or override.

Quantitative outcomes: 8-week approval timeline, engagement of 15 stakeholders across 6 departments, 28-page AIIA, 12 approval conditions, quarterly monitoring requirements, and 6-month mandatory evaluation before scope expansion. The rigorous process built organizational confidence that patient safety and regulatory compliance were thoroughly considered.

Section 6: Lessons Learned

What Worked Well	What Could Be Improved	Recommendations for Others
Structured AIIA template ensured comprehensive risk assessment covering clinical, regulatory, privacy, and ethical dimensions systematically.	Initial timeline estimate significantly underestimated complexity, causing stakeholder frustration. Should have set realistic 8-10 week expectation upfront.	Use risk-based timeline expectations: Low risk (2-3 weeks), Medium (4-6 weeks), High (8-12 weeks). Communicate realistic timeframes early.
Cross-functional stakeholder engagement surfaced diverse perspectives, preventing blind spots and building organizational buy-in.	Vendor transparency negotiations were contentious. Earlier clarity on minimum transparency requirements would have streamlined discussions.	Develop standard vendor transparency requirements before procurement. Include in RFP requirements rather than negotiating post-selection.
Phased approval approach balanced innovation with risk management effectively, providing learning opportunity before broader deployment.	Patient advocacy perspective came late in process. Earlier involvement would have identified consent concerns sooner.	Include patient advocacy representatives from initial AIIA stage. Patient perspective often differs from clinical team assumptions.

Section 7: Takeaway for Practitioners

High-risk AI system approvals require structured, comprehensive review processes that balance innovation with appropriate risk management. The approval timeline should reflect risk level—organizations should resist pressure to accelerate beyond what thorough due diligence requires. Cross-functional stakeholder engagement is essential, but patient/end-user perspectives must be included early. Conditional approvals with phased deployment provide an effective middle ground between approval and rejection. Finally, pre-established organizational standards streamline individual system evaluations and ensure consistency across decisions.

U.3 Case Study 2: Agentic AI Deployment and Oversight

Section 1: Case Overview

Title: Autonomous Fraud Detection Agents with Account Freeze Authority

Organization: SecureBank Financial Services, a regional bank with 2.3 million retail customers and \$45B in assets

Timeline: 12 weeks from concept to initial deployment (January 15 - April 10, 2025)

Key GOVERN Procedures Involved:

- GOVERN 1.5.1: Agentic AI Governance Committee Charter
- GOVERN 1.5.2: Agent Identity and Taxonomy Framework
- GOVERN 1.5.3: Agentic AI Decision Authority Matrix
- GOVERN 1.5.4: Human-Agent Collaboration Protocols
- Appendix F, H, I: Agentic AI Templates

Section 2: Background and Context

SecureBank's fraud prevention team proposed deploying autonomous AI agents capable of analyzing transaction patterns in real-time and independently freezing accounts exhibiting high-probability fraud indicators without human review. The proposal stemmed from increasing fraud sophistication, with losses growing 23% year-over-year.

Traditional fraud detection required human analyst review before account freezes, creating a 15-45 minute window where fraudsters could complete unauthorized transactions. The fraud team estimated autonomous agents acting within seconds could prevent \$12-18M in annual losses. However, concerns arose: What if agents incorrectly freeze legitimate customer accounts? How would customers appeal automated decisions?

This was SecureBank's first agentic AI deployment—previous AI systems were purely recommendation-based. The organization lacked governance frameworks for AI agents with independent decision authority.

Section 3: Governance Process Applied

- Week 1-3: Agentic AI Governance Foundation.** SecureBank established an Agentic AI Governance Committee using Appendix F template. The Committee developed the bank's agent taxonomy and decision authority matrix using Appendix H, defining what decisions different agent types could make independently versus requiring human approval.
- Week 4-5: Agent Risk Assessment.** The fraud team completed a comprehensive Agent Risk Assessment (Appendix I) classifying the agents as 'Level 3 - High Autonomy, High Risk' based on independent decision authority over customer account access.
- Week 6-7: Autonomy Level Determination.** The Committee designed a graduated autonomy approach: Phase 1 (Months 1-3) - Read-only mode; Phase 2 (Months 4-6) - Limited autonomy, 1-hour freeze maximum; Phase 3 (Months 7+) - Full autonomy for 24-hour freezes.
- Week 8-9: Agent Identity Framework.** SecureBank implemented a formal agent identity framework. Each fraud detection agent received a unique agent ID, registered in the enterprise agent registry with complete metadata.
- Week 10-12: Oversight Dashboard and Deployment.** The Committee required a real-time agent oversight dashboard before authorizing deployment. The Committee approved Phase 1 deployment (read-only mode) beginning April 10.

Section 4: Key Documents and Artifacts

Document 1: Agent Risk Assessment Summary (Appendix I - Excerpt)

Assessment Element	Details
Agent Classification	Level 3 - High Autonomy, High Risk. Independent decision authority affecting customer account access.
Authorized Actions - Phase 1	Analyze transaction patterns, calculate fraud probability scores, flag suspicious transactions for human review
Primary Risks	False positives impacting legitimate customers. Discriminatory patterns in freeze decisions. Cascading failures if agents operate incorrectly.
Human Oversight Controls	Phase 1: 100% human review. Real-time oversight dashboard. Weekly Committee performance review.
Success Criteria	Phase 1→2: <5% false positive rate over 90 days, no demographic bias detected.

Document 2: Agent Identity Registry Entry

Field	Value
Agent ID	AGENT-FRAUD-001
Agent Name	Real-Time Transaction Fraud Detector
Risk Classification	Level 3 - High Autonomy, High Risk
Current Autonomy Level	Phase 1: Read-Only (Flag & Recommend)
Responsible Agent Owner	Marcus Thompson, VP Fraud Prevention
Deployment Status	Active - Phase 1 (since April 10, 2025)
Next Review Date	July 10, 2025 (Phase 2 readiness evaluation)

Section 5: Outcome and Results

The Agentic AI Governance Committee approved phased deployment of fraud detection agents beginning April 10, 2025, starting with read-only mode for initial 90 days. This graduated autonomy approach allowed SecureBank to validate agent performance and build organizational confidence before granting independent decision authority.

Phase 1 deployment analyzed approximately 85,000 transactions daily, flagging 1,200-1,500 per day for human review (1.4-1.8% flagging rate). Human analysts confirmed approximately 73% as legitimate (false positives) and 27% as actual fraud attempts. The deployment established critical governance foundations: formal Agentic AI Governance Committee, agent identity registry, standardized risk assessment methodology, oversight dashboard providing real-time visibility, and organizational understanding of graduated autonomy concepts.

Section 6: Lessons Learned

What Worked Well	What Could Be Improved	Recommendations for Others
Graduated autonomy phases balanced innovation with risk management effectively. Read-only phase built confidence before granting decision authority.	Initial timeline underestimated governance foundation work needed for first agentic AI deployment.	For first agentic deployment, allocate time to build governance foundations before system-specific work.
Agent identity framework with unique IDs and formal registry created clear accountability and enabled systematic oversight.	Success criteria for phase progression proved unrealistic based on Phase 1 actual performance.	Set conservative initial performance expectations. Use Phase 1 actual performance to calibrate realistic targets.
Real-time oversight dashboard provided transparency into agent decisions and enabled early issue detection.	Customer appeal/dispute process was afterthought, developed during Phase 1 rather than before deployment.	Design customer-facing processes before deployment, especially for customer-impacting autonomous decisions.

Section 7: Takeaway for Practitioners

Agentic AI deployments require fundamentally different governance approaches than traditional AI systems. Organizations cannot simply extend existing AI governance to autonomous agents—specific frameworks for agent identity, autonomy levels, human oversight, and graduated deployment are essential. The graduated autonomy approach (read-only → limited → full) is particularly valuable for managing uncertainty. For first agentic deployments, invest time establishing governance foundations. Real-time oversight visibility is critical. Finally, customer-facing processes must be designed before deployment, not retrofitted after complaints emerge.

U.4 Case Study 3: Environmental Impact Assessment and Optimization

Section 1: Case Overview

Title: Large Language Model Training Carbon Footprint Assessment and Reduction

Organization: TechVentures AI, a mid-sized technology company developing enterprise AI solutions with 800 employees

Timeline: 10 weeks from assessment initiation to optimization implementation (February 1 - April 12, 2025)

Key GOVERN Procedures Involved:

- GOVERN 1.6.1: Environmental Impact Assessment Requirements
- GOVERN 1.6.2: AI Sustainability Metrics and Monitoring
- GOVERN 1.6.3: Carbon Footprint Reduction Standards
- Appendix G, J, K: Environmental Sustainability Templates

Section 2: Background and Context

TechVentures AI planned to train a large language model with 30 billion parameters to power their enterprise document intelligence platform. Initial planning focused on model performance and training timeline, but the newly-established AI Environmental Sustainability Committee flagged this for mandatory carbon footprint assessment under GOVERN 1.6.1 procedures.

The assessment requirement emerged from TechVentures' corporate sustainability commitments (carbon neutral by 2030) and board-level ESG focus. However, the AI team had limited environmental expertise and initially viewed the assessment as bureaucratic overhead delaying critical product development.

Preliminary calculations suggested the planned training approach would generate approximately 85 metric tons CO₂e—equivalent to 9 gasoline-powered cars driven for one year. This conflicted with the company's sustainability messaging and raised concerns about whether the environmental cost justified the business value.

Section 3: Governance Process Applied

- **Week 1-2: Baseline Carbon Footprint Assessment.** The AI team completed a comprehensive carbon footprint assessment using Appendix J template. The assessment calculated total emissions: 85.3 metric tons CO₂e for training plus ongoing inference emissions estimated at 12 metric tons CO₂e annually.
- **Week 3-4: Impact Analysis and Benchmark Comparison.** The sustainability committee contextualized these numbers: 85 metric tons represented 11% of TechVentures' total annual corporate emissions and would consume 23% of their remaining 2025 carbon budget. The committee compared this to industry benchmarks—similar-scale models ranged from 45-120 metric tons CO₂e.
- **Week 5-6: Optimization Options Analysis.** The sustainability committee partnered with the AI team to identify optimization approaches using Appendix K template. Five major optimization categories emerged: data center selection, computational efficiency, model architecture, training schedule, and carbon offsets.
- **Week 7-8: Decision and Implementation Planning.** After detailed analysis, the sustainability committee recommended a combination approach: relocate training to renewable-powered data center, invest 3 weeks in training optimization, and purchase certified carbon offsets. Executive leadership approved the optimization plan despite extending the project timeline by 3 weeks.
- **Week 9-10: Optimized Training Execution.** The AI team implemented the approved optimizations: migrated to renewable-powered data center, implemented mixed-precision training and gradient accumulation, optimized batch sizes. The optimized training achieved target performance using 35% fewer FLOPs, completed in 4.5 weeks instead of 6, and generated 51 metric tons CO₂e—a 40% reduction.

Section 4: Key Documents and Artifacts

Document 1: Carbon Footprint Assessment Summary

Metric	Baseline	Optimized
Metric	Baseline Estimate	Optimized Actual
Total Computation	2.5 x 10 ²³ FLOPs	1.6 x 10 ²³ FLOPs (-36%)
Training Duration	6 weeks (1,008 hours)	4.5 weeks (756 hours)
Energy Consumption	437,500 kWh	285,000 kWh (-35%)
Training Emissions	85.3 metric tons CO ₂ e	51.2 metric tons CO ₂ e (-40%)
% of Corporate Emissions	11.0% of annual	6.6% of annual

Document 2: Optimization Recommendations Summary

Optimization	Emissions Impact	Cost Impact	Timeline Impact
Migrate to renewable data center	-42% emissions via cleaner energy	+\$45K training cost, -\$50K/year inference	+1 week
Training optimization	-30% FLOPs via computational efficiency	Neutral (offset by reduced compute)	+3 weeks, -1.5 weeks actual
Carbon offset purchase	100% offset for unavoidable emissions	+\$770 for 51 tons	No impact

Section 5: Outcome and Results

TechVentures successfully completed carbon-optimized LLM training achieving 40% emissions reduction (85.3 → 51.2 metric tons CO₂e) compared to baseline plan, with remaining emissions fully offset through certified carbon credit purchases. The optimized approach delivered target model performance while reducing environmental impact to 6.6% of corporate annual emissions rather than the baseline 11%. Beyond environmental benefits, the optimization delivered significant business value.

The optimized model required 35% less computational resources for inference, reducing ongoing operational costs approximately \$200K annually—generating positive ROI within 9 months despite the upfront optimization investment.

The process established valuable precedents: formal carbon footprint assessment procedures for major AI training runs (now standard practice), cross-functional collaboration between AI and sustainability teams, environmental optimization as core engineering practice, and executive understanding of AI environmental impact business case.

Section 6: Lessons Learned

What Worked Well	What Could Be Improved	Recommendations for Others
Quantifying emissions in business terms (% of carbon budget, cost implications) made environmental impact concrete for leadership.	Environmental assessment came late (after training plan finalized). Earlier integration would have avoided re-planning work.	Integrate environmental impact assessment into initial AI project planning, not as downstream review.
Demonstrating positive ROI from optimization transformed perception from 'compliance burden' to 'business opportunity.'	Initial resistance from AI team stemmed from perception that sustainability was blocking innovation.	Frame environmental optimization as engineering excellence and cost reduction, not just compliance.
Combination of absolute reduction (40% via optimization) plus offsets for remainder provided pragmatic approach.	Carbon offset selection process was rushed. Should have researched offset quality standards earlier.	Plan carbon offset strategy early. Research reputable providers and certification standards in advance.

Section 7: Takeaway for Practitioners

Environmental impact assessment for AI systems should be integrated into initial project planning, not treated as downstream compliance review. Organizations that frame environmental optimization as engineering excellence and cost reduction achieve better engagement and outcomes. The business case for AI environmental optimization often proves stronger than anticipated: computational efficiency reduces both emissions and operating costs, sustainability credentials enhance competitive positioning, and customers increasingly prioritize environmental responsibility. A pragmatic combination of absolute emissions reduction plus certified offsets provides an effective approach. Finally, cross-functional collaboration between AI/engineering and sustainability teams is essential.

U.5 Case Study 4: AI Ethics Review Board Decision

Section 1: Case Overview

Title: Predictive Policing System Evaluation and Rejection

Organization: Metro City Police Department, municipal law enforcement agency serving 850,000 residents

Timeline: 6 weeks from proposal submission to ethics board decision (March 1 - April 12, 2025)

Key GOVERN Procedures Involved:

- GOVERN 5.1.2: Ethics Review Board Procedures
- GOVERN 1.3: Diversity, Equity, Inclusion & Accessibility in AI
- GOVERN 6.1.8: Fairness and Bias Management

Section 2: Background and Context

Metro City Police Department proposed deploying a predictive policing system using machine learning to forecast crime hotspots and optimize patrol resource allocation. The system analyzed historical crime data, demographic patterns, weather conditions, and other contextual factors to predict where crimes were most likely to occur, generating recommended patrol deployment zones.

The proposal emerged from resource constraints—the department faced 12% budget cuts while responding to increased public safety concerns. Department leadership argued predictive analytics could help 'do more with less' by optimizing limited patrol resources. Vendor demonstrations showed 82% accuracy in retrospective testing, with claimed 15-20% crime reductions in other jurisdictions.

However, the department's AI Ethics Review Board had serious concerns. Community advocates raised alarm about potential algorithmic bias—would the system perpetuate historical discriminatory policing patterns embedded in training data? Civil liberties organizations questioned due process implications. The board needed to carefully evaluate whether purported public safety benefits outweighed significant civil rights risks and ethical concerns.

Section 3: Governance Process Applied

- **Week 1-2: Initial Ethics Review and Fairness Analysis.** The Ethics Review Board conducted initial assessment examining fairness, transparency, accountability, privacy, and community impact. The board requested detailed fairness testing results, but the vendor claimed proprietary concerns and provided only aggregate accuracy metrics without demographic breakdowns—an immediate red flag.
- **Week 3: Community Stakeholder Input.** The board held public hearings to gather community perspectives, a critical component of GOVERN 1.3 DEIA principles. Community advocates testified that historical crime data reflected discriminatory policing patterns, not objective crime rates. One advocate explained: 'If police concentrated in Black neighborhoods for decades because of racial bias, the data shows high arrest rates there—but that reflects where police were, not where crime actually occurred.'
- **Week 4: Technical Deep Dive and External Expert Consultation.** The board engaged independent academic experts in algorithmic fairness to evaluate the vendor's system. The experts identified multiple concerning technical issues: training data quality problems, feature selection bias using socioeconomic and demographic features strongly correlated with race, lack of causal modeling, and no feedback loop mitigation mechanisms.
- **Week 5: Ethics Board Deliberation.** The board conducted intensive deliberations weighing purported benefits against serious risks. Board members representing different perspectives engaged in thoughtful debate: Public safety representatives argued crime reduction benefited all communities. Civil rights representatives countered that ends don't justify means—crime reduction achieved through discriminatory methods violates constitutional principles.
- **Week 6: Board Decision and Alternative Recommendations.** After careful consideration, the Ethics Review Board voted 7-2 to reject the predictive policing system deployment, finding unacceptable risks of algorithmic discrimination, insufficient transparency and fairness safeguards, and likely damage to community trust essential for effective policing. However, the board recognized the department's legitimate resource optimization needs and recommended alternative approaches.

Section 4: Key Documents and Artifacts

Ethics Review Board Decision Summary (April 12, 2025)

DECISION: REJECT deployment of proposed predictive policing system

VOTE: 7 in favor of rejection, 2 opposed, 0 abstentions

PRIMARY CONCERNS:

1. **Algorithmic Discrimination Risk:** The system is highly likely to perpetuate and amplify historical discriminatory policing patterns. Training on biased historical data, combined with feedback loop dynamics and lack of causal reasoning, creates unacceptable risk of directing disproportionate police attention to minority communities based on past discriminatory practices.
2. **Insufficient Transparency and Fairness Safeguards:** The vendor's refusal to provide demographic bias testing results, lack of explainability for zone predictions, and absence of feedback loop mitigation mechanisms fail to meet minimum fairness and transparency standards.
3. **Community Trust Implications:** Deployment would likely damage essential police-community trust in neighborhoods already bearing unfair policing burdens.
4. **Constitutional and Civil Rights Concerns:** The system poses significant risks of Fourth Amendment violations (unreasonable search/seizure) and Fourteenth Amendment violations (equal protection).

Alternative Approaches Recommended:

Alternative Approach	Description and Rationale
Community-Informed Resource Allocation	Partner with neighborhood organizations to identify public safety priorities through community input rather than algorithmic predictions. Build trust and ensure policing serves community needs.
Root Cause Prevention Focus	Invest in addressing underlying crime drivers: economic opportunity programs, mental health services, youth engagement, housing stability support, and community violence interruption programs.
Enhanced Community Policing	Strengthen community policing strategies emphasizing relationship-building, foot patrols, community partnerships, and problem-oriented policing.

Section 5: Outcome and Results

The Ethics Review Board rejected the predictive policing system deployment by 7-2 vote, finding unacceptable risks of algorithmic discrimination, insufficient fairness safeguards, and likely damage to essential police-community trust. The board's decision demonstrated that governance processes can and should reject AI systems when ethical risks outweigh purported benefits.

The department accepted the board's decision and committed to exploring the recommended alternative approaches. The police chief publicly acknowledged community concerns, stating: 'Public safety requires public trust. If a technology damages the community relationships essential for effective policing, it undermines rather than supports our mission.' The department redirected the \$850K budgeted for the predictive policing system toward community policing expansion and youth intervention programs.

The process established important precedents: ethics review boards have authority to reject AI deployments, not just suggest modifications; community stakeholder input is essential for AI systems affecting civil rights; vendor transparency is mandatory for high-stakes public sector AI; and operational efficiency cannot justify systems that perpetuate discrimination or violate civil rights principles.

Section 6: Lessons Learned

What Worked Well	What Could Be Improved	Recommendations for Others
Community stakeholder hearings provided critical perspectives that technical analysis alone would have missed, surfacing concerns about feedback loops and trust implications.	Community engagement came relatively late (Week 3). Earlier involvement would have identified concerns sooner.	Engage affected community stakeholders from the beginning for AI systems with civil rights implications.
Independent expert consultation provided credible technical evaluation of vendor claims, avoiding dependence solely on vendor-provided information.	Vendor selection occurred before ethics review. Procurement process should have included ethics screening criteria upfront.	Integrate ethics review criteria into procurement requirements. Require vendor transparency commitments as bid evaluation factors.
Providing alternative recommendations (vs pure rejection) demonstrated board's constructive role helping achieve legitimate operational goals through ethical approaches.	Initial proposal framed choice as 'deploy system vs nothing.' Earlier board involvement could have shaped alternatives before single-solution focus.	Frame ethics review as collaborative problem-solving. Engage ethics boards early to help design ethical approaches.

Section 7: Takeaway for Practitioners

Ethics review boards must have genuine authority to reject AI systems, not just suggest modifications—credible governance requires real decision power. For AI systems affecting civil rights, affected community stakeholder input is essential and must be incorporated early. Vendor transparency is mandatory for high-stakes public sector AI; proprietary concerns don't justify opacity in systems affecting fundamental rights. Ethics review should be collaborative problem-solving helping organizations achieve legitimate goals through ethical approaches. Organizations should be prepared to reject systems when ethical risks truly outweigh benefits—sometimes the right answer is 'not this system.' Finally, integrating ethics considerations into procurement processes prevents investing resources in fundamentally problematic approaches.

U.6 Case Study 5: Third-Party AI Supplier Risk Assessment

Case Summary

This case study examines RetailCo's evaluation of a third-party AI recommendation engine vendor for their e-commerce platform. The company needed to assess supplier trustworthiness, data handling practices, model transparency, and ongoing support capabilities before integrating a critical customer-facing AI system.

Key elements covered:

- Supplier due diligence using Appendix L toolkit including security assessments, data governance reviews, and technical capability evaluation
- Transparency negotiations when vendor initially refused to provide training data details and model architecture documentation
- Contractual protection development including performance guarantees, bias testing requirements, incident notification obligations, and exit clauses
- Ongoing monitoring framework with quarterly reviews, annual audits, and continuous performance tracking

Outcome: RetailCo proceeded with vendor selection after successfully negotiating additional transparency provisions and contractual protections. The supplier governance framework established through this process became the template for evaluating all future AI suppliers.

Key Takeaway: Third-party AI supplier risk assessment requires balancing vendor intellectual property concerns with organizational transparency needs. Comprehensive due diligence before contract signing is far more effective than attempting to address gaps post-implementation. Organizations should develop standardized supplier governance requirements before procurement rather than negotiating requirements supplier-by-supplier.

U.7 Case Study 6: AI Incident Investigation and Response

Case Summary

This case study details ShopNow E-commerce Platform's response to discovering their recommendation algorithm was showing systematically higher prices to customers in certain zip codes, creating potential discriminatory pricing patterns. The incident was discovered through routine fairness monitoring but required immediate investigation and remediation.

Key elements covered:

- Incident detection through automated fairness monitoring dashboard flagging geographic price disparities exceeding thresholds
- Rapid escalation to AI Governance Committee and senior leadership within 2 hours of detection
- Root cause analysis revealing unintended interaction between dynamic pricing algorithm and zip code-based shipping cost estimation
- Immediate remediation including algorithm suspension, affected customer identification and compensation (refunds plus goodwill credits), and system modification to prevent recurrence
- Transparent communication with affected customers and proactive regulatory disclosure to state consumer protection agencies

Outcome: ShopNow successfully remediated the discriminatory pricing pattern within 48 hours of detection, compensated 12,400 affected customers totaling \$487K in refunds and credits, and implemented enhanced testing protocols to prevent similar issues. The rapid, transparent response helped maintain customer trust and avoided regulatory penalties.

Key Takeaway: Effective AI incident response requires automated monitoring for early detection, clear escalation procedures, rapid investigation capabilities, and willingness to take decisive action including system suspension when necessary. Transparent communication with affected parties and proactive regulatory engagement is essential. The goal is containment, remediation, and learning—not just damage control.

U.8 Case Study 7: Bias Detection and Mitigation in Credit Scoring

Case Summary

This case study examines Community Credit Union's discovery of racial disparate impact in their AI-powered credit scoring model, the comprehensive bias mitigation effort, and regulatory reporting requirements. The institution's internal fairness testing revealed that similarly-qualified Black applicants were approved at rates 12 percentage points lower than White applicants.

Key elements covered:

- Fairness testing methodology using demographic parity, equalized odds, and disparate impact ratio metrics
- Root cause analysis identifying problematic features (neighborhood-level socioeconomic indicators serving as race proxies) and historical training data bias
- Diverse team formation including data scientists, compliance officers, community representatives, and civil rights experts to develop mitigation approaches
- Technical mitigation including model retraining with fairness constraints, feature engineering to remove proxy variables, and threshold adjustments
- Regulatory reporting to CFPB and state banking regulators with remediation plan and timeline

Outcome: Through comprehensive bias mitigation efforts, Community Credit Union reduced the approval rate disparity from 12 percentage points to 3 percentage points (60% improvement) while maintaining overall predictive accuracy. The credit union voluntarily reviewed previously denied applications from affected demographic groups and approved 340 additional applications. The transparent, proactive approach was recognized favorably by regulators.

Key Takeaway: Bias detection requires comprehensive, ongoing fairness testing across multiple metrics and demographic categories—not one-time assessments. Effective mitigation requires diverse teams bringing different perspectives and expertise.

Technical solutions alone are insufficient; meaningful bias reduction requires examining historical training data, identifying proxy variables, and potentially accepting some predictive accuracy tradeoffs to achieve fairness. Proactive regulatory engagement demonstrates good faith and often results in more favorable regulatory treatment than reactive response to complaints.

U.9 Case Study 8: ISO 42001 Certification Journey

Case Summary

This case study chronicles DataSoft Solutions' 12-month journey to achieve ISO 42001 AI Management System certification. As a mid-size AI software provider competing for enterprise contracts, DataSoft needed third-party certification to demonstrate governance maturity and differentiate from competitors.

Certification Journey Timeline:

Phase	Activities and Outcomes
Month 1-3: Gap Analysis	Engaged certification consultant to conduct comprehensive gap analysis against ISO 42001 requirements. Identified 47 gaps across documentation, processes, controls, and evidence. Prioritized gaps into critical (15), high (18), and medium (14) categories.
Month 4-9: Remediation	Systematically addressed identified gaps. Developed missing policies and procedures. Implemented technical controls. Established governance committees. Created evidence collection systems. Most challenging: establishing comprehensive AI system inventory and risk assessment processes for all deployed systems.
Month 10-11: Pre-Assessment	Conducted internal pre-assessment audit using external consultant. Identified 8 remaining gaps requiring correction. Refined documentation and evidence. Trained personnel on audit protocols and responses.
Month 12: Certification Audit	External certification body conducted 3-day on-site audit examining documentation, interviewing personnel, and reviewing evidence. Auditors issued 3 minor non-conformances requiring correction before certification grant. DataSoft addressed non-conformances within 2 weeks. Achieved ISO 42001 certification.

Key elements covered:

- Gap analysis revealing most significant deficiencies in documentation completeness, risk assessment consistency, and governance committee formalization
- Remediation challenges including resource constraints, competing priorities, and cultural change management
- Audit preparation emphasizing evidence organization, personnel training, and process consistency
- Business value realization through enhanced market positioning, customer confidence, and internal process improvements

Outcome: DataSoft achieved ISO 42001 certification in 12 months with total investment of approximately \$185K (consultant fees, personnel time, tools, audit fees). The certification became a significant competitive differentiator—DataSoft won 3 major contracts totaling \$4.2M where certification was an explicit evaluation criterion. Beyond market positioning, the certification process drove meaningful governance improvements: more consistent risk assessments, better documentation, clearer accountability, and enhanced organizational AI maturity.

Key Takeaway: ISO 42001 certification requires significant investment of time, resources, and organizational commitment—but provides substantial business value beyond the certificate itself. Gap analysis should occur early to realistically plan remediation effort and timeline. Organizations should view certification as governance improvement opportunity, not just compliance exercise. The most challenging aspect is often cultural change and consistent implementation across the organization, not technical requirements. Finally, certification provides external validation of governance maturity that resonates strongly with enterprise customers and regulators.

Appendix U Summary

This appendix has presented eight comprehensive case studies demonstrating the AI RMF 2026 GOVERN framework in action across diverse governance scenarios:

- Case Study 1: High-risk AI system approval balancing innovation with patient safety in healthcare clinical decision support
- Case Study 2: Agentic AI deployment establishing governance foundations for autonomous fraud detection agents in financial services
- Case Study 3: Environmental impact assessment and optimization achieving 40% emissions reduction in large language model training
- Case Study 4: Ethics review board rejection of predictive policing system protecting civil rights and community trust
- Case Study 5: Third-party AI supplier risk assessment ensuring vendor transparency and contractual protections
- Case Study 6: AI incident investigation and response remediating discriminatory pricing patterns rapidly and transparently
- Case Study 7: Bias detection and mitigation in credit scoring addressing racial disparate impact through comprehensive fairness testing
- Case Study 8: ISO 42001 certification journey establishing governance maturity and competitive differentiation

These case studies illustrate several common themes across effective AI governance:

- Risk-proportionate processes: High-risk systems require more rigorous review than low-risk systems. Timeline and resource allocation should match risk level.
- Cross-functional collaboration: Effective governance requires diverse perspectives from technical, legal, ethical, business, and community stakeholders.
- Phased approaches: Graduated deployment, conditional approvals, and iterative improvement allow organizations to manage uncertainty while learning.
- Transparency and accountability: Clear decision criteria, documentation, and assignment of responsibilities enable effective oversight and continuous improvement.
- Stakeholder engagement: Early involvement of affected parties (patients, customers, community members) surfaces concerns and builds trust.
- Pragmatic balance: Effective governance balances innovation enablement with appropriate risk management—not obstruction, but thoughtful facilitation.

Organizations implementing the AI RMF 2026 GOVERN framework can draw on these case studies to understand how governance procedures, templates, and decision-making processes work in practice. The case studies demonstrate that comprehensive governance is achievable, provides substantial value, and strengthens rather than hinders responsible AI deployment.

Appendix V: Agent Identity and Credential Management Protocol

Update 1.2v: AI RMF 2026 GOVERN Procedural Manual

V.1 Overview and Purpose

This appendix provides comprehensive technical protocols for implementing agent identity and credential management frameworks as required by GOVERN 1.5 (Agentic AI Governance). It extends the existing agent identity principles with operational procedures, technical specifications, and NIST IR 8596 alignment.

Alignment with Standards:

- NIST IR 8596: Cybersecurity Framework Profile for Artificial Intelligence (Draft)
- NIST AI RMF 1.0: GOVERN 1.2 (Roles and Responsibilities)
- ISO/IEC 27001:2022: Access control and identity management requirements
- Singapore Model AI Governance Framework for Agentic AI (2026 Edition)

V.2 Agent Identity Core Principles

1. Unique Identification Requirement

Every AI agent deployed in production must have a globally unique identifier that distinguishes it from:

- All human user accounts
- All other AI agents in the organization
- Service accounts and system processes
- External entities and third-party systems

Identity Format Requirements:

- Namespace: Use 'agent-' prefix to distinguish from human identities
- Format: agent-[deployment-domain]-[agent-type]-[unique-id]
- Example: agent-finance-creditdecision-7f3a92b1
- Immutability: Identity remains constant across agent lifecycle

2. Hierarchical Ownership Chain

Every agent identity must have documented ownership tracing to accountable humans:

- Primary Owner: Individual responsible for agent deployment and behavior
- Organizational Unit: Department or team with operational authority
- Executive Sponsor: Senior leader accountable for agent risk
- Multi-Agent Systems: Parent agent or orchestration owner for agent swarms

3. Capacity and Authority Recording

Agent identity registry must document:

- Independent Agent: Acts autonomously within pre-defined boundaries
- Human-Delegated Agent: Acts on behalf of specific named individual
- System Agent: Executes automated organizational processes
- Delegation Authority: Maximum decision-making scope and escalation rules

4. Permission Inheritance Model

Agents cannot possess permissions exceeding those of their authorizing human:

- Maximum Privilege Ceiling: Agent permissions \leq Owner permissions
- Least Privilege Default: Agents start with minimal necessary permissions
- Explicit Grant Requirement: All permissions must be explicitly documented
- Delegation Audit: Quarterly review of agent vs. owner permission alignment

5. Comprehensive Audit Trail

All agent identity operations must be logged and retained:

- Identity Creation: Timestamp, creator, authorization basis
- Permission Grants: Each authorization with justification and approver
- Permission Revocations: Deauthorizations with reason and date
- Identity Modifications: Changes to ownership, capacity, or scope
- Identity Deactivation: Retirement date and archival disposition
- Retention: Minimum 7 years or per regulatory requirement

V.3 Agent Identity Registry Implementation

Required Registry Fields

Field Name	Description	Data Type
agent_id	Globally unique agent identifier	String (UUID)
agent_name	Human-readable agent designation	String
agent_type	Classification: Independent, Delegated, System	Enum
primary_owner	Individual accountable for agent behavior	User_ID
organizational_unit	Department/team with operational authority	String
executive_sponsor	Senior leader accountable for risk	User_ID
deployment_domain	Business domain (finance, HR, operations, etc.)	String
risk_classification	Risk level: Low, Medium, High, Critical	Enum
creation_date	Timestamp of identity creation	DateTime
status	Active, Suspended, Deactivated	Enum
permissions	Array of granted permissions with timestamps	Array[Object]
last_activity	Timestamp of most recent agent action	DateTime

V.4 Credential Management Framework

This section defines technical requirements for agent authentication and credential lifecycle management, aligned with NIST IR 8596 guidance on credentials for autonomous cyber actors.

V.4.1 Supported Credential Types

Credential Type	Use Case	Security Requirements
API Keys	Internal system-to-system agent authentication	Minimum 256-bit entropy, encrypted at rest, rotated every 90 days
OAuth 2.0 Tokens	Agents accessing third-party APIs and cloud services	JWT format, short-lived access tokens (15 min), refresh token rotation
X.509 Certificates	High-security agents, mutual TLS authentication	PKI-issued, 2048-bit minimum, annual renewal, certificate pinning
Service Account Credentials	Cloud platform agents (AWS, Azure, GCP)	Platform-native IAM, principle of least privilege, MFA where supported

V.4.2 Credential Lifecycle Management

Phase 1: Credential Issuance

1. Authorization: Primary owner submits credential request with business justification
2. Approval: Security team reviews and approves based on risk classification
3. Generation: Credentials generated using cryptographically secure methods
4. Secure Delivery: Credentials transmitted via encrypted channel to authorized recipient only
5. Registry Update: Credential metadata recorded in agent identity registry
6. Initial Audit: First use logged and verified within 24 hours

Phase 2: Credential Storage

- Encryption: All credentials encrypted at rest using AES-256 or equivalent
- Access Control: Credentials accessible only to authorized agent runtime and security team
- Secrets Management: Use dedicated secrets management system (e.g., HashiCorp Vault, AWS Secrets Manager)
- No Hardcoding: Credentials never embedded in source code or configuration files
- Environment Isolation: Production and non-production credentials completely separated

Phase 3: Credential Rotation

Rotation Schedule by Risk Classification:

- Critical Risk Agents: 30-day rotation
- High Risk Agents: 60-day rotation
- Medium Risk Agents: 90-day rotation
- Low Risk Agents: 180-day rotation

Rotation Procedure:

1. Generate new credential
2. Deploy new credential to agent runtime
3. Verify agent functionality with new credential
4. Revoke old credential after 24-hour grace period
5. Update registry with new credential metadata
6. Audit rotation completion

Phase 4: Credential Revocation

Immediate revocation required for:

- Suspected credential compromise or exposure
- Agent deactivation or retirement
- Owner departure or role change
- Security incident involving the agent
- Audit finding requiring permission reduction

Revocation Procedure:

1. Disable credential in authentication system (< 15 minutes)
2. Verify agent can no longer authenticate
3. Update registry status to 'Revoked'
4. Document revocation reason and timestamp
5. Notify primary owner and security team
6. Archive revocation event for audit trail

V.5 Continuous Monitoring and Audit Requirements

Per NIST IR 8596 requirements for continuous monitoring of autonomous cyber actors, implement the following surveillance mechanisms:

V.5.1 Real-Time Activity Logging

All agent authentication and authorization events must be logged in real-time:

- Authentication Attempts: Timestamp, agent_id, credential type, success/failure, source IP
- Authorization Decisions: Resource requested, permission evaluated, grant/deny decision
- Credential Usage: Each credential use with full context (what, when, where)
- Permission Escalations: Any attempt to access resources beyond granted permissions

- Anomalous Patterns: Unusual access times, locations, or frequencies

V.5.2 Alert Thresholds and Response

Automated alerts triggered for:

- Failed Authentication: 3 failures within 5 minutes → Alert + temporary lockout
- Unauthorized Access Attempt: Any attempt → Immediate alert to security team
- Credential Compromise Indicators: Multiple simultaneous locations → Alert + revocation
- Permission Boundary Violation: Access beyond scope → Alert + suspension
- Dormant Agent Activity: First use after 30+ days inactivity → Alert + verification

V.5.3 Audit Log Requirements

- Retention: Minimum 7 years (or per regulatory requirement)
- Immutability: Logs append-only, tamper-evident
- Encryption: Logs encrypted in transit and at rest
- Access Control: Audit logs accessible only to security team and auditors
- Review Frequency: Weekly for high-risk agents, monthly for others

V.6 Implementation Checklist

Use this checklist to verify complete implementation of agent identity and credential management protocols:

Identity Registry

- Agent identity registry database established with required fields
- Unique agent_id format and namespace defined
- Ownership chain documented for all existing agents
- Permission inheritance model implemented and tested

Credential Management

- Secrets management system deployed
- Credential rotation schedules defined by risk classification
- Automated rotation procedures tested
- Emergency revocation procedures documented and tested
- All existing agent credentials inventoried and classified

Monitoring and Audit

- Real-time logging infrastructure deployed
- Alert thresholds configured and tested
- Audit log retention and encryption verified
- Security team trained on alert response procedures
- Quarterly audit review process established

□ Integration and Testing

- Integration with existing IAM systems completed
- Test environment deployed mirroring production
- Credential lifecycle tested end-to-end
- Incident response procedures validated
- Documentation updated and accessible to operations team

V.7 Related Resources and Cross-References

AI RMF 2026 References:

- GOVERN 1.5: Agentic AI Governance (Section 3.5)
- Appendix H: Agentic AI Risk Assessment Template
- Appendix S: AI Incident Response Playbook
- GOVERN 1.2: Roles and Responsibilities (ownership chains)

External Standards:

- NIST IR 8596: Cybersecurity Framework Profile for Artificial Intelligence (Draft)
- NIST SP 800-53 Rev 5: Security and Privacy Controls (IA-2, IA-4, AC-2)
- ISO/IEC 27001:2022: Annex A.5 (Access Control), A.9 (Access Control)
- Singapore Model AI Governance Framework for Agentic AI: Section 4 (Identity Management)

APPENDIX X: AI RMF 2026 CROSSWALK TO SINGAPORE MODEL AI GOVERNANCE FRAMEWORK FOR AGENTIC AI

Overview

This appendix maps AI RMF 2026 GOVERN function categories to the SINGAPORE (Model AI Governance Framework for Agentic AI, 2026) and the WEF ('AI Agents in Action: Foundations for Evaluation and Governance, 2025). The Singapore Framework provides guidance on multi-agent coordination, governance boundaries, and system-level oversight that complements AI RMF 2026's comprehensive approach.

Singapore Framework Background

Published: February 2026

Developer: Singapore

Scope: Multi-agent ecosystems, autonomous agent coordination, system-level governance

Key Focus Areas:

- Governance boundaries across organizational and jurisdictional lines
- Governor and auditor agent patterns for automated oversight
- Multi-agent coordination and orchestration challenges
- Cross-organizational participation frameworks
- System-level risk assessment for agent ecosystems

Framework Comparison: AI RMF 2026 vs SINGAPORE

AI RMF 2026 GOVERN	Singapore Framework Element	Integration Notes
GOVERN-1.1: Legal and Regulatory Requirements	Governance Boundaries	SINGAPORE defines which actors have regulatory authority; AI RMF 2026 provides compliance tracking templates
GOVERN-1.2: Organizational Structure	Cross-Organizational Participation	AI RMF 2026 RACI matrix extended for multi-organizational agent ecosystems
GOVERN-2.1: Accountability Structures	Multi-Party Accountability Mechanisms	AI RMF 2026 Agent Profile Cards document accountability; SINGAPORE provides contractual liability guidance
GOVERN-5.1: Oversight & Monitoring	Governor/Auditor Agent Patterns	AI RMF 2026 integrates automated oversight patterns with human governance structures
MAP Function: Agent Profiling	System-Level Risk Assessment	Agent Profile Cards (individual) + System-Level Agentic Profile (ecosystem)
MEASURE-4: System Monitoring	Cross-Agent Metrics	Monitor conflict incidents, transaction anomalies, governor intervention rates
MANAGE-4: Incident Response	Multi-Agent Incident Management	Cover orchestration failures, cascades, governor agent compromise

Key Integration Points

- Governance Charters:** AI Governance Committee charter templates (Appendix E) should explicitly cover multi-agent oversight per SINGAPORE guidance. Add provisions for governor/auditor agent oversight and cross-organizational coordination.
- Policy Development:** AI policies (template in Appendix E) must address governor/auditor agent authority and cross-organizational participation rules.

Document liability allocation and Data Processing Agreements (DPAs) for inter-organizational agent interactions.

3. **Risk Assessment:** Agent Risk Assessment Template (Appendix I) incorporates SINGAPORE system-level risk factors including emergent behaviors, cross-agent conflicts, and orchestration failures.
4. **Monitoring Dashboard:** KPI Dashboard (Appendix T) includes SINGAPORE-recommended metrics for governor/auditor agent intervention rates, cross-agent conflict incidents, and multi-agent transaction anomalies.
5. **Agent Profile Cards:** Section 1.6 of Agent Profile Card template documents SINGAPORE governance requirements: regulatory classification, legal entity responsible, liability limits, governance boundaries, and cross-organizational participation frameworks.

Singapore Framework Core Governance Practices

The Singapore Model AI Governance Framework organizes agentic AI governance into five core practice areas:

1. Governance Boundaries Definition

Purpose: Establish which actors have formal authority over which parts of multi-agent ecosystems

AI RMF 2026 Implementation:

- GOVERN-1.1: Document regulatory obligations across organizational boundaries in Compliance Register
- Agent Profile Card Section 1.6: Identify legal entity responsible for each agent
- GOVERN-1.2: Define organizational authority structures for multi-agent oversight

2. System-Level Oversight Patterns

Purpose: Implement automated oversight through governor and auditor agents

Governor Agents:

- Monitor, validate, and intervene in multi-agent operations
- Authority to veto or roll back agent actions violating governance policies
- Real-time policy enforcement across heterogeneous agent populations

Auditor Agents:

- Continuous monitoring of agent behaviors, transactions, outcomes
- Anomaly detection for cross-agent conflicts and escalation patterns
- Automated reporting of governance violations and emergent risks

AI RMF 2026 Implementation:

- GOVERN-5.1: Define governance policies that governor/auditor agents enforce
- MEASURE-4: Track governor/auditor intervention rates as key governance metric
- MANAGE-4: Establish fallback procedures if governor/auditor agents fail or are compromised

3. Multi-Agent Risk Assessment

Purpose: Evaluate both individual agent risks and emergent system-level behaviors

AI RMF 2026 Implementation:

- MAP Function: Agent Profile Cards for individual agents + System-Level Agentic Profile for ecosystems
- Risk Assessment Template (Appendix I): Includes multi-agent coordination risks, emergent behavior risks
- MEASURE-2: Monitor cross-agent conflict incidents, transaction anomalies

4. Cross-Organizational Participation Rules

Purpose: Define participation frameworks for agents from different vendors/domains

AI RMF 2026 Implementation:

- GOVERN-1.1: Establish contractual liability allocation before agent deployment
- GOVERN-1.2: Create Data Processing Agreements (DPAs) for inter-organizational agent interactions
- Third-Party Governance Toolkit (Appendix L): Vendor assessment for agent providers

5. Multi-Agent Incident Management

Purpose: Respond to orchestration failures, cascades, and cross-agent incidents

AI RMF 2026 Implementation:

- MANAGE-4: Extend incident management to cover multi-agent scenarios
- Incident Response Template: Includes orchestration failures, governor agent compromise
- KPI Dashboard (Appendix T): Track multi-agent incident frequency and response times

Singapore Framework Gap: Environmental Sustainability

CRITICAL FINDING:

The Singapore Model AI Governance Framework for Agentic AI provides comprehensive technical guidance for multi-agent systems but completely omits environmental sustainability considerations. Analysis reveals ZERO references to:

- Carbon footprint or greenhouse gas emissions from AI agent operations
- Energy consumption or computational efficiency of multi-agent ecosystems
- Water usage for data center cooling supporting agent infrastructure
- Environmental impact assessment of AI agent deployment
- Integration with environmental management systems (ISO 14001)
- Lifecycle environmental impact of agent hardware and software

Universal Gap Across Frameworks

This gap is systematic across ALL major agentic AI governance frameworks released 2024-2026:

Framework	Release Date	Agentic AI Coverage	Environmental Sustainability
Singapore Model AI Governance	2026	✓ Comprehensive	✗ Zero Coverage
Singapore MGF Agentic	Jan 2026	✓ Comprehensive	✗ Zero Coverage
NIST AI RMF 1.0	Jan 2023	⚠ Limited	⚠ Minimal
ISO/IEC 42001	Dec 2023	⚠ Limited	⚠ General Only
AI RMF 2026	2026	✓ Comprehensive	✓ Comprehensive

Implementation Recommendation

Organizations implementing Singapore Framework guidance should adopt AI RMF 2026 as their integrating framework:

6. **Use Singapore Framework for multi-agent technical guidance:** Governance boundaries, governor/auditor agent patterns, cross-organizational participation rules
7. **Adopt AI RMF 2026 as comprehensive solution:** Integrates SINGAPORE agentic AI guidance with environmental sustainability layer that SINGAPORE completely omits
8. **Implement environmental governance:** Use AI RMF 2026 Environmental Sustainability Committee (Appendix G), Carbon Footprint Assessment (Appendix J), and Sustainability Metrics (Appendix K) to address Singapore Framework gap
9. **Achieve unified governance:** Single AI Governance Board oversees both agentic AI (SINGAPORE) and environmental sustainability (AI RMF 2026), avoiding governance fragmentation
10. **Gain cost efficiency:** 50-60% cost savings and 40% faster implementation vs. managing separate frameworks for agentic AI and environmental compliance

Conclusion

The Singapore Model AI Governance Framework for Agentic AI provides valuable technical guidance for multi-agent systems and complements AI RMF 2026's comprehensive approach. However, the systematic omission of environmental sustainability across SINGAPORE, Singapore MGF, and other recent frameworks validates AI RMF 2026's unique value proposition as the only framework addressing both dimensions.

Organizations implementing AI RMF 2026 gain:

- Complete SINGAPORE agentic AI technical guidance through AI RMF 2026 integration
- Comprehensive environmental sustainability governance absent from all other frameworks
- Unified approach avoiding governance fragmentation and redundant audits
- ISO certification pathway (42001, 27001, 23894, 14001)
- Significant cost and time savings through integrated implementation

AI RMF 2026 is the only comprehensive framework for responsible AI governance in 2026 and beyond.

Appendix Y: AI RMF 2026 – MULTI-AGENT ECOSYSTEMS

Updated: February 2026 (v1.4 - SINGAPORE Integration)

1. Overview

Multi-agent ecosystems are an emerging class of AI deployments in which multiple agents—often from different vendors and domains—interact, coordinate, and sometimes transact with one another to achieve complex goals. These ecosystems can span organizational boundaries and combine software agents, embodied systems, and human actors into tightly coupled socio-technical systems.

This section extends AI RMF 2026 guidance to cover multi-agent ecosystems (MAEs), building on existing generative and agentic AI coverage, NIST AI RMF profiles, ISO/IEC 42001, and emerging governance work from the World Economic Forum AI Agents in Action: Foundations for Evaluation and Governance, Singapore IMDA (Model AI Governance Framework for Agentic AI), and other international bodies. It describes characteristic risks, introduces system-level profiling concepts, and clarifies how GOVERN, MAP, MEASURE, and MANAGE functions apply when multiple agents operate together.

Version 1.4 Update:

This section has been enhanced to explicitly integrate Singapore Model AI Governance Framework for Agentic AI (2026), which provides comprehensive guidance on governance boundaries, governor/auditor agent patterns, and cross-organizational coordination. The WEF Framework also complements Singapore MGF's technical guidance while AI RMF 2026 adds the critical environmental sustainability layer completely absent from both frameworks.

2. Characteristics and Example Use Cases

2.1 Key Characteristics of Multi-Agent Ecosystems

Multi-agent ecosystems exhibit the following characteristics:

Heterogeneity

- Agents may differ in capabilities, autonomy levels, authority, vendor, and underlying models
- Example: Financial ecosystem combining fraud detection agents (Vendor A), trading agents (Vendor B), and compliance monitoring agents (Vendor C)

Distributed Ownership and Control

- Different organizations may own, deploy, and control different agents within the same ecosystem
- **SINGAPORE Governance Challenge:** Determining governance boundaries and regulatory authority when agents cross organizational lines

Dynamic Interactions and Emergent Behaviors

- Agents coordinate, negotiate, and make collective decisions that may not be predictable from individual agent behavior
- Risk: Emergent system-level behaviors that exceed intended scope or create unintended consequences

Governance Boundaries (Singapore Framework)

- Which actors (organizations, regulators, consortia) have formal authority or obligations over which parts of the ecosystem?
- SINGAPORE guidance emphasizes establishing clear governance boundaries BEFORE deploying multi-agent systems

System-Level Oversight Requirements

- **Singapore Framework:** Governor and auditor agents designed to monitor, validate, and intervene in multi-agent operations
- **Singapore MGF:** Emphasis on continuous monitoring and assurance mechanisms
- **AI RMF 2026:** Integration of automated oversight with human governance structures

2.2 Example Use Cases

Supply Chain Coordination

- Multiple organizations' agents coordinate inventory, logistics, and demand forecasting
- Governance Challenge: Who is liable if coordinated agent decisions cause stockouts or excess inventory?
- **SINGAPORE Requirement:** Contractual liability allocation documented before deployment

Smart City Infrastructure

- Traffic management, energy grid, public safety, and emergency response agents interact
- Governance Challenge: Agents from different municipal departments and private vendors operating in shared physical environment
- **Singapore MGF Requirement:** Clear accountability mechanisms and safety assurance

Healthcare Care Coordination

- Diagnostic agents, treatment planning agents, scheduling agents, and insurance processing agents coordinate patient care
- Governance Challenge: HIPAA compliance when agents share patient data across organizational boundaries
- **AI RMF 2026 Requirement:** Data Processing Agreements (DPAs) between all participating organizations

3. System-Level Profiling for Multi-Agent Ecosystems

The system-level profile complements individual Agent Profile Cards. In risk assessment and governance decisions, organizations SHOULD consider both the highest-risk individual agents and the emergent risks arising from agent interactions.

3.1 System-Level Agentic Profile Template

For multi-agent ecosystems, document the following system-level characteristics:

Profile Element	Documentation Requirements
Ecosystem Participants	List all organizations deploying agents in the ecosystem, their roles, and legal entities responsible
Agent Count and Types	Total number of agents, breakdown by TACO+ classification, highest autonomy/authority levels
Governance Boundaries (SINGAPORE)	Which organizations have authority over which agents? Document regulatory obligations across organizational boundaries
Coordination Mechanisms	How do agents communicate? Protocols, APIs, message formats, orchestration layer if present
Governor/Auditor Agents (SINGAPORE)	List oversight agents, their authority (monitor/veto/rollback), and who operates them
Emergent Risk Assessment	Identify risks that only emerge from agent interactions (e.g., cascading failures, coordination deadlocks, emergent behaviors)
Data Sharing Agreements	Data Processing Agreements (DPAs) between organizations, GDPR/privacy compliance mechanisms
Liability Allocation (SINGAPORE)	Contractual agreements specifying which organization is liable for different failure scenarios
Environmental Impact (AI RMF 2026)	Aggregate carbon footprint, energy consumption, water usage for entire ecosystem (NOT covered by SINGAPORE or Singapore MGF)

4. Applying GOVERN-MAP-MEASURE-MANAGE to Multi-Agent Ecosystems

GOVERN – System-Level Oversight and Accountability

Singapore Framework Integration:

- **Establish multi-agent oversight patterns:** Deploy governor or auditor agents designed to monitor, validate, and, when necessary, veto or roll back agent actions. Governance MUST ensure that governor/auditor agents themselves are subject to oversight to prevent governance bypass.
- **Define governance boundaries:** Document which organizations have authority over which agents, regulatory obligations, and liability allocation before ecosystem deployment.
- **Update governance charters:** AI Governance Committee charter (Appendix E) must explicitly cover multi-agent oversight, governor/auditor agents, and cross-organizational participation rules.
- **Establish Data Processing Agreements:** When agents share data across organizational boundaries, create DPAs ensuring GDPR/privacy law compliance.

Singapore MGF Integration:

- Dimension 2 (Govern for Transparency): Document all governance decisions, agent interactions, and accountability mechanisms

MAP – System-Level Context and Risk Assessment

- **Create Agent Profile Cards:** Document each individual agent using Agent Profile Card template (Section 1.6 includes governance boundaries, legal entity responsible, liability limits)
- **Create System-Level Agentic Profile:** Document ecosystem-level characteristics using template in Section 3.1 above
- **Assess emergent risks:** Identify risks that only appear at system level - cascading failures, coordination deadlocks, multi-agent conflicts, orchestration failures
- **SINGAPORE Risk Factors:** Evaluate heterogeneity risks (different vendors/models), governance boundary ambiguities, cross-organizational coordination challenges

MEASURE – System-Level Monitoring Metrics

Extend monitoring (MEASURE-4) with system-level metrics:

- **Cross-agent conflict incidents:** Number of times agents reach incompatible decisions or deadlocks
- **Agent-initiated transaction anomalies:** Transactions outside expected patterns or authority limits
- **Governor/auditor intervention rates (SINGAPORE):** Frequency of oversight agent veto/rollback actions
- **Ecosystem-level performance:** Collective achievement of goals vs. individual agent performance
- **Environmental metrics (AI RMF 2026 unique):** Aggregate carbon footprint, total energy consumption, water usage for entire ecosystem

Implementation: Add these metrics to KPI Dashboard (Appendix T) and ensure they are reviewed by AI Governance Committee quarterly.

MANAGE – Multi-Agent Incident Response

Extend incident management to cover multi-agent incidents:

- **Orchestration failures:** Central coordination layer fails, agents lose ability to synchronize
- **Multi-agent cascades:** Failure in one agent triggers sequential failures across ecosystem
- **Security incidents (SINGAPORE):** Protocol-level compromise, governor/auditor agent takeover, cross-organizational data breach
- **Emergent behavior incidents:** System exhibits unexpected behaviors not predictable from individual agents

Cross-Organizational Incident Response:

- Establish incident notification protocols across all participating organizations
- Define authority for incident response: who can shut down agents in emergency?
- Document post-incident liability: which organization bears costs of multi-agent failures?

5. Framework Integration Summary

Framework	Key Contribution	AI RMF 2026 Integration
Singapore Model AI Governance Framework	<ul style="list-style-type: none"> • Governance boundaries • Governor/auditor agents • Cross-organizational coordination 	Integrated into GOVERN-5.1, Agent Profile Cards, System-Level Profiling
Singapore MGF Agentic AI	<ul style="list-style-type: none"> • 4-dimension framework • Safety assurance • Transparency requirements 	Integrated into MAP function risk assessment and MEASURE monitoring
AI RMF 2026 (UNIQUE)	<ul style="list-style-type: none"> • Environmental sustainability • Carbon footprint tracking • Energy/water metrics 	ONLY framework addressing environmental impact - completely absent from SINGAPORE and Singapore MGF

6. Critical Finding: Environmental Sustainability Gap

Analysis of SINGAPORE and Singapore MGF agentic AI frameworks reveals ZERO coverage of environmental sustainability:

- **Singapore Framework (2026):** No references to carbon footprint, energy consumption, water usage, or climate impact
- **Singapore MGF Agentic (January 2026):** No environmental sustainability dimension or metrics
- **AI RMF 2026 UNIQUE Value:** Only framework integrating agentic AI governance with comprehensive environmental sustainability requirements

Organizations implementing multi-agent ecosystems need AI RMF 2026 to address BOTH technical governance (SINGAPORE/Singapore MGF) AND environmental impact (completely omitted by other frameworks).

Conclusion

Multi-agent ecosystems require comprehensive governance addressing technical coordination (SINGAPORE), safety assurance (Singapore MGF), and environmental sustainability (AI RMF 2026 unique). AI RMF 2026 integrates guidance from SINGAPORE and Singapore Frameworks while adding the critical environmental layer completely absent from both, providing organizations with a unified governance approach that addresses all dimensions of responsible AI deployment.

Appendix Z: Your AI Agents Don't Have Identities.

Your Governance Framework Doesn't Know They Exist.

Why the GOVERN Function Is Where Agentic AI Governance Lives or Dies—and What Every Current Framework Is Missing

| Bluefox Global Consulting Service, LLC – U.S. Virginia

AI Risk Management Framework 2026 (AI RMF 2026) — GOVERN Implementation Manual

The Question Nobody Is Asking About Their AI Agents

In a *previous article*, I explored why decision survivability—not model reliability—is the real frontier of agentic AI governance. That conversation surfaced the concepts of authority latency, intervention half-life, and the silent drift that makes governance structures decorative rather than functional over time.

This article goes deeper into the structural layer beneath those concepts: the **GOVERN function**—the foundation on which everything else either holds or collapses.

Because here's the uncomfortable reality most organizations haven't confronted: your AI agents are already operating in your environment. They have API keys to your financial systems. They're reading your customer data. They're making decisions that affect real people. And your governance framework—the one designed for traditional AI systems where a human operator sits between every model output and every consequential action—has no structural mechanism for knowing they exist, tracking what they're authorized to do, or intervening when they exceed their boundaries.

That's not a policy gap. It's a governance architecture failure. And it starts at the GOVERN function.

The Agent Identity Problem: Autonomous Actors Without Credentials

Traditional Identity and Access Management (IAM) systems were built for two types of actors: humans and the applications those humans explicitly operate. An employee has credentials. A service account has a defined scope. The relationship between actor and action is traceable.

Agentic AI systems break this model. An AI agent is neither a human nor a traditional application. It plans across multiple steps, executes actions autonomously, accesses databases, calls external APIs, and interacts with other agents—all without a human in the decision loop for each individual action. It's an autonomous cyber actor, as NIST IR 8596 correctly frames it.

And yet, most organizations have deployed agents using shared service accounts, generic API keys, and permission structures inherited from the human who happened to set up the integration. The agent has no unique identity. Its actions are logged under someone else's credentials. Its permissions aren't bounded by its actual purpose—they're bounded by whatever the integrating engineer had access to.

When something goes wrong—and in agentic systems, the question is when, not if—accountability becomes un-reconstructable. Not because nobody was responsible, but because the identity infrastructure was never designed to make responsibility legible.

The AI RMF 2026 GOVERN Implementation Manual addresses this directly through a comprehensive Agent Identity Management Framework built on five foundational principles: unique identification distinguishable from human users, hierarchical ownership linking every agent to a supervising entity, capacity recording documenting whether an agent acts independently or on behalf of a specific human, permission inheritance ensuring agents never receive permissions exceeding those of their authorizing human, and complete audit trails logging all identity delegations and permission grants.

This isn't aspirational guidance. It's an operational protocol with 12 required registry fields per agent, credential lifecycle management from issuance through revocation, rotation schedules calibrated to risk classification, and real-time monitoring with defined alert thresholds—such as automated lockout after three failed authentication attempts within five minutes. These are the controls NIST says organizations need but doesn't tell them how to implement.

The Automation Bias Trap: When Oversight Becomes Rubber-Stamping

In the previous article, I described how authority latency develops when governance structures remain technically intact but response velocity degrades under accumulated trust. The GOVERN Manual reveals the mechanism driving that degradation: **automation bias**.

Automation bias is the tendency to over-trust automated systems, particularly after they've performed reliably for an extended period. For traditional AI systems, this manifests as operators accepting model recommendations without adequate scrutiny. For agentic AI, the stakes are fundamentally different—because the agent isn't just making recommendations. It's executing actions.

When a human overseer is responsible for reviewing an agent's proposed actions, the pattern is predictable: early in deployment, every action is scrutinized. Three months in, the overseer is approving routine actions with a glance. Six months in, they're approving batches. A year in, the human-in-the-loop has become a human-in-name-only.

The GOVERN Manual confronts this directly with structural countermeasures rather than policy aspirations. It requires randomized oversight checks—detailed review of randomly selected agent workflows, not just failures. It mandates rotation of oversight responsibilities to prevent complacency from forming around familiar agents. It calls for regular auditing of human approvals to detect rubber-stamping patterns. And critically, it requires organizations to maintain tradecraft—ensuring that the humans overseeing agents retain the actual skills being automated, so they remain capable of meaningful evaluation rather than performative oversight.

This is where governance either functions or becomes theater. An oversight mechanism that can't detect its own degradation isn't an oversight mechanism. It's a compliance artifact.

Why Operational Teams Cannot Certify Their Own Governance

One of the most consequential architectural decisions in the GOVERN Manual is the structural separation between operational performance and governance certification. This isn't a procedural detail—it's the single most important design principle for governance that survives organizational pressure.

The natural organizational tendency is for the team responsible for an AI system's performance to also assess whether governance constraints remain adequate. This creates an inherent conflict: the people most invested in the system's continued operation are the least likely to identify reasons to restrict it. Smooth operation becomes evidence of adequate governance—a conclusion that feels intuitive and is structurally dangerous.

The GOVERN Manual establishes distinct governance bodies with independent authority. The AI Governance Committee provides cross-functional executive oversight, approving policies and reviewing high-risk systems. The AI Ethics Review Board conducts independent ethical assessment of socially impactful systems. The Agentic AI Committee provides specialized oversight for autonomous systems—with decision authority over all high-risk agentic deployments, agent identity frameworks, and agentic incident response. And the Environmental Sustainability Committee oversees AI environmental impact against independently set sustainability targets.

Each body has a defined charter, specified membership composition, and documented decision authority. The people who build and operate the system are consulted. They are not the ones who certify that governance is working. That separation is what prevents authority latency from becoming structurally inevitable.

Three Oversight Patterns for Agentic AI—and the Criteria That Determine Which One Applies

Not every agentic AI system requires the same level of human oversight. But every agentic AI system requires *explicitly defined and documented* oversight—proportionate to the agent’s autonomy, authority, and environmental complexity. The GOVERN Manual standardizes three patterns and defines the conditions under which each applies:

- **Human-in-the-Loop (HITL):** Agents propose actions, but execution requires explicit human review and approval. This is mandatory for high-stakes decisions affecting safety, fundamental rights, or high-value transactions, and for all systems classified in the highest risk tiers. The human isn’t monitoring—they’re authorizing.
- **Human-on-the-Loop (HOTL):** Agents execute autonomously within defined guardrails while humans monitor behavior and retain intervention capability. This is permitted for lower-tier systems in well-defined environments—but only when effective monitoring, alerts, and emergency stop mechanisms are verified as operational, not merely documented.
- **Human-in-Command:** Humans set objectives, constraints, and risk appetite while retaining authority to pause, reconfigure, or decommission systems entirely. This pattern is required for all multi-agent ecosystems and for any governor or auditor agents with broad visibility or control—because the systems overseeing other systems must themselves remain under human authority.

The critical governance requirement is that these patterns aren’t optional labels. They must be documented in governance policies, Agent Profile Cards, and System-Level Agentic Profiles. Changes to oversight patterns for high-tier agents require AI Governance Committee review and approval. You don’t drift from HITL to HOTL because the system has been running smoothly. You formally propose, review, and approve the transition—with documented rationale and updated risk assessment.

Governor and Auditor Agents: When AI Oversees AI

As multi-agent ecosystems scale—with agents from different vendors and domains interacting, coordinating, and transacting across organizational boundaries—human oversight alone becomes insufficient for real-time governance. The Singapore’s Model AI Governance Framework for Agentic AI introduces two critical concepts that the AI RMF 2026 GOVERN Manual integrates into its operational governance architecture:

- **Governor agents** monitor, validate, and intervene in multi-agent operations in real time. They hold authority to veto or roll back agent actions that violate governance policies, enforcing policy compliance across heterogeneous agent populations.
- **Auditor agents** provide continuous monitoring of agent behaviors, transactions, and outcomes—detecting anomalies, cross-agent conflicts, and escalation patterns, and generating automated reporting of governance violations and emergent risks.

But here’s the governance challenge most organizations miss: governor and auditor agents are themselves AI agents. They require their own identity management, risk assessment, oversight patterns, and—critically—human-in-command governance. A governor agent with veto authority over financial transactions may itself require financial services licensing. An auditor agent with broad visibility across systems creates its own data governance and privacy considerations.

The GOVERN Manual addresses this recursion explicitly. It requires organizations to define governance policies that governor and auditor agents enforce (GOVERN 5.1), track their intervention rates as key governance metrics with defined escalation thresholds (governor intervention rates exceeding 10% or response times exceeding 15 minutes trigger immediate committee review), and establish fallback procedures for when governor or auditor agents themselves fail or are compromised.

Governance of AI that governs AI. That’s not a thought experiment anymore. It’s an operational requirement.

The Gap Every Framework Shares—Except One

While developing the GOVERN Manual and analyzing the latest international frameworks for integration, we identified a finding that should concern every organization planning agentic AI deployment:

Every major agentic AI governance framework released between 2024 and 2026 completely omits environmental sustainability.

The Singapore Model AI Governance Framework for Agentic AI—published February 2026 with comprehensive technical guidance for multi-agent systems—contains zero references to carbon footprint, energy consumption, water usage, or environmental impact assessment. The World Economic Forum-AI Agents in Action: Foundations for Evaluation and Governance—also comprehensive on agentic governance—also has

zero coverage. NIST AI RMF 1.0 and ISO 42001 provide minimal or general-only coverage. Agentic AI Risk-Management Standards Profile (UC Berkeley CLTC) but contains zero coverage of environmental sustainability (carbon footprint, energy consumption, climate impact), reinforcing AI RMF 2026's unique position as the only framework integrating both agentic AI governance and environmental considerations.

This isn't a minor oversight. Agentic AI systems are among the most computationally intensive applications being deployed today. Multi-agent ecosystems multiply that intensity. Large language models serving as agent foundations consume significant energy for both training and inference. And as organizations scale from single-agent deployments to multi-agent ecosystems operating across extended time horizons, the environmental footprint grows—not linearly, but compounding.

The AI RMF 2026 GOVERN Manual is the only framework that integrates environmental sustainability as a core governance function. GOVERN 1.6 establishes a dedicated Environmental Sustainability Committee with oversight authority over AI carbon footprint, energy efficiency targets, resource optimization, and sustainability reporting. It requires organizations to track environmental metrics throughout the AI lifecycle—from model training through operational deployment through decommissioning.

Environmental sustainability isn't a reporting exercise in this framework. It's a governance obligation with the same structural rigor—committee charters, RACI matrices, control mappings, KPIs, and board reporting—that applies to risk management, security, and human oversight.

From Binary Classification to Risk-Proportionate Governance

Most frameworks classify AI systems as either high-risk or not high-risk. The EU AI Act adds prohibited and limited-risk categories. But for agentic AI, binary or tripartite risk classifications are insufficient because the risk surface is multidimensional—action scope, autonomy level, data sensitivity, reversibility, and cascading impact all vary independently.

The GOVERN Manual introduces a four-tier agentic risk classification—Automated Tools, Advisory Agents, Operational Agents, and Critical Agentic Systems—each with proportionate governance requirements. The tier classification is determined through structured risk assessment across nine dimensions: domain and use case risk, data access risk, system access risk, action scope risk, reversibility risk, autonomy risk, task complexity risk, threat modeling, and cascading impact.

Critically, the approved tier isn't just a label. It determines the depth of evaluation in the MAP function, the monitoring expectations in the MEASURE function, and the oversight models (HITL, HOTL, human-in-command) in the MANAGE function. The classification flows through the entire lifecycle. A Critical Agentic System requires Agentic AI Committee approval, extensive human oversight, real-time monitoring, staged rollout, and external audit. A Low-risk Automated Tool requires standard approval processes and basic monitoring.

This proportionality matters for practical adoption. Organizations deploying dozens or hundreds of agents need a governance architecture that applies rigorous oversight where it matters and avoids bureaucratic overhead where it doesn't. Risk-proportionate governance is what makes comprehensive governance scalable.

Glossary

This glossary defines key terms used throughout the AI RMF 2026 GOVERN Procedural Manual. Terms marked with **[NEW]** are additions in version 1.1.

Accountability

The obligation to explain and justify actions and decisions to relevant stakeholders. In AI governance, accountability means clear assignment of responsibility for AI system decisions, outcomes, and impacts throughout the lifecycle.

Action-Space [NEW]

The range of actions an AI agent is permitted to take, determined by the tools, systems, APIs, and transaction authority granted to the agent. A narrow action-space limits agent to read-only operations, while a broad action-space might include database writes, external API calls, and financial transactions.

Agent [NEW]

An AI system that can autonomously plan and execute sequences of actions to achieve specified objectives. Agents typically have access to tools (APIs, databases, search engines), can make multi-step plans, and adapt their approach based on intermediate results.

Agent Boundaries [NEW]

The defined limits on what an AI agent can do, including action-space boundaries (which tools and systems), autonomy boundaries (level of independent decision-making), and environmental boundaries (sandboxed vs. production access).

Agent Identity [NEW]

A unique, trackable identity assigned to an AI agent that enables authentication, authorization, and audit logging. Agent identities are linked to supervising entities (humans or other agents) and cannot exceed the permissions of their supervisors.

Agent Owner [NEW]

The individual or team responsible for a specific AI agent's behavior, actions within defined scope, and ongoing monitoring. The Agent Owner ensures the agent operates safely and appropriately, and responds to agent incidents.

Agent-to-Agent (A2A) [NEW]

Communication protocol enabling multiple AI agents to interact, coordinate, and collaborate on tasks. A2A protocols allow agents to request assistance, share information, and delegate subtasks to other specialized agents.

Agentic AI [NEW]

AI systems that possess autonomous planning and action-taking capabilities to achieve user-defined goals across multiple steps. Agentic AI systems can use tools, interact with external systems, adapt to new information, and operate with varying degrees of human oversight.

AI System

A machine-based system that can make predictions, recommendations, or decisions influencing real or virtual environments for a given set of human-defined objectives. AI systems operate with varying levels of autonomy and may learn from data to improve performance.

AI System Lifecycle

The complete progression of an AI system from initial conception through decommissioning, typically including the following stages: (1) Design and planning - defining objectives, requirements, and architecture; (2) Data collection and preparation - acquiring and processing training data; (3) Model development - training, testing, and validation; (4) Deployment - integration into production environment; (5) Monitoring and maintenance - ongoing performance tracking and updates; (6) Retirement - graceful decommissioning when system is no longer needed or fit for purpose. Each lifecycle stage requires specific governance controls and risk management activities.

Autonomy [NEW]

The degree to which an AI agent can decide when and how to act toward a goal without human intervention at each step. Low autonomy means following detailed Standard Operating Procedures (SOPs); high autonomy means independent judgment and planning.

Automation Bias [NEW]

The tendency for humans to over-trust automated systems, especially after they perform reliably, leading to inadequate oversight. In agentic AI, automation bias can cause human overseers to rubber-stamp agent decisions without proper scrutiny.

Bias

Systematic errors in AI system outputs that create unfair outcomes for different groups. Bias can originate from training data, algorithm design, or deployment context. Harmful bias results in discriminatory treatment based on protected characteristics.

Carbon Footprint [NEW]

The total amount of greenhouse gas emissions (measured in tonnes of CO₂ equivalent) generated by AI system operations throughout its lifecycle, including model training, inference, and infrastructure. Includes both direct emissions (Scope 2) and indirect emissions (Scope 3).

Carbon Intensity [NEW]

The amount of CO₂ equivalent emissions per unit of output, typically measured as grams of CO₂e per inference or per kilowatt-hour of energy consumed. Lower carbon intensity indicates more environmentally efficient AI operations.

Computational Efficiency [NEW]

The ratio of useful computational work performed to energy consumed, typically measured in floating point operations per watt (FLOPs/Watt). Higher computational efficiency means less energy is required to achieve the same AI performance.

Data Poisoning

A type of adversarial attack where malicious actors intentionally corrupt or manipulate training data to cause an AI system to learn incorrect patterns or behaviors. Data poisoning can be used to create backdoors in models, reduce model accuracy, introduce biases, or cause models to misclassify specific inputs. Defenses include data validation, anomaly detection in training data, and diverse data sourcing. Particularly concerning for AI systems that continuously learn from user inputs or publicly available data.

DEIA

Diversity, Equity, Inclusion, and Accessibility - principles ensuring AI teams and systems serve all stakeholders fairly. DEIA in AI development involves diverse team composition, inclusive design practices, and accessible system interfaces.

Differential Privacy

A mathematical framework and set of techniques that provide formal privacy guarantees when analyzing datasets containing personal information. Differential privacy adds carefully calibrated noise to data or query results such that the inclusion or exclusion of any single individual's data does not significantly affect the output, thereby protecting individual privacy while enabling meaningful analysis. Widely used in AI systems to train models on sensitive data without exposing individual records. The privacy guarantee is quantified by the epsilon parameter - smaller epsilon provides stronger privacy but may reduce model utility.

Energy Monitoring [NEW]

The continuous tracking and measurement of electrical energy consumption by AI workloads, including model training and inference operations. Energy monitoring enables identification of optimization opportunities and carbon footprint calculation.

Environmental Sustainability Officer [NEW]

Individual responsible for tracking, managing, and reporting on the environmental impact of AI systems. Coordinates energy monitoring, carbon footprint assessments, and sustainability initiatives across AI operations.

Explainability

The ability to provide meaningful explanations of AI system behavior and decisions to stakeholders. Explainable AI enables users to understand why a system produced a particular output and how it reached a decision.

Federated Learning

A distributed machine learning approach where models are trained across multiple decentralized devices or servers holding local data samples, without exchanging the raw data itself. Instead of centralizing data in one location, federated learning brings the training algorithm to the data, with each participant training a local model on their data and sharing only model updates (gradients or parameters) with a central server for aggregation. Particularly valuable for privacy-sensitive applications like healthcare, finance, and mobile applications where data cannot leave local devices due to privacy regulations or policies.

High-Risk AI

AI systems that pose significant risk to health, safety, fundamental rights, or have substantial societal impact. High-risk AI requires enhanced governance, rigorous testing, human oversight, and regulatory compliance (e.g., under EU AI Act).

ISO/IEC 23894:2023 [NEW]

International standard providing guidance on AI risk management. Defines risk management processes, techniques, and controls. AI RMF 2026 extends ISO 23894 with specific provisions for agentic AI and environmental sustainability.

ISO/IEC 27001:2022 [NEW]

International standard for information security management systems. Provides security controls applicable to AI systems including access control, cryptography, incident management, and resilience.

ISO/IEC 42001:2023 [NEW]

International standard for AI management systems. Defines requirements for establishing, implementing, maintaining, and continually improving an AI management system. Provides certifiable framework for responsible AI governance.

Model Card

A standardized documentation framework for machine learning models that provides essential information about a model's characteristics, intended uses, performance metrics, limitations, and ethical considerations. Introduced by researchers at Google, model cards typically include: model details (architecture, version, training data), intended use cases and out-of-scope applications, performance metrics across different demographic groups, ethical considerations, caveats and recommendations. Model cards enhance transparency, facilitate informed deployment decisions, and support responsible AI practices by making model characteristics and limitations explicit to users and stakeholders.

Model Context Protocol (MCP) [NEW]

Standardized protocol enabling AI agents to access external data sources, tools, and services in a controlled manner. MCP provides structured interface for agents to interact with databases, APIs, and other systems while maintaining security and access control.

Multi-Agent Systems [NEW]

AI architectures where multiple agents work together to accomplish tasks. Common patterns include sequential (agents work in sequence), supervisor (orchestrating agent coordinates workers), and swarm (agents collaborate dynamically without central control).

Privacy-Enhancing Technologies (PETs) [NEW]

Technical methods that protect individual privacy while enabling AI systems to function effectively. Examples include differential privacy, federated learning, homomorphic encryption, and secure multi-party computation.

Prompt Injection

A security vulnerability specific to large language models and AI systems that process natural language inputs, where attackers craft malicious inputs (prompts) designed to manipulate the AI system into ignoring its instructions, revealing sensitive information, or performing unintended actions. Prompt injection attacks exploit the fact that many AI systems cannot reliably distinguish between system instructions and user-provided content. Examples include: instructing a customer service chatbot to ignore safety guidelines, extracting confidential data from an AI assistant's context, or causing an AI to generate harmful content. Defenses include input validation, output filtering, prompt engineering with clear instruction boundaries, and separating user content from system instructions.

RACI Matrix

A responsibility assignment matrix that identifies who is Responsible (performs work), Accountable (ultimately answerable), Consulted (provides input), and Informed (kept updated) for each activity. Used to clarify roles in AI governance.

Red Teaming

A structured adversarial testing methodology where a dedicated team (the 'red team') attempts to identify vulnerabilities, weaknesses, and failure modes in an AI system by simulating realistic attack scenarios and edge cases. In AI governance, red teaming involves deliberately trying to make AI systems produce harmful outputs, violate safety constraints, exhibit biases, leak sensitive information, or fail in unexpected ways. Red team findings inform system improvements, safety measures, and risk mitigation strategies. Red teaming is particularly important for high-stakes AI systems and should be conducted by individuals with diverse backgrounds and perspectives to identify a wide range of potential issues.

Scope 2 Emissions [NEW]

Indirect greenhouse gas emissions from purchased electricity, heat, or cooling consumed by AI operations. Scope 2 emissions are the primary carbon footprint component for AI model training and inference in data centers.

Scope 3 Emissions [NEW]

Indirect greenhouse gas emissions from sources not owned or controlled by the organization but related to AI operations. For AI systems, Scope 3 includes embodied carbon in hardware manufacturing, cloud provider upstream emissions, and employee travel.

Shadow AI

Unauthorized or unmanaged artificial intelligence tools, systems, and services that employees use within an organization without formal IT approval, governance oversight, or risk assessment. Shadow AI typically emerges when employees adopt publicly available AI tools (such as ChatGPT, Claude, or other generative AI services) to improve productivity without going through official procurement and governance processes. Shadow AI poses significant risks including: data leakage of confidential information, compliance violations, lack of accountability, security vulnerabilities, and inability to enforce organizational AI policies. Organizations should address shadow AI through a combination of education, governance policies that balance innovation with risk management, and providing approved AI tools that meet employee needs.

Singapore Model AI Governance Framework [NEW]

Guidance published by Singapore's Infocomm Media Development Authority (IMDA) providing best practices for AI governance. The January 2026 edition for Agentic AI defines four dimensions: assess and bound risks upfront, make humans meaningfully accountable, implement technical controls, and enable end-user responsibility.

Trustworthiness

The degree to which an AI system demonstrates the characteristics of being valid, reliable, safe, secure, resilient, accountable, transparent, explainable, privacy-enhanced, and fair with harmful bias managed. Trustworthy AI systems merit appropriate levels of trust from users and stakeholders.

References

This manual integrates requirements and guidance from the following authoritative sources:

Primary Framework

- National Institute of Standards and Technology (NIST). (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. <https://doi.org/10.6028/NIST.AI.100-1>
- Bluefox Consulting Service, LLC. (2026). AI Risk Management Framework 2026 - Integrated ISO/APM Edition. Extensions for agentic AI governance and environmental sustainability.

International Standards

- International Organization for Standardization (ISO). (2023). ISO/IEC 42001:2023 - Information technology — Artificial intelligence — Management system. Geneva: ISO.
- International Organization for Standardization (ISO). (2022). ISO/IEC 27001:2022 - Information security, cybersecurity and privacy protection — Information security management systems — Requirements. Geneva: ISO.
- International Organization for Standardization (ISO). (2023). ISO/IEC 23894:2023 - Information technology — Artificial intelligence — Guidance on risk management. Geneva: ISO.
- International Organization for Standardization (ISO). (2018). ISO 14064-1:2018 - Greenhouse gases — Part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals. Geneva: ISO.
- Institute of Electrical and Electronics Engineers (IEEE). (2021). IEEE 7000-2021 - Model Process for Addressing Ethical Concerns During System Design. IEEE Standards Association.
Context: *First in IEEE's 7000 series on ethical AI design. Provides systematic process for identifying and addressing ethical values in system design.*
- Partnership on AI. (2023). Responsible Practices for Synthetic Media: A Framework for Collective Action. <https://partnershiponai.org/responsible-practices-for-synthetic-media/>
Context: *Industry framework for governance of generative AI and synthetic media, addressing provenance, transparency, and responsible deployment.*

Regulatory Frameworks

- European Parliament and Council. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence (Artificial Intelligence Act). Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- U.S. Securities and Exchange Commission (SEC). (2024). The Enhancement and Standardization of Climate-Related Disclosures for Investors. Final Rule. 17 CFR Parts 210, 229, 232, 239, and 249.
- European Parliament and Council. (2022). Directive (EU) 2022/2464 on Corporate Sustainability Reporting (CSRD). Official Journal of the European Union.
- Government of Canada. (2019). Directive on Automated Decision-Making. Treasury Board of Canada Secretariat. <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>

Context: *Canadian government directive requiring impact assessments and transparency for automated decision systems. Provides useful framework for public sector AI governance.*

Governance Frameworks

- Infocomm Media Development Authority (IMDA), Singapore. (2026). Model AI Governance Framework for Agentic AI. Second Edition. Singapore: IMDA. <https://www.imda.gov.sg/resources/model-ai-governance-framework>
- Infocomm Media Development Authority (IMDA), Singapore. (2024). Model AI Governance Framework. Second Edition. Singapore: IMDA.
- Organisation for Economic Co-operation and Development (OECD). (2019). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. Paris: OECD Publishing.

Technical Guidance

- National Institute of Standards and Technology (NIST). (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1. <https://doi.org/10.6028/NIST.AI.600-1>
- National Institute of Standards and Technology (NIST). (2020). Security and Privacy Controls for Information Systems and Organizations. NIST Special Publication 800-53, Revision 5. <https://doi.org/10.6028/NIST.SP.800-53r5>
- Greenhouse Gas Protocol. (2015). Corporate Value Chain (Scope 3) Accounting and Reporting Standard. World Resources Institute and World Business Council for Sustainable Development.

- Greenhouse Gas Protocol. (2004). A Corporate Accounting and Reporting Standard. Revised Edition. World Resources Institute and World Business Council for Sustainable Development.
- National Institute of Standards and Technology (NIST). (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) Playbook. NIST AI 100-2e2023. https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook

Context: *Practical implementation guidance and examples for organizations using the NIST AI RMF. Complements the core framework with actionable steps and real-world applications.*

- Information Commissioner's Office (ICO), United Kingdom. (2024). Guidance on AI and Data Protection. ICO AI Auditing Framework. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/>

Context: *UK regulator's guidance on GDPR compliance for AI systems, including accountability, transparency, and data protection by design principles.*

Research and Best Practices

- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645-3650.
- Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). Carbon Emissions and Large Neural Network Training. arXiv preprint arXiv:2104.10350.
- Wu, C. J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Aga, F., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H. H. S., ... & Hazelwood, K. (2022). Sustainable AI: Environmental Implications, Challenges and Opportunities. Proceedings of Machine Learning and Systems, 4, 795-813.
- Solaiman, I., Talat, Z., Agnew, W., Ahmad, L., Baker, D., Blodgett, S. L., Daumé III, H., Dodge, J., Evans, I., Hooker, S., Jernite, Y., Luccioni, A. S., Lusoli, A., Mitchell, M., Newman, J., Png, M. T., Strait, A., & Vassilev, A. (2023). Evaluating the Social Impact of Generative AI Systems in Systems and Society. arXiv preprint arXiv:2306.05949.
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., ... & Gabriel, I. (2021). Ethical and social risks of harm from Language Models. arXiv preprint arXiv:2112.04359.

U.S. Sector-Specific Privacy and Data Protection Regulations

- California State Legislature. (2018). California Consumer Privacy Act of 2018 (CCPA). California Civil Code §§ 1798.100–1798.199. <https://oag.ca.gov/privacy/ccpa>
- U.S. Department of Health and Human Services. (1996). Health Insurance Portability and Accountability Act of 1996 (HIPAA). Public Law 104-191. 45 CFR Parts 160, 162, and 164. <https://www.hhs.gov/hipaa/>
- U.S. Congress. (1999). Gramm-Leach-Bliley Act (GLBA). Public Law 106-102. 15 U.S.C. §§ 6801–6809. <https://www.ftc.gov/business-guidance/privacy-security/gramm-leach-bliley-act>
- U.S. Congress. (1970). Fair Credit Reporting Act (FCRA). 15 U.S.C. § 1681 et seq. <https://www.ftc.gov/legal-library/browse/statutes/fair-credit-reporting-act>

Additional Resources

- AI Incident Database. (2024). Partnership on AI. <https://incidentdatabase.ai/>
- ML Commons. (2024). AI Safety Benchmark. <https://mlcommons.org/benchmarks/ai-safety/>
- Stanford Institute for Human-Centered Artificial Intelligence (HAI). (2024). AI Index Report 2024. Stanford University.
- Future of Life Institute. (2023). Pause Giant AI Experiments: An Open Letter. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- MITRE Corporation. (2024). ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems. <https://atlas.mitre.org/>

•

-
- Agentic AI Risk-Management Standards Profile (UC Berkeley CLTC)

Context: *Comprehensive knowledge base of adversarial tactics, techniques, and case studies for AI systems. Essential reference for AI security and red teaming.*

END OF DOCUMENT

AI RMF 2026 GOVERN Procedural Manual v1.4

COMPLETED EDITION

February 2026

