CRI Financial Services AI Risk Management Framework   
Ver. 1.0

Control Objective Reference Guide –   
Initial AI Adoption Stage

February 2026

**

Developed in collaboration with the Financial Services Sector Coordinating Council and the U.S. Department of the Treasury



Cyber Risk Institute  
Washington, D.C.

[CyberRiskInstitute.org](http://www.cyberriskinstitute.org/)

**INTRODUCTION**

This document serves as a detailed reference guide, providing a comprehensive breakdown of each of the 21 Control Objectives for the Initial AI Adoption Stage as defined in the Financial Services AI Risk Management Framework (FS AI RMF). This includes the contents of both the FS AI RMF Guidebook and the Risk and Control Matrix.

The details of each Control Objective have been carefully structured to provide a holistic understanding, from foundational alignment to practical implementation guidelines and examples.

**IMPORTANT**: Note that, because this document has been customized to include the Initial AI Adoption Stage only, the numbers will not be sequential. They have been left as-is for the purposes of connecting back to the complete framework which includes all stages.

Each Control Objective contains the following information:

* **NIST AI RMF Alignment:** Identifying the corresponding NIST Function, Category, and Sub-Category, establishing its place within the broader framework.
* **Control Objective Definition:** Including a unique Control Objective ID, a Control Objective Name, and a clear description outlining its purpose and scope.
* **Link to the NIST AI RMF Playbook:** A direct Link to the NIST AI RMF Playbook for additional context and guidance.
* **Adoption Stage Applicability:** Indicating which Adoption Stage(s) apply (Initial, Minimal, Evolving, and/or Embedded) and the rationale, enabling organizations to focus on the most relevant information.
* **Alignment to Risks and AI Trustworthy Principle(s):** Highlighting the underlying AI Trustworthy Principle it upholds, alongside an associated Risk Name and a descriptive Risk Statement to clarify the specific threats the objective aims to mitigate.
* **Actionable Implementation Guidelines:** Providing detailed instructions, practical steps, and recommendations for effectively achieving the Control Objective.
* **Example Controls and Effective Evidence:** Five example Controls, each with a clear description of the action, complemented by four examples of Effective Evidence to demonstrate successful implementation and adherence.

**DISCLAIMER**

Please note that the examples of Controls and Effective Evidence provided in this Control Objective Reference Guide are intended to serve only as illustrative guides. They are not exhaustive, exclusive, or all-encompassing lists and should not be interpreted as a definitive guarantee for meeting regulatory or audit expectations. The suitability and applicability of these examples may vary across organizations, and different Evidence might provide greater value in specific contexts.

Users of the FS AI RMF should critically assess the adequacy and relevance of the Controls and Evidence presented to adequately substantiate their responses, weighing it against the resources and investment required to produce such Evidence. It is important to note that there should be no presumption that all, or even most, of the examples of Controls and Effective Evidence included are necessary, advisable, or appropriate for supporting the organization’s compliance assertions. Organizations are encouraged to tailor their approach to their specific context and regulatory environment.

**CONTENTS**

[UPDATE SCHEDULE AND VERSIONS 4](#_Toc221016937)

[**GOVERN / Establishing Key Policies & Processes / Legal and Regulatory Compliance** 5](#_Toc221016938)

[**GOVERN / Establishing Key Policies & Processes / Defining Trustworthy AI Policies** 11](#_Toc221016939)

[**GOVERN / Establishing Third-Party Risk Management for the AI System / Establishing Third-Party Risk Management Processes** 16](#_Toc221016940)

[**MAP / Understanding the Operating Context / Defining the Operating Context** 18](#_Toc221016941)

[**MAP / Understanding the Operating Context / Bolstering Cross-Functional Perspectives on the Operating Context** 26](#_Toc221016942)

[**MAP / Understanding the Operating Context / Defining the Mission and Goals** 28](#_Toc221016943)

[**MAP / Understanding the Operating Context / Defining Real-World Business Value** 30](#_Toc221016944)

[**MAP / Understanding the Operating Context / Defining and Communicating User Requirements** 32](#_Toc221016945)

[**MAP / Understanding the AI System / Defining AI System Task and Methodology** 34](#_Toc221016946)

[**MAP / Understanding the AI System / Defining AI System Utility and Oversight Needs** 36](#_Toc221016947)

[**MAP / Understanding the AI System / Applying Scientific Methodologies** 38](#_Toc221016948)

[**MEASURE / Methods and Metrics / External Assessment of Measurement Approaches** 40](#_Toc221016949)

[**MEASURE / Evaluating AI Systems / Measuring Privacy** 43](#_Toc221016950)

[**MEASURE / Evaluating AI Systems / Measuring Nondiscrimination** 45](#_Toc221016951)

[**MEASURE / Tracking AI Risks / Enabling Feedback, Appeal, and Override** 48](#_Toc221016952)

# **UPDATE SCHEDULE AND VERSIONS**

The FS AI RMF and this Control Objective Reference Guide are living documents. CRI will review the content and usefulness regularly to determine if an update is or may be necessary. The nature and extent of the update will be reviewed with the CRI Board of Directors, members, stakeholders, trade associations, relevant third parties, and others where appropriate. All input received will be considered.

Updates will be tracked in a Version Control Log using a consistent, two-number versioning schedule. The Version Control Log will include: (1) the version number; (2) the date of the change(s); (3) a description of the change(s); (4) the updated section(s); and (5) the relevant page number(s). For example:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Date of Change(s)** | **Description of Change(s)** | **Relevant Section(s)** | **Page Number(s)** |
| 1.0 | XX/XX/XXX | [Change] | [Section] |  |

General comments and change requests should be sent to [Secretariat@cyberriskinstitute.org](mailto:Secretariat@cyberriskinstitute.org) and will be reviewed at least semi-annually.

**GOVERN / Establishing Key Policies & Processes / Legal and Regulatory Compliance**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| GV-1.1.1 | AI Legal, Regulatory, and Policy Integration | The organization identifies, monitors, and integrates applicable laws, regulations, contractual obligations, and sector requirements into policies, procedures, and operations governing AI. This includes updating governance artifacts and aligning operational practices as requirements evolve.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/govern/#govern-1-1) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations are beginning to recognize the importance of monitoring relevant laws, regulations, and requirements related to AI. However, the process is likely to be informal, ad-hoc, and lacking systematic implementation. |
| **Stage 2: Minimal** | Yes | Organizations at the minimal stage have established a basic process for tracking some relevant AI-related laws and regulations. However, the approach may not be comprehensive, consistently applied, or well-documented. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations have a well-defined and documented process for actively monitoring and maintaining awareness of AI-related legal and regulatory requirements. They likely utilize technological tools or legal expertise to systematically track, update, and apply these requirements. |
| **Stage 4: Embedded** | Yes | Organizations at the embedded stage have fully implemented, automated, or integrated processes for ensuring constant awareness of AI-related legal and regulatory changes. These processes are deeply embedded into their operations and decision-making across various functions. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Regulatory Monitoring Failure | AI laws, regulations, policies, and other obligations are not monitored effectively or adhered to |

**Implementation Guidelines:**

Establish a process to actively monitor and review relevant laws, regulations, contractual obligations, and sector requirements impacting AI on a regular basis. Assign responsibility to a specific team or individual, such as a legal or compliance department, to track updates and communicate them to relevant stakeholders.

Participate in industry forums, subscribe to regulatory alerts, and engage with legal experts to stay informed of emerging developments in the AI regulatory landscape. Ensure that organizational policies and operations are aligned with these requirements, and incorporate safeguards to prevent the use of AI in ways that could violate legal or organizational standards. Restrict activities that may lead to unfair or deceptive practices, and enforce governance around AI deployment.

**Example Controls and Evidence:**

**Example Control 1:** Regulatory Monitoring Program

**Description:** Create a cross-functional regulatory monitoring team (legal, compliance, privacy, security, and AI/ML leads) responsible for identifying, assessing, and communicating new or changing laws, regulations, and sector guidance that affect AI systems. This team should maintain a documented risk register and escalation process for items requiring organizational action.

**Examples of Effective Evidence:**

• Charter or terms of reference for the regulatory monitoring team.

• Meeting minutes, agendas, and action item logs showing regulatory reviews.

• AI regulatory risk register with assessments and assigned owners.

• Escalation records showing regulatory issues forwarded to senior leadership or the board.

**Example Control 2:** External Intelligence Subscriptions and Alerts

**Description:** Subscribe to and maintain active feeds from regulatory databases, legal alert services, industry associations, and standards bodies to receive timely notifications about relevant legal and sector developments. Assign personnel to triage and summarize alerts for stakeholders.

**Examples of Effective Evidence:**

• Subscription invoices/agreements for legal/regulatory intelligence services (e.g., regulatory databases, industry alerts).

• Alert digest emails or summaries distributed to stakeholders.

• Logs or trackers showing alerts received and triage actions taken.

• Assigned roles/responsibilities documentation for alert triage.

**Example Control 3:** Contractual Obligation Management for AI Clauses

**Description:** Implement a contract review and tracking process focused on AI-specific contractual obligations (e.g., data usage, model performance, audit rights, liability, and regulatory compliance) to ensure contractual commitments are captured, assigned, and monitored.

**Examples of Effective Evidence:**

• Repository of AI-related contract clauses or redlines and associated metadata.

• Contract obligation register showing AI-specific obligations, owners, and due dates.

• Contract review checklist or template that includes AI, privacy, and regulatory items.

• Records of contract review meetings and sign-offs.

**Example Control 4:** Policy and Procedure Review Cycle Linked to Regulatory Changes

**Description:** Maintain documented AI governance policies and procedures with a formal review cycle triggered by regulatory or sector changes; ensure updates are versioned, approved, and communicated to affected teams.

**Examples of Effective Evidence:**

• Policy documents with version history and approval signatures.

• Change logs showing policy updates mapped to specific regulatory or guidance changes.

• Communication records (email, intranet posts, training) announcing policy revisions.

• Training completion reports showing staff acknowledged updated policies.

**Example Control 5:** Regulatory Impact Assessments and Control Implementation Plans

**Description:** For material regulatory changes, perform documented regulatory impact assessments (RIAs) for affected AI systems that identify gaps, remediation actions, timelines, and resource requirements; track remediation to closure.

**Examples of Effective Evidence:**

• Completed regulatory impact assessment reports for specific regulations.

• Remediation/action plans with timelines, owners, and status updates.

• Project management artifacts (Gantt charts, tickets) tracking remediation tasks.

• Validation evidence showing implemented changes (test reports, configuration baselines).

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| GV-1.1.2 | AI Compliance Responsibilities Management | The organization clearly assigns, documents, and regularly validates its compliance responsibilities for AI-related legal, regulatory, contractual requirements, and internal organizational AI policies.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/govern/#govern-1-1) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations' responsibilities for compliance are typically assigned in an informal or ad-hoc manner. Documentation and regular validation might not be consistent. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations have started to assign and document certain AI compliance responsibilities. However, the process may be incomplete, inconsistently applied, and lack regular validation. |
| **Stage 3: Evolving** | Yes | Organizations at the evolving stage have comprehensively assigned, documented, and communicated AI compliance responsibilities across relevant functions. They conduct periodic validations to ensure continued accountability and effectiveness. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations assign and document AI compliance responsibilities that are detailed, systematic, and integrated into the organization's overall compliance management framework. A robust validation process ensures consistent adherence and continuous improvement. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Unclear Compliance Responsibilities | Compliance responsibilities for AI laws and other obligations are not clearly assigned or validated |

**Implementation Guidelines:**

Define and document roles and responsibilities to comply with AI-related legal, regulatory, contractual requirements, and internal organizational AI policies. Create a matrix or chart that maps specific compliance tasks to designated individuals or teams, including timelines for completion.

Emphasize the restriction of non-compliant or high-risk AI activities (e.g., assigning oversight roles specifically focused on preventing misuse). Establish accountability mechanisms as new governance practices and controls are implemented.

Conduct regular validation of these responsibilities through periodic reviews or audits to ensure that assigned parties fulfill their obligations effectively, including adherence to internal policies. Establish clear communication channels to facilitate collaboration and information sharing among relevant stakeholders. Consider implementing escalation processes for addressing misuse or non-compliance to help mitigate risks and reinforce commitment to AI Trustworthy Principles.

**Example Controls and Evidence:**

**Example Control 1:** Roles and Responsibilities Matrix (RACI) for AI Compliance

**Description:** Maintain a documented RACI (or equivalent) matrix that maps AI-related legal, regulatory, contractual, and internal policy obligations to specific roles or functions (e.g., data owners, legal, privacy, security, AI/ML engineers, business owners). Review and update the matrix at least annually or when organizational changes occur.

**Examples of Effective Evidence:**

• RACI matrix or responsibility matrix with sign-off from leadership.

• Version history showing periodic reviews and updates.

• Organizational charts or role descriptions aligning named individuals to RACI entries.

• Meeting minutes where RACI changes or assignments were approved.

**Example Control 2:** Documented Ownership and Escalation Procedures

**Description:** Define and publish procedures that specify how ownership of compliance obligations is assigned, how owners accept responsibility, and the escalation path for unresolved compliance gaps or incidents. Include SLA (Service Level Agreement) timelines for owner responses.

**Examples of Effective Evidence:**

• Published ownership and escalation procedure document.

• Records of owner acknowledgements (e.g., signed statements, email confirmations).

• Escalation logs showing issues raised, escalated, and resolved.

• SLA or response-time metrics and reports demonstrating compliance with timelines.

**Example Control 3:** Compliance Responsibility Onboarding and Offboarding Process

**Description:** Integrate AI compliance responsibilities into employee onboarding, role changes, and offboarding so that accountability is transferred and documented when personnel or contractors change roles.

**Examples of Effective Evidence:**

• Onboarding/offboarding checklists that include AI compliance handover tasks.

• Human Resources (HR) or access-management records showing role changes and completed handovers.

• Dated acknowledgement forms or training completion records tied to role assignments.

• Access change tickets or logs showing transfer of system/contract permissions.

**Example Control 4:** Periodic Validation and Assurance Reviews

**Description:** Conduct scheduled validation activities (e.g., internal audits, targeted control testing, compliance assessments) to verify that assigned owners are fulfilling AI compliance responsibilities and that controls are operating effectively. Document findings and remediation.

**Examples of Effective Evidence:**

• Internal audit or assessment reports focused on AI compliance roles and controls.

• Test results and control evidence sample files showing owner activities (e.g., policy reviews, contract checks).

• Issue logs with remediation plans, owners, and closure evidence.

• Management responses and board-level briefings on validation outcomes.

**Example Control 5:** Automated Tracking and Compliance Registry

**Description:** Implement a centralized, searchable compliance registry or tool that records AI-related obligations, assigned owners, due dates, status, and evidence attachments; enable automated reminders and periodic owner attestations.

**Examples of Effective Evidence:**

• Screenshots or exports from the compliance registry showing obligations, owners, and statuses.

• Automated reminder logs and owner attestation records (timestamped).

• Linked evidence attachments (policy approvals, contract excerpts) accessible from the registry.

• Reports demonstrating overdue items, resolution history, and metrics tracked over time.

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| GV-1.1.6 | Data Lifecycle and Retention | The organization ensures that data retention policies, privacy requirements, and legal obligations governing data lifecycle management are integrated into its AI system lifecycle processes. It maintains records of data retention periods, manages data deletion or archiving in compliance with privacy norms, and verifies that training datasets adhere to these policies, especially when data subjects exercise their rights or when legal or privacy obligations change.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/govern/#govern-1-1) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations understand that incorporating data retention and privacy policies into AI lifecycle processes is fundamental for compliance and risk mitigation. At this stage, organizations focus on establishing basic controls to manage data in accordance with existing legal and privacy requirements, ensuring that data used in AI systems is handled responsibly and documented properly. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations implement basic data management policies for AI data, ensuring compliance with privacy requirements in limited contexts. |
| **Stage 3: Evolving** | Yes | Organizations at the evolving stage develop more comprehensive data lifecycle management processes as AI systems handle more sensitive and extensive data, including records of data retention and deletion. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations require automated and integrated data management, continuous compliance verification, and adaptive policies to swiftly respond to legal and privacy changes, ensuring ongoing adherence across the entire data lifecycle. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Privacy-Enhanced | Data Lifecycle Control Gaps | Data lifecycle controls fail to meet retention, privacy, or legal obligations |

**Implementation Guidelines:**

Establish clear procedures for integrating comprehensive data management, privacy requirements, and data retention policies into all stages of the AI system lifecycle, from data collection to deployment and decommissioning. This includes documenting data provenance, tracking data lineage and transformations, implementing information sharing restrictions, and managing data deletion or archiving in accordance with legal, regulatory, and privacy obligations. Develop processes to ensure ongoing compliance with privacy principles and facilitate activities such as honoring data subjects’ rights, including the right to delete data, while maintaining transparency and control over data uses.

Implement automated or manual processes to monitor adherence to data management and retention policies, with safeguards to prevent unauthorized data handling or deletion. Develop policies to enable timely data archiving or deletion when data subjects exercise their rights or when legal or privacy obligations change, ensuring continued compliance and risk mitigation.

Maintain comprehensive records of data processing activities (e.g., including data provenance, lineage, transformations, retention periods, deletion or archiving dates, data sharing decisions) aligned with privacy frameworks. Conduct regular audits to verify that training datasets and other data used in AI systems comply with data management and privacy requirements, and that data handling practices adhere to established procedures. When changes in legal or regulatory requirements occur, promptly review and update policies and practices, providing training to relevant personnel to sustain compliance and privacy protection.

**Example Controls and Evidence:**

**Example Control 1:** Data Retention and Classification Policy for AI Data

**Description:** Establish and publish a data retention and classification policy specifically covering datasets used in AI development, training, and inference, defining retention periods, archival criteria, legal bases, and disposal procedures mapped to data sensitivity and regulatory requirements.

**Examples of Effective Evidence:**

• Published data retention and classification policy with AI-specific sections and approval signatures.

• Data classification inventory showing datasets, sensitivity labels, retention periods, and legal basis.

• Version history linking policy updates to regulatory or business-driver changes.

• Communications/training artifacts announcing the policy to data stewards and AI teams.

**Example Control 2:** Data Lifecycle Controls Integrated into the AI SDLC

**Description:** Integrate retention, anonymization/pseudonymization, and deletion checkpoints into the AI system development lifecycle (requirements, data collection, preprocessing, training, deployment, monitoring, decommissioning) with mandatory attestations before progression.

**Examples of Effective Evidence:**

• SDLC stage-gate checklists showing completed data lifecycle controls and attestation records.

• Workflow/ticketing logs requiring and recording anonymization or retention actions prior to model training or deployment.

• Data processing maps or dataflow diagrams annotated with lifecycle control points.

• Sign-offs from data protection/privacy owners at relevant SDLC stages.

**Example Control 3:** Automated Retention Enforcement and Deletion/Archival Mechanisms

**Description:** Implement technical controls and tooling that enforce retention schedules (automatic archival or deletion), track deletion requests, and produce tamper-evident logs demonstrating actions taken on training and production datasets.

**Examples of Effective Evidence:**

• System logs or audit trails showing automated deletion/archival jobs with timestamps and dataset identifiers.

• Tooling configuration screenshots (retention rules, schedules) and change history.

• Reports of completed deletion/archival actions in response to retention expiry or data subject requests.

• Hashes/checksums or snapshots proving dataset state before and after deletion/archival (where permitted).

**Example Control 4:** Data Subject Rights and Workflow for AI Datasets

**Description:** Maintain documented processes and tooling to handle data subject requests (access, correction, deletion, portability) that ensure requests affecting AI training or inference data are identified, propagated to owners, and remedied across datasets/models (including retraining or model explainability adjustments where necessary).

**Examples of Effective Evidence:**

• Logged data subject requests showing intake, impact analysis on AI datasets/models, and closure evidence.

• Workflow tickets and communications documenting coordination between privacy, data engineering, and ML teams for fulfillment.

• Records of model retraining, feature removal, or inference adjustments performed as a result of requests.

• Standard operating procedures (SOPs) and SLA metrics for handling requests affecting AI data.

**Example Control 5:** Periodic Verification, Auditing, and Reconciliation of AI Training Data Compliance

**Description:** Conduct regular audits and reconciliations to verify training and production datasets comply with retention and privacy policies (sampling, lineage checks, consent verification), and update records when legal obligations change; feed findings into remediation plans.

**Examples of Effective Evidence:**

• Audit reports showing dataset samples, lineage verification, consent/ legal-basis checks, and findings.

• Reconciliation logs linking datasets in storage to the retention inventory and noting discrepancies and remediation actions.

• Remediation tickets and closure evidence for non-compliant datasets (e.g., deleted/archived or reclassified).

• Governance reports showing tracking of policy-change-driven updates (e.g., changes in retention periods) and confirmation of applied changes across datasets and models.

**GOVERN / Establishing Key Policies & Processes / Defining Trustworthy AI Policies**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| GV-1.2.1 | AI Principles and Policy Integration | The organization defines and documents its own AI Trustworthy Principles, which are then integrated into relevant policies, processes, and practices across key functions (e.g., HR, IT, legal, compliance).  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/govern/#govern-1-2) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | Organizations at the initial stage start defining high-level AI principles to provide an essential foundation for responsible AI deployment. Early-stage organizations typically formalize these principles to guide future risk management efforts, laying groundwork for consistent and responsible AI practices across various functions. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations have started to fully define and document their own AI Trustworthy Principles. However, these principles may not be fully developed or consistently integrated into relevant policies, processes, and practices across key functions. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations have defined and documented comprehensive AI Trustworthy Principles. These principles are integrated into relevant policies, processes, and practices across key functions, although their implementation may not yet be fully consistent or mature. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations define and document AI Trustworthy Principles and integrate them into relevant policies, processes, and practices. and it is fully ingrained in the organization's culture and operations. These principles are consistently applied, regularly reviewed and updated based on feedback and best practices, and are deeply embedded into all aspects of the organization's AI-related activities. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Undefined AI Trustworthy Principles | AI Trustworthy Principles are not defined or integrated organization-wide |

**Implementation Guidelines:**

Define and document your organization's AI Trustworthy Principles, aligning them with industry best practices and your specific context and values. Communicate these principles to relevant stakeholders and integrate them into key functions such as business lines, HR, IT, legal, and compliance.

Update existing policies, create new ones as necessary, and provide training to ensure the principles are understood and consistently applied throughout the organization.

**Example Controls and Evidence:**

**Example Control 1:** AI Trustworthy Principles Statement and Governance Approval

**Description:** Create a formal, published statement of the organization’s AI Trustworthy Principles (e.g., fairness, transparency, accountability, privacy, safety) and obtain governance-level approval (executive/board) to ensure organizational endorsement and visibility.

**Examples of Effective Evidence:**

• Published AI Trustworthy Principles document with date and version.

• Board or executive meeting minutes recording approval or endorsement.

• Internal communications (intranet announcement, emails) announcing the principles.

• Executive sponsor or C-suite sign-off on the principles.

**Example Control 2:** Integration of Principles into Policies and Standards

**Description:** Map the AI Trustworthy Principles to existing and new organizational policies, standards, and standard operating procedures (SOPs) (privacy, data governance, procurement, vendor management, HR) so the principles are operationalized and enforced through formal documents.

**Examples of Effective Evidence:**

• Policy documents (privacy, data retention, procurement) with sections referencing or embedding AI principles.

• Standards or SOPs showing specific controls derived from the principles (e.g., fairness testing in ML testing standards).

• Policy mapping matrix linking each principle to affected policies/processes and owners.

• Version-controlled policy change records showing when principles were integrated.

**Example Control 3:** Role-Based Procedures and Responsibilities for Principle Implementation

**Description:** Define role-specific procedures and responsibilities (e.g., hiring, performance reviews, model development, legal review, procurement) that translate high-level principles into concrete actions and assign accountable owners.

**Examples of Effective Evidence:**

• RACI or responsibility matrices detailing owner tasks tied to each principle.

• Job descriptions, SOPs, or checklists for roles (ML engineers, product owners, legal) showing principle-related duties.

• Signed attestations or acknowledgements from role owners accepting responsibilities.

• Operational runbooks or playbooks demonstrating how roles apply principles in routine tasks.

**Example Control 4:** Principle-Based Requirements in Procurement and Third-Party Management

**Description:** Embed AI Trustworthy Principles criteria into procurement, vendor selection, and contract clauses (e.g., requirements for model interpretability, data provenance, compliance with principles) and require evidence from suppliers.

**Examples of Effective Evidence:**

• Procurement templates and Requests for Information/Proposal (RFI/RFP) including AI Trustworthy Principles and mandatory supplier attestations.

• Contracts or contract clauses referencing compliance with the organization’s AI Trustworthy Principles.

• Supplier assessments or evidence packages demonstrating adherence (model cards, certifications, SOC reports).

• Vendor approval records showing principle-based evaluation outcomes.

**Example Control 5:** Training, Awareness, and Compliance Monitoring Program

**Description:** Deliver role-specific training on the AI Trustworthy Principles, embed principle checks into SDLC/gates, and run monitoring/audit programs to measure adoption and effectiveness, with mechanisms to remediate gaps.

**Examples of Effective Evidence:**

• Training materials, attendance records, and completion certificates for staff and relevant stakeholders.

• SDLC stage-gate checklists or automated controls that include principle compliance attestations.

• Monitoring or audit reports assessing principle adoption, with remediation plans and closure evidence.

• Metrics dashboards showing KPIs (training completion rates, number of principle-related findings, remediation times)

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| GV-1.2.2 | AI Terminology and Scope Glossary | The organization establishes, maintains, and periodically updates a glossary of AI-related terms and concepts and clearly defines the scope, purpose, and intended/acceptable uses of its AI systems. This information is referenced and reflected consistently across organizational policies, standards, other relevant documentation, and training programs.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/govern/#govern-1-2) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | Organizations at the initial stage begin establishing a shared understanding of AI terminology and scope is crucial for effective communication and policy development. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations have started to establish a glossary of AI-related terms and concepts and define the scope, purpose, and intended/acceptable uses of its AI systems. However, this information may not be fully developed, consistently referenced, or reflected across organizational policies, standards, documentation, and training programs. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations have established a comprehensive glossary of AI-related terms and concepts and clearly defined the scope, purpose, and intended/acceptable uses of its AI systems. This information is referenced and reflected consistently across organizational policies, standards, documentation, and training programs, although its implementation may not yet be fully mature or integrated. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations establish and maintain a glossary of AI-related terms and concepts and define the scope, purpose, and intended/acceptable uses of AI systems, and it is fully integrated into the organization's knowledge management and governance processes. This information is consistently applied, regularly reviewed and updated, and deeply embedded into all aspects of the organization's AI-related policies, standards, documentation, and training programs. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Inconsistent Use of AI Terms | AI terms are used inconsistently across the organization, its policies, and its documentation |

**Implementation Guidelines:**

Develop and maintain a glossary of AI-related terms and concepts to promote a common understanding across the organization. Include clear definitions of key terms and descriptions of the scope, purpose, and intended uses of the AI systems.

Regularly review and update the glossary to reflect changes in AI practices and the broader AI landscape. Ensure that definitions and concepts from the glossary are consistently referenced in organizational policies, standards, documentation, and training programs.

**Example Controls and Evidence:**

**Example Control 1:** Centralized AI Glossary and Taxonomy

**Description:** Maintain a centrally managed, version-controlled glossary and taxonomy of AI terms (e.g., model, training data, inference, bias, explainability) with definitions, usage guidance, and links to related policies and artifacts; review and update at regular intervals.

**Examples of Effective Evidence:**

• Published glossary document or intranet page with version history and approval stamps.

• Change log showing updates tied to reviews or regulatory/standards changes.

• Access/audit logs demonstrating who edited or approved glossary entries.

• Cross-reference matrix showing links from glossary terms to policies, playbooks, and system dossiers.

**Example Control 2:** AI System Use-Case and Scope Register

**Description:** Maintain a register that documents each AI system’s scope, purpose, intended/acceptable uses, boundaries, and prohibited uses, and require owners to update the register whenever capabilities or use cases change.

**Examples of Effective Evidence:**

• System register or catalog entries for sampled AI systems specifying scope, purpose, allowed and disallowed uses.

• Owner attestations or approvals confirming accuracy of system scope and use-case entries.

• Change logs showing updates to scope/use entries after system changes or governance reviews.

• Communications to stakeholders when system uses are added/modified (emails, release notes).

**Example Control 3:** Policy and Documentation Cross-Referencing Requirement

**Description:** Require that all AI-related policies, standards, SOPs, procurement templates, and training materials reference glossary definitions and the system use register to ensure consistent terminology and aligned intent across documents.

**Examples of Effective Evidence:**

• Policy and SOP documents with explicit references or hyperlinks to glossary terms and system use entries.

• Policy mapping matrix demonstrating consistent term usage across documents.

• Documentation review records showing checks for consistency with the glossary during updates.

• Approval records indicating reviewers confirmed alignment to glossary and scope register.

**Example Control 4:** Integration into SDLC Gates and Approval Workflows

**Description:** Embed checks into SDLC stage-gates and change control workflows that validate system scope, purpose, and acceptable uses against the register and require glossary-based term usage in design/specifications before approval.

**Examples of Effective Evidence:**

• Stage-gate checklist items and approval forms requiring confirmation of scope/use alignment and correct term usage.

• Ticketing/workflow logs showing gates blocked until scope and glossary checks are completed.

• Design/specification documents and product requirements document (PRD)s using standardized glossary terms with reviewer sign-offs.

• Audit samples demonstrating SDLC enforcement for multiple projects.

**Example Control 5:** Training, Communications, and Stakeholder Awareness Program

**Description:** Provide role-specific training and communications that teach the glossary terms, expected interpretations, and the organization’s definitions of acceptable AI uses. Include assessments and periodic refreshers to ensure ongoing consistency.

**Examples of Effective Evidence:**

• Training materials, completion records, and assessment results referencing glossary terms and acceptable uses.

• Awareness campaign artifacts (intranet posts, FAQs, quick-reference cards) linking to the glossary and system register.

• Periodic refresher attendance logs and quiz/assessment outcomes demonstrating retained understanding.

• Survey or audit results showing reduced terminology inconsistencies and improved policy adherence after training.

**GOVERN / Establishing Third-Party Risk Management for the AI System / Establishing Third-Party Risk Management Processes**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| GV-6.1.1 | Third-Party AI Evaluation and Selection | The organization establishes processes for evaluating and selecting third-party AI technologies based on criteria that assess security and privacy implications, integrating AI-specific considerations into third-party procurement planning, due diligence, evaluation, and contracting practices to create alignment with organizational risk management policies.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/govern/#govern-6-1) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations are beginning to identify third-party AI technologies for potential use but lack formal processes for their evaluation. Considerations for security and privacy are largely informal or project-specific, without established criteria or integration into broader organizational risk management policies. |
| **Stage 2: Minimal** | Yes | Organizations at the minimal stage establish basic processes for evaluating third-party AI technologies with initial focus on security and privacy considerations, creating a foundational approach to responsible AI procurement that can be expanded and refined over time. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations develop criteria that incorporate security and privacy considerations into third-party AI evaluation processes. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations have automated, standardized procedures that consistently assess AI vendors and technologies based on these criteria, ensuring alignment with risk management policies and promoting responsible AI adoption, supporting AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Insufficient Third-Party Evaluation | Processes for evaluating third-party AI technologies are insufficient or missing |

**Implementation Guidelines:**

Establish clear processes for evaluating and selecting third-party AI technologies to be integrated into the AI system. These processes should be based on a set of criteria that assess the security and privacy implications of each technology, ensuring alignment with the organization's overall risk management policies.

Develop a standardized evaluation framework that includes a set of questions or requirements that each potential third-party provider must address. Key areas to assess should include data handling practices, model training and testing procedures, and transparency and accountability measures.

Engage internal stakeholders, such as legal, security, and compliance teams, to ensure that the evaluation criteria are comprehensive and align with the organization's values and priorities. Establish a review and approval process for each third-party technology, including documentation of the evaluation results and any identified risks or concerns. Ensure that the selection process is transparent and that relevant stakeholders are involved in the decision-making process, promoting collaboration and informed choices.

**Example Controls and Evidence:**

**Example Control 1:** Development of a Third-Party AI Evaluation Framework

**Description:** Create a formal framework that defines criteria and processes for assessing third-party AI solutions on security and privacy, ensuring comprehensive risk-based evaluation.

**Examples of Effective Evidence:**

• Evaluation templates or checklists incorporating AI-specific security and privacy metrics.

• Approval workflows for third-party assessments aligned with risk policies.

• Detailed criteria outlining minimum standards and scoring mechanisms.

• Version-controlled evaluation protocols.

**Example Control 2:** Due Diligence and Vendor Assessment Processes

**Description:** Implement thorough due diligence activities, including risk assessments, code reviews, and compliance checks, tailored to AI-specific risks, during vendor evaluation and before procurement.

**Examples of Effective Evidence:**

• Risk assessment reports for each AI vendor or solution.

• Security and privacy compliance documentation review records.

• Technical assessments, including model transparency and bias evaluation.

• Due diligence checklists with AI-specific questions.

**Example Control 3:** Contracting and Risk Mitigation Clauses

**Description:** Incorporate contractual clauses addressing security and privacy, assessments, and liability, aligned with organizational policies and standards.

**Examples of Effective Evidence:**

• Contract templates with AI-specific risk mitigation provisions.

• Signed agreements explicitly covering security, data governance, and responsible use.

• Auditing and monitoring clauses for ongoing compliance.

• Documentation of contract reviews and approvals.

**Example Control 4:** Integration with Organizational Risk Policies and Oversight

**Description:** Ensure evaluation outcomes inform risk management decisions, with integration into the broader enterprise risk management and oversight frameworks.

**Examples of Effective Evidence:**

• Risk register entries linking third-party AI solutions with identified issues.

• Management approval records for deployment decisions based on evaluated risks.

• Reports showing alignment between evaluation results and organizational risk appetite.

• Oversight meeting minutes discussing third-party AI risks.

**Example Control 5:** Continuous Monitoring and Reassessment

**Description:** Establish ongoing monitoring processes for third-party AI solutions, including periodic reassessment of security and privacy, with updates embedded into contract terms.

**Examples of Effective Evidence:**

• Monitoring schedules and compliance audit reports.

• Post-deployment risk reassessment documentation.

• Incident or breach reports linked to third-party AI solutions.

• Updated evaluation criteria reflecting evolving standards and lessons learned.

**MAP / Understanding the Operating Context / Defining the Operating Context**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-1.1.1 | AI Purpose and Context Documentation | The organization identifies and documents the AI system’s purpose, context, beneficial uses, user types and expectations, relevant laws, norms, and prospective deployment settings.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-1-1) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations are beginning to identify and document basic information about the AI system's purpose, context, and intended uses. The process is likely informal and high-level at this stage. |
| **Stage 2: Minimal** | Yes | Organizations at the minimal stage have established a process for documenting key information about the AI system, but it may not be comprehensive or consistently applied across all systems. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations have a well-defined and consistently applied process for thoroughly identifying and documenting the AI system's purpose, context, uses, applicable laws/norms, and deployment settings. |
| **Stage 4: Embedded** | Yes | Organizations at the embedded stage maintain comprehensive, standardized documentation that is regularly updated, ensuring transparency and responsible AI deployment aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Undocumented AI System Purpose and Context | AI system purpose and context are undocumented or doesn't consider deployment settings |

**Implementation Guidelines:**

To establish a clear understanding of the AI system's context, document its intended purpose, beneficial uses, and prospective deployment settings. Engage relevant stakeholders such as business owners, end-users, and subject matter experts to gather input on the specific goals and objectives of the AI system, as well as its potential benefits.

Analyze the legal and regulatory landscape, and identify applicable laws, regulations, and/or industry standards. Document user types, expectations, and requirements, ensuring this information is communicated clearly and is accessible to relevant parties involved in the AI system's development and deployment.

Regularly update this documentation as needed to reflect any changes, maintaining its relevance and accuracy.

**Example Controls and Evidence:**

**Example Control 1:** AI System Purpose and Use-Case Statement

**Description:** Require a formal, system-level purpose statement for every AI project that defines intended benefits, primary/secondary use-cases, business objectives, and constraints; this statement must be approved by the system owner and product sponsor before development.

**Examples of Effective Evidence:**

• Approved system purpose/use-case document or PRD.

• Sign-off records from system owner, product sponsor, and relevant governance body.

• Version history showing updates when the system scope or objectives change.

• Linked registry entry in the AI catalog referencing the purpose statement.

**Example Control 2:** Contextual Deployment and Environment Profile

**Description:** Document the operational context and deployment settings for the AI system (e.g., intended users, geographic jurisdictions, technical environment, integration points, decision autonomy level) to assess constraints and risk exposure.

**Examples of Effective Evidence:**

• Deployment/environment profile forms detailing jurisdictions, integration architecture, user types, and autonomy level.

• System architecture diagrams and environment inventories tied to the profile.

• Records of environment-specific legal or regulatory requirements identified for each jurisdiction.

• Approval or review notes from operations/security teams confirming environment suitability.

**Example Control 3:** Stakeholder and User Expectation Analysis

**Description:** Conduct and document stakeholder mapping and user expectation analyses (including vulnerable or special-interest groups) to identify needs, likely interactions, transparency requirements, and potential harms; include documented mitigation strategies.

**Examples of Effective Evidence:**

• Stakeholder mapping documents and user personas with documented expectations and risk considerations.

• Notes or reports from stakeholder interviews, workshops, or surveys.

• Impact/risk matrices linking user expectations to mitigation strategies and owners.

• Records showing incorporation of findings into design requirements or mitigation plans.

**Example Control 4:** Legal, Normative and Regulatory Applicability Record

**Description:** Maintain a documented assessment that maps applicable laws, regulations, industry norms, contractual obligations, and guidelines to the AI system and its deployment contexts, with identified compliance actions and owners.

**Examples of Effective Evidence:**

• Legal/regulatory applicability checklist or mapping for the system listing statutes, standards, and contractual clauses.

• Assigned owner records for each identified legal/contractual obligation and planned compliance actions.

• Legal review memos or compliance sign-offs referencing the mapping.

• Updates to the mapping when new jurisdictions or deployment contexts are added.

**Example Control 5:** Acceptable and Prohibited Use Cases and Boundary Conditions Register

**Description:** Publish a register for each AI system specifying acceptable uses, prohibited uses, boundary conditions, and fail-safe behavior; require owners to validate that designs and controls enforce these boundaries prior to deployment.

**Examples of Effective Evidence:**

• System use register entries enumerating allowed/prohibited uses and boundary conditions with owner attestations.

• Test plans and results demonstrating enforcement of prohibited-use controls and fail-safe scenarios.

• Change-control records showing updates to acceptable uses following reviews or incidents.

• Communications and training materials sent to user groups describing acceptable uses and expectations.

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-1.1.2 | Impact and Stakeholder Impact Assessment | The organization considers the potential positive and negative impacts of the AI system in collaboration with a multidisciplinary set of relevant stakeholders (e.g., domain experts, human factors experts, affected communities) throughout its lifecycle.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-1-1) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations can prioritize inclusion of multidisciplinary perspectives (e.g., domain experts, human factors experts) in first internal and vendor meetings and in planning and design work. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations are starting to consider the AI system's potential impacts, but the assessment may be limited in scope and not consistently involve a multidisciplinary set of stakeholders. |
| **Stage 3: Evolving** | Yes | Organizations at the evolving stage conduct more thorough assessments of the AI system's potential impacts in collaboration with a multidisciplinary group of relevant stakeholders. The process is well-defined and consistently applied. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations incorporate ongoing, structured collaboration with a broad range of stakeholders into their standard processes, ensuring responsible and inclusive AI development aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Fair | Insufficient Stakeholder Impact Consideration | Stakeholder collaboration is insufficient in AI impact assessment |

**Implementation Guidelines:**

Engage a multidisciplinary set of relevant stakeholders (e.g., domain experts, human factors experts, and representatives from potentially affected communities) to inform and support ongoing impact considerations of the AI system. These considerations should focus on ensuring that the AI system works as intended and adheres to principles of fairness, transparency, accountability, and safety, particularly in regulated environments.

Use the impact consideration results to guide the design, development, deployment, and continual improvement of the AI system. This includes identifying any risks and implementing appropriate mitigation strategies. The process should be transparent, well-documented, and communicated to stakeholders, with clear criteria for consistency. Implement a follow-up mechanism after deployment to reassess impacts as the system evolves.

The primary goal is to utilize the AI system effectively to benefit customers, ensuring that the impact considerations support valid, fit-for-purpose solutions that align with regulatory requirements related to fair treatment and non-discrimination.

**Example Controls and Evidence:**

**Example Control 1:** Stakeholder Engagement Plan and Forum

**Description:** Establish a documented stakeholder engagement plan and recurring forum(s) that identify relevant internal and external stakeholders (domain experts, human factors, community representatives, civil society, regulators) and schedule consultations at key lifecycle stages to gather input on impacts and mitigation measures.

**Examples of Effective Evidence:**

• Stakeholder engagement plan and roster listing participants, roles, and engagement schedule.

• Agendas, minutes, and attendee lists from stakeholder forums, workshops, or consultations.

• Submitted stakeholder feedback logs and action-tracking entries showing responses and resolutions.

• Communications confirming outreach to affected communities (invitations, outreach materials).

**Example Control 2:** Inclusive Impact Assessment Process

**Description:** Require documented impact assessments (economic, privacy, safety) that include inputs from multidisciplinary stakeholders and consider disparate impacts; assessments must be reviewed and approved before deployment.

**Examples of Effective Evidence:**

• Completed impact assessment reports with sections documenting stakeholder inputs and identified impacts.

• Approval/sign-off records from a multidisciplinary review panel including external/community representatives.

• Mitigation plans derived from the assessment with assigned owners and status tracking.

• Evidence of follow-up consultations with stakeholders on proposed mitigations.

**Example Control 3:** Participatory Design and Co-creation Activities

**Description:** Integrate participatory design methods (co-creation workshops, user testing with representative populations, lived-experience advisory panels) into design and testing phases to surface real-world impacts, usability concerns, and contextual nuances.

**Examples of Effective Evidence:**

• Workshop artifacts (design prototypes, user-testing scripts), participant recruitment records, and consent forms.

• Usability and feedback reports summarizing findings from representative user groups and resulting design changes.

• Records of advisory panel recommendations and documented incorporation into design specifications.

• Photographs or recordings (with consent) and debrief notes from participatory sessions.

**Example Control 4:** Impact Monitoring with Community-Linked Feedback Channels

**Description:** Implement post-deployment monitoring that tracks impact indicators and provide accessible feedback channels for affected users to report harms or concerns; ensure reports trigger investigation and remediation workflows.

**Examples of Effective Evidence:**

• Monitoring dashboards capturing agreed impact indicators (complaints, equity metrics, misuse reports) and periodic impact reports.

• Public-facing feedback/contact mechanisms and logs of incoming reports from users.

• Investigation records, incident tickets, and remediation action logs tied to reported harms.

• Periodic reports to governance bodies summarizing feedback, trends, and remediation status.

**Example Control 5:** Community Benefit and Harm Mitigation Commitments

**Description:** Define and publish commitments to maximize community benefits and mitigate harms (e.g., compensation, access improvements, transparency measures), integrate these into project plans and contracts, and track fulfillment through KPIs.

**Examples of Effective Evidence:**

• Documented community benefit/harm mitigation plans with KPIs, owners, and timelines included in project plans.

• Contractual clauses or MOUs with community partners outlining commitments and accountability.

• KPI reports demonstrating progress on benefit delivery and harm reduction (service access metrics, reduced complaint rates).

• Evidence of completed commitments (e.g., improved access features deployed, published transparency reports).

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-1.1.3 | Assumptions, Risks, and TEVV Documentation | The organization identifies and documents assumptions, limitations, and potential risks throughout the AI system’s lifecycle. Testing, evaluation, verification, and validation (TEVV) requirements and system metrics to identify, monitor, and manage associated system risks are identified.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-1-1) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations acknowledge that all AI systems have assumptions and limitations, and factor that knowledge into preparations for AI adoption. |
| **Stage 2: Minimal** | Yes | Organizations at the minimal stage have started identifying some assumptions, limitations, and risks, but the process is not comprehensive or consistently applied throughout the AI lifecycle. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations have established processes for systematically identifying and documenting assumptions, limitations, and risks, including those related to TEVV and metrics, at key points in the AI lifecycle. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations have integrated, automated processes for continuous assessment, verification, and validation, ensuring comprehensive risk management aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Undocumented Limitations and Risks | AI assumptions, limitations, risks, and testing requirements are undocumented |

**Implementation Guidelines:**

Identify and document assumptions, limitations, and potential risks associated with the AI system throughout its lifecycle. This includes assumptions about data quality, model performance, and user behavior, as well as limitations in the AI system's capabilities or context applicability.

Conduct a risk assessment to identify potential risks during development, testing, deployment, and use of the AI system, including testing, evaluation, verification, and validation (TEVV) and system metrics. Develop and implement risk mitigation strategies for each identified risk and monitor their effectiveness throughout the AI system's lifecycle.

Regularly review and update the assumptions, limitations, and risks in collaboration with relevant stakeholders as the AI system evolves and new information becomes available.

**Example Controls and Evidence:**

**Example Control 1:** Assumptions and Limitations Register

**Description:** Maintain a living register for each AI system that records documented assumptions, known limitations, boundary conditions, and accepted residual risks, reviewed and updated at key lifecycle milestones.

**Examples of Effective Evidence:**

• Assumptions & limitations register entries tied to specific system versions with timestamps and owner names.

• Change logs showing updates to assumptions/limitations after design reviews, tests, or incidents.

• Review/sign-off records from system owners, product managers, and risk/legal reviewers.

• Traceability matrix linking assumptions/limitations to design documents, requirements, and test cases.

**Example Control 2:** TEVV Plan and Requirements Document

**Description:** Produce a formal Testing, Evaluation, Verification, and Validation (TEVV) plan per system that specifies TEVV objectives, scope, test methods, acceptance criteria, datasets, metrics, test schedules, and roles/responsibilities.

**Examples of Effective Evidence:**

• Approved TEVV plan documents with versioning and sign-offs from QA, ML engineering, privacy, and risk owners.

• Test matrices mapping requirements/assumptions to specific TEVV activities and expected outcomes.

• Scheduled TEVV timelines and resource allocations recorded in project plans.

• Documented acceptance criteria that gate progression to production.

**Example Control 3:** Metric Definition and Monitoring Framework

**Description:** Define and instrument system-specific metrics (performance, calibration, fairness, robustness, drift, privacy leakage, reliability) with thresholds, alerting rules, and dashboards to support continuous risk detection and TEVV monitoring.

**Examples of Effective Evidence:**

• Metric catalogue listing defined metrics, formulas, thresholds, and responsible owners.

• Monitoring dashboards and automated alert logs showing metric values, threshold breaches, and timestamps.

• Configuration records for alerting rules and escalation procedures.

• Sample reports showing trend analysis and metric-driven decisions (retrain/rollback).

**Example Control 4:** TEVV Execution Records and Validation Artifacts

**Description:** Execute TEVV activities per the TEVV plan and retain validation artifacts—test datasets, scripts, execution logs, results, defect/issue tickets, and remediation evidence—that demonstrate how tests mapped to risks and acceptance criteria.

**Examples of Effective Evidence:**

• Test execution reports with inputs, outputs, pass/fail results, and links to datasets and scripts.

• Issue/defect trackers documenting findings, severity, assigned owners, remediation actions, and closure evidence.

• Signed validation certificates or deployment gate approvals referencing completed TEVV artifacts.

• Reproducibility artifacts (notebooks, environment manifests, random seeds) showing tests can be rerun.

**Example Control 5:** Risk-Based Periodic Reassessment and TEVV Feedback Loop

**Description:** Implement a periodic reassessment process that uses TEVV results, monitoring alerts, incident analyses, and stakeholder feedback to update risk ratings, assumptions, TEVV scopes, and mitigation plans; document lessons learned and policy/process changes.

**Examples of Effective Evidence:**

• Periodic reassessment reports or risk register updates showing changes to risk ratings and updated mitigation plans.

• Lessons-learned or post-mortem documents linking TEVV findings to process or design changes.

• Updated TEVV plans or test cases created in response to monitoring/incident findings with approval records.

• Governance meeting minutes showing review of reassessments, resource decisions, and closure of remediation items.

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-1.1.4 | Non-AI Alternatives Evaluation | The organization considers non-AI alternatives and documents the comparative analysis and decision rationale.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-1-1) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations acknowledge non-AI solutions may be a better fit for certain applications, and factor that knowledge into preparations for AI adoption. |
| **Stage 2: Minimal** | Yes | Organizations at the minimal stage establish more formal documentation of alternatives and decision rationale to support transparency. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations expand the analysis process to include comprehensive evaluations and systematic comparisons as AI projects become more strategic. |
| **Stage 4: Embedded** | Yes | Organizations at the embedded stage routinely integrate non-AI alternatives into their decision workflows, with thorough documentation and consistent review, ensuring responsible and transparent AI deployment aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Valid & Reliable | Alternatives Not Considered | Non-AI alternatives are not considered or documented |

**Implementation Guidelines:**

Evaluate non-AI alternatives to the proposed AI system to determine if they provide more valid outcomes, considering options such as traditional rule-based systems, human decision-making processes, or other non-AI technologies.

Conduct a comparative review of the AI system and its alternatives, using established criteria to assess factors such as accuracy, fairness, transparency, and accountability.

Document the results of this review, including the rationale for selecting the AI system, and communicate this information to relevant stakeholders. Regularly update the comparative review in collaboration with stakeholders as new information becomes available or the AI system evolves.

**Example Controls and Evidence:**

**Example Control 1:** Alternatives Analysis and Decision Rationale Document

**Description:** Require formal documentation of non-AI and AI alternatives during project initiation, including criteria such as cost, complexity, accuracy, explainability, and risk, with explicit comparison and decision justifications.

**Examples of Effective Evidence:**

• Completed alternatives analysis reports comparing AI and non-AI options with documented criteria and scoring.

• Decision rationale memos signed off by project sponsors, risk owners, and relevant stakeholders outlining selected options and reasoning.

• Supporting evidence such as cost estimates, performance benchmarks, and stakeholder input inform the analysis.

• Version-controlled documents including updates when project scope or options change.

**Example Control 2:** Alternative Options Review Workflow

**Description:** Implement a structured review process involving cross-functional stakeholders (e.g., technical, legal, business) to evaluate and approve or reject the use of AI versus non-AI options, with documented approval records.

**Examples of Effective Evidence:**

• Review and approval checklists with documented deliberations during project initiation stages.

• Meeting minutes and sign-offs from stakeholders endorsing the chosen approach.

• Record of dissenting opinions or alternative considerations that influenced the decision.

• Action logs tracking follow-up on alternative evaluations or re-assessments when project scope evolves.

**Example Control 3:** Documented Criteria for AI Adoption

**Description:** Define and document explicit criteria for considering AI, including use-case appropriateness, transparency, data availability, risk levels, and alignment with organizational values; review criteria during assessments.

**Examples of Effective Evidence:**

• Criteria checklist incorporated into project charters or proposals.

• Policy or standard referencing criteria for AI vs. non-AI decision-making.

• Training or briefing materials emphasizing the importance of comparing alternatives.

• Records of review sessions verifying criteria application in project assessments.

**Example Control 4:** Comparative Evaluation Framework

**Description:** Use a standardized framework or scoring model to evaluate alternatives systematically, capturing quantitative and qualitative factors, and documenting the resulting scores and rationale for AI adoption.

**Examples of Effective Evidence:**

• Completed evaluation matrices with scores, weights, and justifications for each criterion.

• Stakeholder review reports analyzing comparisons and highlighting key decision drivers.

• Summaries or dashboards presenting the comparative analysis outcomes.

• Records of revisions or sensitivity analyses demonstrating robustness of the choice.

**Example Control 5:** Post-Decision Review and Justification Maintenance

**Description:** Conduct periodic reviews of past decisions comparing actual outcomes with the initial analysis, updating justification documentation to reflect lessons learned, and ensuring transparency and accountability.

**Examples of Effective Evidence:**

• Review reports analyzing decision outcomes versus initial rationale.

• Updated decision records capturing lessons learned, arising risks, or changes in organizational priorities.

• Stakeholder feedback and audit logs confirming review effectiveness.

• Website or internal repository providing accessible history of AI vs. alternative decisions for accountability.

**MAP / Understanding the Operating Context / Bolstering Cross-Functional Perspectives on the Operating Context**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-1.2.1 | Multidisciplinary AI Impact Teams | The organization recruits, supports, and facilitates documented collaboration among multidisciplinary teams throughout the AI system lifecycle (from problem formulation to monitoring). This enhances comprehensive risk analysis, bias mitigation, and thorough consideration of potential impacts.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-1-2) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | Organizations at the initial stage begin to recognize the importance of multidisciplinary teams and make initial efforts to recruit multidisciplinary members. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations recognize the importance of different perspectives and begin forming multidisciplinary teams, although team composition may not yet reflect the full range of necessary expertise. |
| **Stage 3: Evolving** | Yes | Organizations at the evolving stage have well-defined processes for establishing multidisciplinary teams with a wide range of demographics, disciplines, experiences, and expertise to thoroughly inform AI system context and impacts. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations systematically integrate multidisciplinary teams into ongoing AI risk analysis processes, with structured support and continuous improvement, ensuring responsible AI aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Fair | Ineffective AI Teams or Collaboration | Teams are not multidisciplinary, are not effectively supported, or collaboration across disciplines is undocumented or inconsistently practiced |

**Implementation Guidelines:**

Form and support multidisciplinary teams with different demographics, disciplines, and expertise (e.g., engineering, legal, external stakeholders) to ensure broad perspectives that reduce risk blind spots and thoroughly consider biases and impacts. Provide these teams with necessary resources and foster effective, documented collaboration throughout the AI system lifecycle, from problem formulation to monitoring.

Prioritize establishing clear communication channels and collaboration platforms, encouraging open dialogue and psychological safety. Ensure that insights and recommendations from these collaborative efforts are carefully documented and incorporated into the AI system's design, development, and deployment. Regularly review team composition to address gaps in expertise or perspectives, and continuously monitor the effectiveness of multidisciplinary collaboration, making adjustments as needed.

**Example Controls and Evidence:**

**Example Control 1:** Policy for Multidisciplinary AI Teams

**Description:** Establish and communicate a formal policy emphasizing recruitment of multidisciplinary team members across demographics, disciplines, and expertise areas involved in AI projects, with clear targets and accountability measures.

**Examples of Effective Evidence:**

• Policy documentation with specific goals for multidisciplinary team composition.

• Recruitment strategies and outreach programs targeting multidisciplinary groups and varied disciplines.

• Workforce reports and hiring records demonstrating progress toward multidisciplinary teams.

• Performance reviews or accountability reports linking objectives to team development.

**Example Control 2:** Multidisciplinary Team Formation and Roles Documentation

**Description:** Mandate the formation of multidisciplinary teams for AI initiatives, and maintain documentation outlining team member roles, expertise areas, and contributions to risk and impact assessments at each lifecycle stage.

**Examples of Effective Evidence:**

• Team composition records, including bios highlighting differing backgrounds, demographics, stakeholder perspectives.

• Roles and responsibilities matrices documenting expertise areas aligned to risk and bias analysis activities.

• Meeting minutes and collaboration logs showing multidisciplinary team participation in reviews and risk assessments.

• Team formation approval records tied to project initiation documentation.

**Example Control 3:** Support Mechanisms for Multidisciplinary Teams

**Description:** Provide ongoing support, training, and resources (e.g., bias awareness training, competency, stakeholder engagement workshops) to ensure team members can effectively identify and address biases and impacts.

**Examples of Effective Evidence:**

• Training completion certificates and participation logs for bias-awareness programs.

• Resource repositories (guidelines, toolkits) accessible to teams for cross-functional design and impact analysis.

• Feedback surveys indicating increased awareness and application of broad concepts.

• Event records of workshops or forums fostering multidisciplinary dialogue and understanding.

**Example Control 4:** Bias and Impact Assessment Processes with multidisciplinary Inputs

**Description:** Integrate inputs from a wide range of experts into risk and impact assessment processes, with documented analyses and mitigation strategies.

**Examples of Effective Evidence:**

• Impact assessment reports explicitly referencing contributions from multidisciplinary stakeholders and experts.

• Stakeholder engagement records, including interview transcripts, focus group summaries, and consultation logs.

• Sign-off records from multidisciplinary reviewers endorsing bias mitigation and impact considerations.

• Documentation of actions taken to incorporate multidisciplinary perspectives into design or policy updates.

**Example Control 5:** Continuous Monitoring and Feedback on Impact Effectiveness

**Description:** Monitor the effectiveness of cross-functional efforts and impact analyses through feedback channels, regular reviews, and metrics, and adjust team composition or processes accordingly.

**Examples of Effective Evidence:**

• Metrics dashboards tracking stakeholder participation and bias mitigation effectiveness.

• Review meeting minutes and improvement plans addressing identified gaps or biases uncovered during project lifecycle.

• Stakeholder feedback and survey results regarding consideration of impacts.

• Records of process enhancements implemented in response to impact review findings.

**MAP / Understanding the Operating Context / Defining the Mission and Goals**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-1.3.1 | AI Alignment with Mission and Values | The organization identifies and documents the AI system’s alignment and contribution to organizational mission, values, goals (including relevant goals for AI technology), and AI Trustworthy Principles.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-1-3) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations are starting to consider how AI systems align with and contribute to their overall mission and goals, but the process is likely informal and high-level. |
| **Stage 2: Minimal** | Yes | Organizations at the minimal stage have begun documenting the AI system's alignment with mission, values, goals, and principles, but the process may not be comprehensive or consistently applied. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations have established clear processes for thoroughly assessing and documenting how each AI system aligns with and contributes to the organizational mission, values, goals, and AI Trustworthy Principles. |
| **Stage 4: Embedded** | Yes | Organizations at the embedded stage systematically embed this alignment into their lifecycle management, with automated tracking and continuous updates ensuring responsible AI development aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | AI System Misalignment | AI system alignment with mission, values, and goals is undocumented or lacking |

**Implementation Guidelines:**

To determine that the AI system aligns with and contributes to the organization's mission, values, goals, and AI Trustworthy Principles, clearly document these elements. Engage with senior leadership and relevant stakeholders to gain a thorough understanding of the organization's overall direction and priorities.

Review existing documentation, such as mission statements, strategic plans, and value statements to identify key themes and objectives. Conduct workshops or interviews with individuals across the organization to gather additional perspectives on how the AI system can support and advance the organization's mission and goals.

Document the specific ways in which the AI system aligns with and contributes to these elements, ensuring this documentation is easily accessible to relevant parties involved in the AI system's development and deployment. Regularly review and update this documentation as the organization's mission and goals evolve over time to maintain its relevance and accuracy.

**Example Controls and Evidence:**

**Example Control 1:** Establishment of Documentation Standards for AI System Alignment

**Description:** Develop standards and templates for documenting how AI systems align with the organization’s mission, values, strategic goals, and AI Trustworthy Principles.

**Examples of Effective Evidence:**

• Standardized templates capturing alignment with mission, values, and principles.

• Guideline documents outlining documentation requirements.

• Training materials for teams responsible for alignment documentation.

• Version-controlled repository of documented AI system alignments.

**Example Control 2:** Systematic Identification and Documentation of Alignment

**Description:** Conduct assessments to determine how each AI system supports or enhances organizational objectives, values, and AI Trustworthy Principles, then document these findings.

**Examples of Effective Evidence:**

• Alignment reports correlating AI features with organizational goals and values.

• Records of stakeholder consultations assessing alignment.

• Evidence of review and approval of alignment documentation.

• Change logs when systems’ alignment status or contribution is updated.

**Example Control 3:** Integration of AI Contribution into Strategic Planning

**Description:** Incorporate AI system alignment and contribution information into strategic planning, oversight, and decision-making processes.

**Examples of Effective Evidence:**

• Strategic planning documents referencing AI contribution assessments.

• Minutes from strategic review meetings discussing AI alignment.

• Reports linking AI system performance or impact to organizational goals.

• Dashboards illustrating AI contribution to key metrics aligned with AI Trustworthy Principles.

**Example Control 4:** Ongoing Monitoring and Updates of Alignment Documentation

**Description:** Regularly review and update documentation to reflect changes in AI systems or organizational priorities, ensuring continued alignment.

**Examples of Effective Evidence:**

• Periodic review schedules and logs.

• Updated documentation reflecting recent system modifications or strategic shifts.

• Feedback or audit reports identifying gaps or misalignments.

• Records of re-assessment after major system changes or organizational initiatives.

**Example Control 5:** Transparency and Communication of AI System Alignment

**Description:** Communicate AI system alignment and contribution to relevant stakeholders, reinforcing transparency and trustworthiness.

**Examples of Effective Evidence:**

• Internal reports or dashboards sharing alignment status with leadership.

• Communication plans or newsletters explaining AI contribution to organizational goals.

• Stakeholder feedback on transparency and understanding of AI role.

• Training sessions or workshops on organizational principles and AI alignment.

**MAP / Understanding the Operating Context / Defining Real-World Business Value**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-1.4.1 | AI Business Value and Context Communication | The organization clearly defines, documents, and communicates the business value, objectives, and intended context of use for each AI system to relevant stakeholders, promoting alignment with organizational goals and facilitating transparency and accountability in how the AI system supports operational success.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-1-4) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations begin defining and documenting the basic purpose, value, and context of each AI system. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations have defined and documented the business value and context of use for AI systems, but the information may not be comprehensive or consistently communicated to all relevant stakeholders. |
| **Stage 3: Evolving** | Yes | Organizations at the evolving stage have established clear processes for thoroughly defining, documenting, and communicating the business value and context of use for each AI system to promote transparency and accountability. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations have automated, comprehensive processes that continuously inform stakeholders, ensuring responsible and transparent AI deployment aligned with AI Trustworthy Principles |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Undefined Business Value | Business value, objectives, and context of AI systems are not clearly defined |

**Implementation Guidelines:**

To confirm that the business value, objectives, and context of use for each AI system are clearly defined, documented, and communicated, engage with relevant stakeholders—such as business owners, end-users, subject matter experts, and other interested parties—to gather multidisciplinary input and perspectives.

Conduct workshops or interviews to identify how the AI system is expected to deliver tangible business outcomes, such as increasing efficiency, reducing costs, or enhancing customer satisfaction. Clearly articulate these value propositions, ensuring they align with and support the organization’s overall goals and strategies.

Additionally, document the specific context of use, including the intended users, operational scenarios, constraints, and key assumptions. Make this documentation readily accessible to all relevant parties and communicate it through multiple channels, such as training sessions, user manuals, and internal communications. Regularly review and update this documentation to ensure ongoing accuracy and relevance, supporting transparency and alignment with organizational objectives.

**Example Controls and Evidence:**

**Example Control 1:** Development of Clear Documentation for AI Business Value, Objectives, and Context

**Description:** Establish standardized processes and templates to define and record the intended business value, objectives, and context of use for each AI system.

**Examples of Effective Evidence:**

• Standardized documentation templates capturing purpose, value, and use cases.

• Approved records of initial and updated documentation reflecting current understanding.

• Training materials for teams responsible for documenting AI systems.

• Centralized repository of documented purpose and context descriptions.

**Example Control 2:** Communication of AI Purpose and Goals to Stakeholders

**Description:** Systematically disseminate AI system purpose, business value, objectives, and context of use to all relevant stakeholders through meetings, reports, or dashboards to ensure shared understanding.

**Examples of Effective Evidence:**

• Stakeholder presentation slides or reports summarizing AI purpose and objectives.

• Internal newsletters or communication channels sharing updates.

• Access logs or views of dashboards displaying AI system descriptions.

• Meeting minutes documenting stakeholder briefings and feedback.

**Example Control 3:** Alignment with Organizational Goals and Transparency

**Description:** Ensure that documentation and communication explicitly link AI system objectives to broader organizational goals, promoting transparency and accountability.

**Examples of Effective Evidence:**

• Documentation linking AI objectives to strategic goals in reports or portals.

• Maps or matrices showing alignment with organizational priorities.

• Audit trail of updates reflecting ongoing alignment efforts.

• Stakeholder acknowledgment or sign-off records indicating understanding and agreement.

**Example Control 4:** Regular Review and Updating of Documentation and Communication

**Description:** Periodically review and update the documented purpose, goals, and context, and re-communicate changes to ensure ongoing alignment and clarity.

**Examples of Effective Evidence:**

• Review schedules and change logs for purpose and objectives documents.

• Updated communication materials or stakeholder briefings.

• Feedback records indicating clarity or gaps in understanding.

• Revision records reflecting adjustments tied to organizational changes.

**Example Control 5:** Facilitating Transparency and Accountability through Accessible Information

**Description:** Maintain accessible, transparent records of AI system purposes and objectives to foster accountability, support audits, and enable informed decision-making across the organization.

**Examples of Effective Evidence:**

• Centralized, accessible documentation repositories or portals.

• Audit logs showing stakeholder access and queries.

• Reports demonstrating use of purpose and objectives for oversight and decision-making.

• Evidence of training or awareness sessions emphasizing transparency.

**MAP / Understanding the Operating Context / Defining and Communicating User Requirements**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-1.6.1 | AI Requirements from Stakeholders | The organization engages relevant AI stakeholders (e.g., end-users, domain experts, impacted communities) to elicit and document system requirements that address AI risks and promote trustworthy characteristics (e.g., security, fairness, transparency, accountability).  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-1-6) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations recognize the importance of engaging AI stakeholders and start the process of eliciting initial requirements. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations have begun engaging some relevant AI stakeholders to gather system requirements, but the process may not comprehensively address AI risks and trustworthiness or be consistently documented. |
| **Stage 3: Evolving** | Yes | Organizations at the evolving stage have established clear processes for engaging a wide range of relevant AI stakeholders to thoroughly elicit and document requirements focused on mitigating AI risks and promoting trustworthy characteristics. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations embed continuous, structured engagement with multidisciplinary AI stakeholders into their workflows, ensuring ongoing, stakeholder-informed development aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Unelicited AI Stakeholder Requirements | AI stakeholder system requirements that address AI risks aren't elicited or documented |

**Implementation Guidelines:**

To effectively elicit and document system requirements that address AI risks and promote trustworthy characteristics, engage a multidisciplinary range of relevant AI stakeholders, including end-users, domain experts, and impacted communities. Use a variety of methods to gather input, such as interviews, focus groups, surveys, and workshops, ensuring that participants can provide feedback and raise concerns.

Develop a standardized framework for categorizing and prioritizing system requirements based on their potential impact on AI risks and trustworthy characteristics, such as security, fairness, transparency, and accountability. Document the system requirements using clear and concise language, ensuring that this information is easily accessible to relevant parties.

Regularly review and update the system requirements to maintain their accuracy and relevance, and communicate changes to relevant stakeholders to ensure alignment.

**Example Controls and Evidence:**

**Example Control 1:** Stakeholder Engagement Planning for AI Requirements Gathering

**Description:** Develop structured plans to identify and involve relevant AI Stakeholders, ensuring their input is integrated into system requirements development.

**Examples of Effective Evidence:**

• Stakeholder analysis documents identifying relevant parties.

• Engagement plans and timelines outlining methods for elicitation.

• Records of outreach and communication to involved stakeholders.

• Feedback collection and response documentation.

**Example Control 2:** Structured Elicitation and Documentation of Requirements

**Description:** Use standardized methods (e.g., interviews, workshops, surveys) to elicit system requirements from AI Stakeholders, focusing on risks and trustworthiness aspects, and systematically document these inputs.

**Examples of Effective Evidence:**

• Records of elicitation sessions, including agendas and participant lists.

• Requirement specification documents capturing stakeholder needs and concerns.

• Follow-up action logs for clarifications or additional input.

• Version-controlled repository of requirement documents.

**Example Control 3:** Alignment of Requirements with AI Trustworthy Principles

**Description:** Ensure that elicited requirements explicitly incorporate considerations related to security, fairness, transparency, and accountability, aligning system design with trustworthiness goals.

**Examples of Effective Evidence:**

• Requirement checklists linked to AI Trustworthy Principles.

• Review and approval records demonstrating requirement alignment.

• Requirement traceability matrices mapping stakeholder inputs to trustworthiness criteria.

• Evaluation reports validating the coverage of trust factors.

**Example Control 4:** Validation and Approval of Requirements

**Description:** Obtain validation and formal approval from relevant stakeholders, including AI Stakeholders, to confirm that documented requirements adequately address risks and trustworthiness concerns.

**Examples of Effective Evidence:**

• Signed requirement approval signatures or electronic endorsements.

• Validation workshop or review meeting minutes.

• Feedback or objection logs from stakeholders.

• Records of requirement revisions following stakeholder input.

**Example Control 5:** Integration of Requirements into System Design and Development

**Description:** Incorporate elicited requirements into the design, development, and testing phases of AI systems to ensure stakeholder concerns about risks and trustworthy characteristics are systematically addressed.

**Examples of Effective Evidence:**

• Requirement traceability matrices linked to design specifications.

• Design review records referencing stakeholder requirements.

• Testing plans and reports confirming requirements fulfillment.

• Change management logs demonstrating requirement-based modifications.

**MAP / Understanding the AI System / Defining AI System Task and Methodology**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-2.1.1 | AI Learning and Decision Tasks | The organization defines and documents the AI system's existing and potential future learning and decision-making tasks (e.g., classification, generation, recommendation, prediction), considering fit-for-purpose, scalability and adaptability.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-2-1) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | Organizations at the initial stage begin identifying basic learning and decision-making tasks of AI systems. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations have begun defining and documenting the AI system's current learning and decision-making tasks, but may not yet consider future potential tasks or scalability and adaptability factors. |
| **Stage 3: Evolving** | Yes | Organizations at the evolving stage have established clear processes for comprehensively defining and documenting both current and potential future AI system tasks, taking into account scalability and adaptability considerations. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations systematically integrate detailed task documentation into lifecycle management, ensuring responsible AI development aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Valid & Reliable | Undefined Learning Tasks | AI system learning and decision-making tasks are undefined or undocumented |

**Implementation Guidelines:**

To effectively define and document the AI system's existing and potential future learning and decision-making tasks, engage relevant stakeholders such as data scientists, domain experts, and business leaders to gather input on the system's intended use and capabilities.

Consider the specific types of tasks the AI system is expected to perform, such as classification, generation, recommendation, or prediction, and document each task using clear and concise language. Assess the validity, scalability and adaptability of each task, considering factors such as the volume and complexity of data involved, the computational resources required, and the potential for the task to evolve or expand over time.

Document the scalability and adaptability considerations for each task, ensuring this documentation is easily accessible to relevant parties. Store detailed information about specific tasks and their attributes in the model documentation within the AI inventory (see GOVERN 1.6), to avoid inventory overload and to support traceability.

Regularly review and update the task documentation as needed to maintain its accuracy and relevance and communicate any changes to relevant stakeholders.

**Example Controls and Evidence:**

**Example Control 1:** Establishment of a Documentation Framework for AI Tasks

**Description:** Develop a structured approach and templates for documenting all AI system tasks, specifying scope, objectives, and technical descriptions, both current and future plans.

**Examples of Effective Evidence:**

• Standardized templates capturing task type, objectives, inputs, outputs, and performance criteria.

• Policy documents outlining documentation standards for AI tasks.

• Versioned records of task descriptions aligned with system development stages.

• Guidelines for identifying potential future tasks and their evaluation criteria.

**Example Control 2:** Identification and Assessment of Current Tasks

**Description:** Inventory and evaluate existing AI tasks, assessing their purpose, fit-for-purpose, and suitability for current operational needs.

**Examples of Effective Evidence:**

• Task inventories with detailed descriptions and performance metrics.

• Requirement and design documents corresponding to existing tasks.

• Assessment reports confirming task relevance and fitness for current use.

• Stakeholder review notes endorsing task deployment.

**Example Control 3:** Planning for Future Tasks and Scalability

**Description:** Outline plans for potential new tasks, scalability, and adaptability, including technical feasibility and strategic alignment considerations.

**Examples of Effective Evidence:**

• Roadmaps and project proposals for future AI tasks.

• Technical feasibility studies and impact assessments.

• Scalability and adaptability criteria integrated into planning and design documents.

• Evaluation matrices for prioritizing future task development.

**Example Control 4:** Integration of Task Documentation into Development and Governance Processes

**Description:** Embed task descriptions and assessments into development workflows, risk management practices, and oversight activities to ensure alignment and traceability.

**Examples of Effective Evidence:**

• Task documentation linked to design artifacts and requirements.

• Risk assessments considering new or evolving tasks.

• Traceability matrices connecting tasks to organizational goals and compliance standards.

• Periodic review records updating task descriptions.

**Example Control 5:** Continuous Review and Refinement Based on Feedback and Changes

**Description:** Regularly review and update task definitions and assessments in response to feedback, system evolution, and changing organizational strategies.

**Examples of Effective Evidence:**

• Scheduled review schedules with documented updates.

• Feedback logs from users, developers, and stakeholders influencing task descriptions.

• Change management records reflecting task refinement.

• Lessons learned from deployment and maintenance activities incorporated into future planning.

**MAP / Understanding the AI System / Defining AI System Utility and Oversight Needs**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-2.2.1 | AI Usage and Boundary Documentation | The organization documents the settings, environments, and conditions that are within the AI system’s intended use, defining the boundaries of its applicability while considering knowledge limits and areas where human intervention and irreversibility of decisions is crucial.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-2-2) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations begin identifying basic settings and boundaries of AI intended use. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations have begun documenting the intended use settings and conditions for AI systems, but the boundaries and knowledge limits may not be comprehensively defined and areas for human intervention may not be clearly identified. |
| **Stage 3: Evolving** | Yes | Organizations at the evolving stage have established clear processes for thoroughly documenting AI system intended use boundaries, knowledge limits, and areas requiring human intervention, providing a comprehensive picture of the system's applicability and limitations. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations integrate comprehensive boundary documentation into lifecycle management, ensuring responsible and transparent AI deployment aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Explainable & Interpretable | Undocumented Use Boundaries | AI system's intended use boundaries, environments, and conditions are undocumented |

**Implementation Guidelines:**

To effectively document the boundaries, environments, and conditions within the AI system's intended use, engage relevant stakeholders, such as data scientists, domain experts, and end-users, to gather input on the system's capabilities and limitations. Develop a clear and concise definition of the AI system's intended use, including the specific tasks it is designed to perform, the types of data it can process, and the expected outputs or outcomes.

Identify and document any known limitations or constraints on the system's applicability, such as data quality requirements, computational resource limitations, or domain-specific assumptions. Clearly define the boundaries of the AI system's knowledge and capabilities, specifying areas where human intervention or oversight are necessary to ensure safe and effective operation.

Document this information in clear and concise language, making it easily accessible to relevant parties. Regularly review and update the documentation as needed to maintain its accuracy and relevance and communicate changes to relevant stakeholders.

**Example Controls and Evidence:**

**Example Control 1:** Development of a Comprehensive Use Case Document

**Description:** Create a detailed record specifying the operational settings, deployment environments, and conditions under which the AI system is intended to operate, including constraints and known knowledge limits.

**Examples of Effective Evidence:**

• Formal use case documents outlining environment parameters and operational boundaries.

• Specifications detailing system assumptions, scope, and knowledge limitations.

• Diagrams illustrating deployment contexts and boundary conditions.

• Stakeholder review and approval records.

**Example Control 2:** Definition of Boundaries and Applicability Criteria

**Description:** Establish explicit criteria delineating where and how the AI system is appropriate, including environmental, data, and user-related conditions, as well as areas requiring human oversight.

**Examples of Effective Evidence:**

• Use boundary matrices listing allowed contexts, data types, and user roles.

• Operational manuals describing environmental parameters and limits.

• Risk assessments identifying areas where human intervention is mandatory.

• Policy documents on system scope restrictions.

**Example Control 3:** Identification of Knowledge Limits and Uncertainty Areas

**Description:** Document areas where the AI system’s knowledge is limited or unreliable, providing guidance for deployment and human oversight needs.

**Examples of Effective Evidence:**

• Knowledge limit assessments noting data gaps, model uncertainties, and performance thresholds.

• Decision matrices mapping knowledge limits to acceptable use conditions.

• Operational reports highlighting areas where human reviews are essential.

• Risk registers with entries on knowledge-based uncertainties.

**Example Control 4:** Specification of Human Intervention Points and Responsibilities

**Description:** Clearly specify situations where human intervention is required, including roles, decision points, and escalation procedures.

**Examples of Effective Evidence:**

• Operator guidelines delineating intervention points.

• Escalation workflows for system alerts or knowledge uncertainties.

• Training and onboarding materials emphasizing intervention responsibilities.

• Incident logs documenting human alerts and responses.

**Example Control 5:** Regular Review and Utilization of Use and Boundary Documentation

**Description:** Schedule periodic reviews to confirm that use boundaries, knowledge limits, and human intervention protocols remain accurate and appropriate, updating them based on operational feedback and system evolution.

**Examples of Effective Evidence:**

• Review schedules with documented outcomes and updates.

• Feedback logs from deployment and operational teams.

• Updated use case and boundary documents with revision history.

• Lessons learned from incidents or operational changes incorporating boundary adjustments.

**MAP / Understanding the AI System / Applying Scientific Methodologies**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MP-2.3.1 | Development and Operation Documentation | The organization documents assumptions, limitations, techniques, and metrics used for development or operation of the AI system throughout the lifecycle, aligning with data governance policies. This includes documenting data selection, curation, preparation, and analysis techniques, as well as identifying modeled constructs and methods for inferring reasonable relationships between constructs and dataset attributes.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/map/#map-2-3) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | Organizations at the initial stage begin recognizing the importance of documenting key assumptions, limitations, and techniques used in AI development and operation. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations have begun establishing TEVV protocols for AI models and systems, but they may not cover the full lifecycle or all relevant subcomponents and may not be consistently documented. |
| **Stage 3: Evolving** | Yes | Organizations at the evolving stage have established comprehensive, well-documented TEVV protocols that cover the entire AI lifecycle and all relevant models, systems, and subcomponents, ensuring thorough testing and validation. |
| **Stage 4: Embedded** | Yes | At the embedded stage, organizations systematically integrate comprehensive documentation into every phase of the AI lifecycle, enabling transparency, accountability, and responsible management aligned with AI Trustworthy Principles. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Valid & Reliable | Undocumented Development and Operation Details | Assumptions, limitations, and techniques for AI system development and operation are undocumented |

**Implementation Guidelines:**

Document the assumptions, techniques, and metrics used for the development, deployment, and ongoing operation of the AI system across its entire lifecycle, ensuring consistency with the organization’s data governance policies. This includes detailed documentation of data selection, curation, preprocessing, and analysis techniques, as well as identifying modeled constructs and methods for inferring meaningful relationships between dataset attributes and constructs.

Engage relevant stakeholders, such as data scientists, data engineers, data governance professionals, and domain experts, in the documentation process to gather input on appropriate techniques and metrics at each stage of the AI lifecycle. This collaboration helps ensure that assumptions are valid, techniques are suitable, and metrics effectively measure trustworthiness and performance.

Develop a standardized framework for documenting data-related assumptions, methods, and metrics, ensuring it aligns with organizational data governance standards and best practices. This framework should be applied consistently throughout the lifecycle.

Regularly review and update the documentation to reflect changes in the system, data sources, techniques, and organizational standards, maintaining accuracy and relevance over time. Communicate updates to relevant stakeholders and provide training or support to foster organizational understanding and adherence to documented practices, supporting transparency and accountability throughout the AI lifecycle.

**Example Controls and Evidence:**

**Example Control 1:** Development of a Comprehensive Documentation Framework

**Description:** Establish standardized templates and processes for recording assumptions, limitations, techniques, and metrics at each lifecycle stage, aligned with data governance standards.

**Examples of Effective Evidence:**

• Formalized templates capturing assumptions, limitations, techniques, and metrics.

• Version-controlled documentation linked to system development and deployment phases.

• Guidelines aligning technical practices with data governance policies.

• Checklists ensuring completeness and consistency.

**Example Control 2:** Documentation of Data Selection, Curation, and Preparation Techniques

**Description:** Record detailed descriptions of data sourcing, selection criteria, curation processes, and preparation methods used throughout the lifecycle.

**Examples of Effective Evidence:**

• Data provenance and sourcing documentation.

• Records of data cleaning, transformation, and feature engineering techniques.

• Data quality assessment reports.

• Data processing workflows with version control.

**Example Control 3:** Identification of Modeled Constructs and Relationships

**Description:** Clearly document the modeled constructs, underlying assumptions, and the methods used to infer relationships between dataset attributes and constructs, including rationale and limitations.

**Examples of Effective Evidence:**

• Definitions and descriptions of key constructs and attributes.

• Modeling assumptions and hypotheses documentation.

• Methods and algorithms used for inferring relationships, with justifications.

• Rationale for model choices in relation to dataset features.

**Example Control 4:** Metrics and Validation Techniques

**Description:** Record the metrics and validation techniques used to evaluate model performance, data adequacy, and relationship inference accuracy, ensuring alignment with data governance policies.

**Examples of Effective Evidence:**

• Model performance metrics and validation reports.

• Data validation reports assessing attribute quality and integrity.

• Documentation of cross-validation, sensitivity analysis, and robustness checks.

• Logs of model tuning and metric tracking over time.

**Example Control 5:** Periodic Review and Update of Documentation

**Description:** Schedule regular reviews to update assumptions, limitations, techniques, and metrics, particularly as new data sources, methods, or system features evolve.

**Examples of Effective Evidence:**

• Review schedules and meeting minutes.

• Version histories of documents reflecting updates.

• Feedback from data scientists, model auditors, or governance reviews.

• Updated data governance policies incorporated into technical documentation.

**MEASURE / Methods and Metrics / External Assessment of Measurement Approaches**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MS-1.3.1 | Independent Expert Evaluation and Stakeholder Consultation | The organization involves internal experts (external to front-line development) and independent assessors, including robust testing functions, in the regular assessment and updating of AI metrics and risk controls. Concurrently, it determines the need for consultation with domain experts, users, AI stakeholders external to the development team, and affected communities, based on risk appetite. This approach leverages multidisciplinary, unbiased insights to ensure a comprehensive understanding of potential impacts and the adequacy of risk management measures.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/measure/#measure-1-3) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | Organizations at the initial stage begin the foundational process of involving internal experts (separate from the development team) and/or independent assessors in the basic assessment of AI metrics and risk controls. |
| **Stage 2: Minimal** | Yes | At the minimal stage, organizations more formally involve internal experts and/or independent assessors in the assessment and update AI metrics and risk controls for their limited AI applications. This includes engaging a broader range of domain experts and users to gather multidisciplinary perspectives on the appropriateness and effectiveness of these controls. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations consistently involve internal experts who were not part of the front-line development team and/or independent assessors in the regular assessment and updating of AI metrics and risk controls, leveraging their insights for an unbiased evaluation. This includes engaging domain experts, users, and other relevant AI stakeholders external to the development and deployment team to gather multidisciplinary perspectives on the appropriateness and effectiveness of AI metrics and risk controls. |
| **Stage 4: Embedded** | Yes | Organizations at the embedded stage have advanced, automated processes for seamlessly involving internal experts who were not part of the front-line development team and/or independent assessors in the continuous assessment and updating of AI metrics and risk controls, with real-time leveraging of their insights for an unbiased evaluation based on data-driven insights. The engagement of domain experts, users, and other relevant AI stakeholders external to the development and deployment team is deeply integrated into these processes, ensuring the highest level of multidisciplinary perspectives, appropriateness, and effectiveness of AI metrics and risk controls. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Explainable & Interpretable | Inadequate Independent Review and Stakeholder Consultation | Independent experts are not involved in AI metrics and risk controls, or stakeholder consultation is insufficient |

**Implementation Guidelines:**

Identify and engage internal experts (e.g., risk management, legal, compliance, independent testing functions) who were not part of the front-line development team, to participate in the regular assessment and updating of AI metrics and risk controls. Establish processes for involving independent assessors, such as third-party auditors or dedicated testing units with clear authority and independence, defining clear selection criteria based on expertise, reputation, and independence.

Determine the need for consultation with domain experts, users, AI stakeholders external to the development and deployment team, and affected communities during assessments, aligning with the organization's risk appetite and tolerance. Identify specific stakeholder groups based on the AI system's nature, potential impact, and risk management priorities.

Develop a comprehensive framework for collaboration among internal experts, independent assessors, and the AI development/deployment team. Concurrently, create stakeholder engagement plans outlining methods and frequency of consultation, considering system complexity, potential adverse impacts, and stakeholder interest. Establish clear communication channels and feedback mechanisms to effectively capture and incorporate all expert and stakeholder input into the assessment process.

**Example Controls and Evidence:**

**Example Control 1:** Independent Expert Assessment of AI Metrics

**Description:** Engage independent experts to conduct regular evaluations of AI performance metrics and risk controls, ensuring objectivity and adherence to industry best practices in risk management.

**Examples of Effective Evidence:**

• Reports detailing the findings from independent assessments of AI metrics.

• Documentation of metrics evaluated, including historical comparisons and performance trends.

• Records of expert qualifications, including credentials and areas of expertise.

• Action plans developed in response to expert recommendations, including implementation timelines.

**Example Control 2:** Stakeholder Consultation Framework

**Description:** Establish a framework for consulting with stakeholders, including domain experts, users, and affected communities, to gather insights and feedback on AI system performance and associated risks.

**Examples of Effective Evidence:**

• Consultation plans that outline stakeholder engagement strategies and timelines.

• Minutes or summaries of meetings held with stakeholders detailing discussions and insights.

• Feedback surveys distributed to stakeholders, including analysis of responses received.

• Documentation of how stakeholder feedback influenced risk management decisions or practices.

**Example Control 3:** Comprehensive Risk Assessment Processes

**Description:** Implement robust risk assessment processes that involve multidisciplinary teams, including both internal experts and independent assessors, to evaluate potential impacts and effectiveness of existing risk controls.

**Examples of Effective Evidence:**

• Risk assessment reports generated from team evaluations, including methodologies used.

• Risk matrix or dashboard tracking identified risks versus implemented controls.

• Records of team composition, including roles and expertise areas represented in assessments.

• Follow-up actions taken based on assessment results, with corresponding documentation.

**Example Control 4:** Training and Development for Expert Evaluators

**Description:** Provide ongoing training and professional development for internal experts to ensure they remain knowledgeable about the latest trends, challenges, and best practices in AI risk management and assessment.

**Examples of Effective Evidence:**

• Training materials and agendas used in professional development sessions for experts.

• Attendance records showing participation in training programs and workshops.

• Feedback from participants regarding the relevance and effectiveness of training provided.

• Certifications or qualifications obtained by internal experts through completed training.

**Example Control 5:** Documentation and Review of Consultation Outcomes

**Description:** Maintain comprehensive documentation of all stakeholder consultations and expert evaluations, regularly reviewing outcomes to enhance risk management practices and adapt to evolving contexts.

**Examples of Effective Evidence:**

• Comprehensive records of consultation sessions, including participant lists and discussion points.

• Evaluation reports summarizing the integration of stakeholder input into AI system design or risk management.

• Change logs that detail modifications made to risk controls based on expert and stakeholder feedback.

• Reviews conducted on the effectiveness of consultation processes, including actions taken to address identified gaps.

**MEASURE / Evaluating AI Systems / Measuring Privacy**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MS-2.10.1 | Initial Privacy Risk Assessment | The organization conducts an initial examination of the privacy risks associated with its AI systems, as identified during the MAP function, and documents the results of this examination, including any identified privacy issues, concerns, or potential violations specific to AI systems. It establishes accountability mechanisms for managing AI-related privacy risks and incidents, such as data breaches or unauthorized access, and assigns clear roles and responsibilities for privacy risk management.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/measure/#measure-2-10) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | Organizations at the initial stage conduct an initial examination of the privacy risks associated with their AI systems. They document the results of this examination, including any identified privacy issues, concerns, or potential violations specific to AI systems and establish basic accountability mechanisms for managing AI-related privacy risks. |
| **Stage 2: Minimal** | Yes | Organizations at the minimal stage conduct a more detailed examination of the privacy risks associated with their AI systems. They document any identified privacy issues, concerns, or potential violations specific to AI systems and establish accountability mechanisms for managing AI-related privacy risks and incidents. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations conduct an initial examination of the privacy risks associated with their AI systems, as identified during the MAP function, and comprehensively document the results of this examination, including any identified privacy issues, concerns, or potential violations specific to AI systems. The organization establishes accountability mechanisms for managing AI-related privacy risks and incidents, such as data breaches or unauthorized access, and assigns clear roles and responsibilities for privacy risk management. |
| **Stage 4: Embedded** | Yes | Organizations at the embedded stage have advanced, automated processes for continuously conducting an initial examination of the privacy risks associated with their AI systems, as identified during the MAP function, and dynamically documenting the results of this examination, including any identified privacy issues, concerns, or potential violations specific to AI systems, with real-time updates based on data-driven insights. The organization establishes accountability mechanisms for managing AI-related privacy risks and incidents, such as data breaches or unauthorized access, and assigns clear roles and responsibilities for privacy risk management, ensuring the highest level of proactive privacy risk identification and management. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Privacy-Enhanced | Unexamined Privacy Risks | Privacy risks are not examined, documented, or managed for AI systems |

**Implementation Guidelines:**

Establish a process for conducting an initial examination of the privacy risks associated with AI systems, as identified during the MAP function. Develop a standardized methodology for this examination, including criteria for evaluating the severity and likelihood of identified privacy issues, concerns, or potential violations specific to AI systems.

Document the results of the initial privacy risk examination, including a clear description of each identified risk, its potential impact on individuals and the organization, and any recommended mitigation strategies. Establish accountability mechanisms for managing AI-related privacy risks and incidents, such as data breaches or unauthorized access, and assign clear roles and responsibilities for privacy risk management to ensure that identified risks are properly addressed.

**Example Controls and Evidence:**

**Example Control 1:** Conduct of a Formal Privacy Risk Examination

**Description:** Develop and execute a structured process to evaluate privacy risks associated with AI systems, focusing on data handling, consent, access controls, and potential privacy violations.

**Examples of Effective Evidence:**

• Privacy risk assessment protocols aligned with privacy standards and legal requirements.

• Documentation of identified privacy issues, concerns, or violations specific to AI.

• Risk registers highlighting privacy vulnerabilities.

• Version-controlled examination reports.

**Example Control 2:** Documentation of Privacy Risks and Concerns

**Description:** Record findings related to privacy issues, potential violations, and concerns, with detailed descriptions, risk levels, and contextual information to create a comprehensive understanding of the privacy landscape surrounding AI systems.

**Examples of Effective Evidence:**

• Privacy risk logs with descriptions, severity, and affected data or individuals.

• Impact analysis reports on privacy violations or potential breaches.

• Data flow diagrams illustrating privacy-sensitive data.

**Example Control 3:** Establishment of Accountability Mechanisms

**Description:** Define and formalize accountability structures and procedures for managing AI-related privacy risks, including incident response, breach management, and data security protocols to ensure preparedness and effective handling of privacy incidents.

**Examples of Effective Evidence:**

• Privacy governance policies detailing roles and responsibilities.

• Incident response plans specific to privacy breaches.

• Appointment letters or role descriptions for privacy officers or data protection officers.

• Escalation and communication protocols for privacy incidents.

**Example Control 4:** Clear Assignment of Roles and Responsibilities

**Description:** Assign specific roles for privacy risk management, including monitoring, enforcement, incident investigation, and reporting, with documented responsibilities to clarify expectations and promote accountability across the organization.

**Examples of Effective Evidence:**

• Responsibility matrices or RACI charts.

• Training records emphasizing privacy responsibilities.

• Standard operating procedures (SOPs) for privacy incident handling.

• Role-specific KPIs linked to privacy management.

**Example Control 5:** Ongoing Monitoring, Review, and Improvement

**Description:** Regularly revisit privacy risk assessments, incident management procedures, and roles to incorporate lessons learned, policy updates, and evolving threats, ensuring dynamic risk management that adapts to changing circumstances and improves overall privacy protection.

**Examples of Effective Evidence:**

• Review schedules and meeting minutes with updates.

• Incident logs and post-incident reports to inform improvements.

• Updated policies and procedures reflecting new risks or standards.

• Feedback from stakeholders on privacy risk handling effectiveness.

**MEASURE / Evaluating AI Systems / Measuring Nondiscrimination**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MS-2.11.1 | Context-Specific Fairness and Bias Evaluation | The organization conducts evaluations of fairness and bias in its AI systems, focusing on any regulatory requirements and the risks and potential impacts identified during the MAP function. The organization employs techniques and tools to assess fairness and bias, examining potential sources of bias throughout the AI lifecycle, including systemic bias, statistical and computational bias, and human cognitive bias. Based on the evaluation results, the organization develops and implements strategies to mitigate or manage identified biases and fairness issues. The organization documents the results of its fairness and bias evaluations, including identified issues and mitigation strategies.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/measure/#measure-2-11) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | At the initial stage, organizations begin to conduct a preliminary evaluation of fairness and bias in their AI systems, even if limited in scope. They employ basic techniques to identify potential sources of bias, and document the results. Establishing this awareness early is crucial, even if formal mitigation strategies aren't immediately implemented. |
| **Stage 2: Minimal** | Yes | Organizations at the minimal stage conduct initial evaluations of fairness and bias on its AI systems. They employ simple techniques and tools to assess fairness and bias and begin examining potential sources of bias throughout the AI lifecycle. Based on these preliminary evaluation results, organizations start to develop and implement basic strategies to mitigate or manage identified biases and fairness issues, and document the results of its fairness and bias evaluations. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations consistently conduct evaluations of fairness and bias in their AI systems, focusing on the risks and potential impacts identified during the MAP function. The organization employs techniques and tools to assess fairness and bias, examining potential sources of bias throughout the AI lifecycle. Based on the evaluation results, the organization develops and implements strategies to mitigate or manage identified biases and fairness issues. The organization comprehensively documents the results of its fairness and bias evaluations, including identified issues and mitigation strategies. |
| **Stage 4: Embedded** | Yes | Organizations at the embedded stage have advanced, automated processes for continuously conducting evaluations of fairness and bias in their AI systems, focusing on the risks and potential impacts identified during the MAP function, with real-time employment of techniques and tools to assess fairness and bias, examining potential sources of bias throughout the AI lifecycle based on data-driven insights. Based on the evaluation results, the organization dynamically develops and implements strategies to mitigate or manage identified biases and fairness issues. The organization dynamically documents the results of its fairness and bias evaluations, including identified issues and mitigation strategies, ensuring the highest level of proactive fairness and bias assessment and management. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Fair | Unconducted Fairness Evaluations | Context-specific fairness and bias evaluations are not conducted or documented for AI systems |

**Implementation Guidelines:**

Establish a process for conducting context-specific evaluations of fairness and bias in AI systems, focusing on the risks and potential impacts identified during the MAP function. It is also helpful to have clear definitions of systemic, computational, and human cognitive bias. Areas of increased risk tend to include lending and employment applications. Develop a standardized methodology for these evaluations, including the selection and use of appropriate techniques and tools to assess fairness and bias at various stages of the AI lifecycle.

Document the results of fairness and bias evaluations, including identified issues, potential sources of bias, and recommended mitigation strategies while addressing any applicable regulatory reporting requirements. Use these findings to inform the development and implementation of targeted interventions to address identified biases and fairness concerns, such as data preprocessing techniques, model adjustments, or post-processing methods.

**Example Controls and Evidence:**

**Example Control 1:** Development of Fairness and Bias Evaluation Framework

**Description:** Create formal procedures, criteria, and tools for assessing fairness and bias across the AI lifecycle, aligned with regulatory requirements and organizational risk thresholds.

**Examples of Effective Evidence:**

• Standardized assessment protocols covering data, model, and output biases.

• Criteria checklists for bias detection tailored to use cases.

• Version-controlled evaluation frameworks.

• Guidelines for identifying bias sources at each stage.

**Example Control 2:** Application of Bias and Fairness Testing Techniques

**Description:** Use validated techniques such as statistical bias testing, fairness metrics, adversarial testing, and data audits to systematically evaluate biases and fairness throughout development and deployment

**Examples of Effective Evidence:**

• Test reports displaying fairness and bias metrics (e.g., demographic parity, equal opportunity).

• Data audit logs identifying biased data sources or features.

• Results from adversarial or scenario-based bias tests.

• Model validation reports highlighting fairness issues.

**Example Control 3:** Documentation of Evaluation Results and Issues

**Description:** Record findings from bias and fairness assessments, including mapped issues, sources, and severity, providing a basis for risk management and mitigation planning

**Examples of Effective Evidence:**

• Bias assessment reports with detailed issue descriptions.

• Risk registers linking fairness issues with potential impacts.

• Source analysis including data, model, or implementation-related biases.

• Issue tracking dashboards.

**Example Control 4:** Development and Implementation of Bias Mitigation Strategies

**Description:** Based on evaluation outcomes, design and execute mitigation actions such as data balancing, model adjustments, or policy changes, documenting decisions and actions take

**Examples of Effective Evidence:**

• Bias mitigation plans with assigned responsibilities.

• Model retraining or data augmentation records.

• Policy updates affecting fairness considerations.

• Follow-up testing to verify mitigation effectiveness.

**Example Control 5:** Ongoing Monitoring, Review, and Improvement

**Description:** Continuously monitor bias and fairness metrics post-deployment, review evaluation processes periodically, and refine mitigation strategies based on new insights, regulations, and lessons learned

**Examples of Effective Evidence:**

• Monitoring dashboards tracking fairness metrics over time.

• Regular review reports assessing bias levels and mitigation success.

• Lessons learned documentation from bias-related incidents.

• Revisions to evaluation frameworks guided by evolving standards.

**MEASURE / Tracking AI Risks / Enabling Feedback, Appeal, and Override**

**Control Objective:**

|  |  |  |
| --- | --- | --- |
| **Id** | **Name** | **Description** |
| MS-3.3.1 | Feedback and Appeals Processes | The organization establishes feedback processes for relevant stakeholders to report problems, concerns, or unintended consequences related to AI system outcomes, integrating with existing processes for reporting security issues and potential vulnerabilities. This includes implementing an appeals mechanism that allows users to challenge or request reviews of outcomes they believe to be inaccurate, unfair, or harmful. Both processes are documented clearly, including channels, timeframes, responsibilities, escalation procedures, and criteria for decision-making, with all reported issues tracked to closure.  [Link to NIST AI RMF Playbook](https://airc.nist.gov/airmf-resources/playbook/measure/#measure-3-3) |

**Adoption Stage Applicability:**

| **Adoption Stage** | **Applicability** | **Rationale** |
| --- | --- | --- |
| **Stage 1: Initial** | Yes | Organizations at the initial stage start to establish basic feedback processes for relevant stakeholders to report any problems, concerns, or unintended consequences related to initial AI system outcomes. This might involve setting up a simple mechanism for users to flag potentially inaccurate or unfair outcomes. |
| **Stage 2: Minimal** | Yes | Organizations at the minimal stage establish basic feedback processes for relevant stakeholders to report any problems, concerns, or unintended consequences related to AI system outcomes. This can include an appeals mechanism that allows users to challenge or request basic reviews of outcomes they believe to be inaccurate, unfair, or harmful, with processes documented clearly, including channels, timeframes, responsibilities, and criteria for decision-making. |
| **Stage 3: Evolving** | Yes | At the evolving stage, organizations expand AI usage into external-facing solutions that involve sensitive data, so establishing robust feedback processes becomes vital. In this stage, while AI might not yet drive critical business decisions, any unintended outcomes could affect external stakeholders or broader operations, making documented reporting and appeals processes essential for mitigating risks and ensuring accountability. |
| **Stage 4: Embedded** | Yes | Organizations at the embedded stage have advanced, automated processes for continuously establishing feedback processes for relevant stakeholders to report problems, concerns, or unintended consequences related to AI system outcomes, with real-time implementation of an appeals mechanism that allows users to challenge or request reviews of outcomes they believe to be inaccurate, unfair, or harmful, with both processes documented clearly, including channels, timeframes, responsibilities, and criteria for decision-making based on data-driven insights. These processes are seamlessly integrated into the organization's risk management and stakeholder engagement frameworks, ensuring the highest level of proactive stakeholder feedback and appeals mechanism establishment and documentation. |

|  |  |  |
| --- | --- | --- |
| **AI Principle** | **Risk Name** | **Risk Statement** |
| Accountable & Transparent | Insufficient Feedback Mechanisms | Feedback processes and appeals mechanisms for AI outcomes are insufficient |

**Implementation Guidelines:**

Develop and implement feedback processes that enable relevant stakeholders, such as end-users and impacted communities, to report problems, concerns, or unintended consequences related to AI system outcomes. Integrate these processes with existing mechanisms for reporting security issues and potential vulnerabilities. Establish clear channels for submitting feedback, such as dedicated email addresses, online forms, or helpline numbers, and ensure these are easily accessible and widely communicated. Such processes may be mandated for credit lending (e.g., adverse action notices) or other high-impact applications.

Note: To align with MEASURE 2.8 and MEASURE 2.9, incorporate assessments to verify stakeholders’ ability to understand AI communications and interpret explainability data, ensuring feedback channels effectively capture whether users can comprehend and challenge AI decisions based on clear explanations.

Design an appeals mechanism that allows users to challenge or request reviews of AI system outcomes they believe to be inaccurate, unfair, or harmful. Define clear criteria for determining which outcomes are eligible for appeal, and establish a standardized process for reviewing and resolving appeals, including timeframes for response and escalation procedures for complex cases. Document both the feedback and appeal processes clearly, outlining channels, timeframes, responsibilities, escalation procedures, and criteria for decision-making. Ensure this documentation is regularly reviewed and updated to reflect changes in the organization's AI systems and stakeholder needs, with all reported issues tracked to closure.

**Example Controls and Evidence:**

**Example Control 1:** Development of Feedback and Appeals Process Documentation

**Description:** Create formal, accessible procedures for collecting stakeholder feedback and handling appeals, with comprehensive documentation of channels, timeframes, roles, and decision criteria.

**Examples of Effective Evidence:**

• Process manuals outlining steps for feedback submission and appeal handling.

• Flowcharts showing escalation paths and decision points.

• Templates for feedback and appeal forms.

• Documented roles and responsibilities of staff managing feedback and appeals.

**Example Control 2:** Implementation of Feedback and Appeal Channels

**Description:** Deploy multiple, easily accessible channels for stakeholders to submit feedback and appeal outcomes (e.g., online portals, contact centers, dedicated email addresses) to ensure a user-friendly experience that encourages reporting and engagement from all relevant parties.

**Examples of Effective Evidence:**

• Website interfaces, portals, or apps with feedback and appeal options.

• Communication logs documenting received feedback and appeals.

• Helpdesk or contact center records.

• Channel usage metrics and response times.

**Example Control 3:** Clear Timeframes and Responsibilities for Response

**Description:** Define and document standard timeframes for acknowledgment, investigation, resolution, and communication of feedback and appeals, assigning responsible personnel or teams to promote accountability and ensure timely responses to reported issues.

**Examples of Effective Evidence:**

• Service level agreements (SLAs) with time targets.

• Response tracking logs showing adherence to timeframes.

• Standardized response scripts and templates.

• Incident and appeal case logs.

**Example Control 4:** Criteria and Decision-Making Processes for Resolution

**Description:** Establish and document the criteria, guidelines, and decision processes used to evaluate reported issues and make determinations during appeals to maintain transparency and consistency in how feedback and appeals are handled.

**Examples of Effective Evidence:**

• Decision matrices or guidelines for assessing fairness, accuracy, or harm.

• Decision review meeting minutes with documented rationale.

• Audit logs of decision approvals or escalations.

• Appeals resolution summaries with explanations.

**Example Control 5:** Monitoring, Evaluation, and Continuous Improvement

**Description:** Track feedback and appeal trends, evaluate resolution effectiveness, and regularly review processes for improvements, incorporating lessons learned and stakeholder input to enhance the overall effectiveness of feedback and appeals mechanisms and ensure they meet user needs.

**Examples of Effective Evidence:**

• Feedback and appeal logs with analysis of common issues.

• Satisfaction surveys from stakeholders.

• Process review reports outlining improvements.

• Updated procedures reflecting lessons learned.