Artificial Intelligence Risk Management Framework (AI RMF 1.0)

NIST AI 100-1

Controls Checklist

Version: 0.1 (Technical Preview Draft)

Last Update: 09 June 2025

TABLE OF CONTENT

Contents

DISCLAIMER

This document is intended exclusively for professional community development. The information provided herein is for GRC professionals’ reference only and does not constitute professional advice.

No Warranties or Liability

This document is provided "as is" without any express or implied warranties. We assume no responsibility and disclaim all liability for any damages arising from the use of the information contained in this document.

Seek Professional Guidance or Advice

The information contained in this document should not be applied to practical situations without consulting a qualified professional.

Verify Before Use

This content includes AI-assisted information. It may contain errors or incorrect information. Please verify all information carefully before use.

Reference Source

Artificial Intelligence Risk Management Framework (AI RMF 1.0)

https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

Version

Version	Updated by	Remarks
0.1	Helena (GRC Library AI Virtual Consultant)	First draft (Tech Preview Draft)

INTRODUCTION

This document is a compliance control checklist designed to assist compliance managers and auditors in their compliance or internal control checks. It includes summary information to help streamline the process, making it a valuable reference.

The checklist maps compliance requirements directly to control activities. Be aware that there may be duplicated controls within the document.

This document is intended to support Governance, Risk, and Compliance (GRC) professionals in their work. We hope this checklist can serve as reference material for saving GRC professionals studying time on regulation requirements.

This document is prepared by GRC Library AI Virtual Consultant, Helena. She specializes in preparing EU Regulation and Directive document deliverables. More information about Virtual Consultant, including her skills and work performance, is available in the GRC Library.

Helena’s profile: https://grclibrary.com/team_profile.php?profile=helena

We hope you find it helpful with this reference template.

Best Regards,

GRC Library

URL: https://grclibrary.com
LinkedIn: https://www.linkedin.com/company/grclibrary
YouTube: https://www.youtube.com/@grclibrary

GRC LIBRARY

Thank you for taking the time to read this Controls Checklist document. More resources and tools can be found in the GRC Library. We also appreciate any feedback or comments you may have on the document. Your input helps us improve our resources for everyone. Don't hesitate to visit the GRC Library to explore other resources that can support your GRC activities.

Best regards,

GRC Library https://grclibrary.com

CONTROLS CHECKLIST

1. AI RMF Review and Version Control Process

Control Actor: NIST

Control Types: administrative, preventive

Related Entities: NIST

Control Description:

Establish and maintain a comprehensive process for regularly reviewing the content and usefulness of the AI Risk Management Framework (AI RMF) and its companion documents (e.g., AI RMF Playbook) to determine if updates are appropriate and to ensure frequent updates. This process should include mechanisms for formal input from the AI community, such as receiving and reviewing comments on the AI RMF Playbook via email (e.g., AIframework@nist.gov) and integrating relevant comments on a semi-annual basis, with a comprehensive review of the AI RMF expected no later than 2028. Additionally, implement and utilize a two-number versioning system (major.minor) for the AI RMF and its companion documents, maintaining a Version Control Table that documents the history of changes, including version number, date, and description.

Enhanced Implementation Guide:

The enhanced description for the "AI RMF Review and Version Control Process" control is as follows:

Control Statement: Establish and maintain a robust, auditable process for the continuous review, update, and version management of the AI Risk Management Framework (AI RMF) and its companion documents (e.g., AI RMF Playbook), ensuring their ongoing relevance, accuracy, and responsiveness to community feedback.

High-Level Control Implementation Guide:
1. **Review and Update Process:**
    * Designate a standing committee or team responsible for overseeing the AI RMF and companion document lifecycle, including regular reviews and incorporation of feedback.
    * Establish formal communication channels (e.g., a dedicated email inbox like `AIframework@nist.gov`) for receiving external community comments and feedback.

    *   Develop and document a clear workflow for the collection, categorization, evaluation, and prioritization of all received comments, specifying criteria for determining "relevance" and "appropriateness" of updates.
    *   Define an approval authority and process for all proposed modifications and updates to the AI RMF and its companion documents, ensuring stakeholder consensus.
    *   Schedule and conduct semi-annual reviews of companion documents to integrate relevant feedback, with a comprehensive review of the core AI RMF explicitly planned for completion no later than 2028 (and subsequent periodic reviews).
2. **Versioning Control System:**
    *   Implement a mandatory two-number versioning system (major.minor) for the AI RMF and all its companion documents. Major version increments (e.g., 1.0 to 2.0) should denote significant changes or comprehensive updates, while minor version increments (e.g., 1.0 to 1.1) should denote smaller revisions, corrections, or integration of minor feedback.
    *   Establish and maintain a centralized, accessible Version Control Table or similar system for each document. This table must meticulously document the history of all changes, including:
        *   Unique Version Number (major.minor)
        *   Date of Change
        *   Clear and concise Description of Changes (e.g., summary of edits, reasons for update, integrated feedback details)
        *   Author/Approver of Change
    *   Ensure that all published versions of the AI RMF and companion documents prominently display their current major.minor version number.

Control Frequency:
*   **Content Review & Input Integration:** Semi-annually for companion documents; Ad-hoc for critical updates or corrections.
*   **Comprehensive AI RMF Review:** No later than 2028, with subsequent reviews planned quinquennially or as deemed necessary by significant technological or policy shifts.
*   **Versioning:** Immediately upon any approved change or update to a document.

Monitoring Frequency: Quarterly, to verify adherence to review schedules, feedback integration, and version control procedures.

Control Type: Semi-Automated (involves manual collection/review of input and decision-making, supported by automated systems for email management, document storage, and version control tools).

Guide for Evaluating Control Effectiveness:

Effectiveness can be evaluated by:

1. **Review Process Adherence:** Verifying documented evidence of semi-annual review meetings/activities for companion documents and tracking progress towards the comprehensive AI RMF review by 2028.
2. **Feedback Integration:** Inspecting comment logs or similar records to confirm formal receipt, review, and disposition of community feedback, correlating integrated changes with relevant comments.
3. **Versioning System Integrity:** Confirming that all AI RMF and companion documents consistently utilize the major.minor versioning system. Auditing Version Control Tables for completeness, accuracy, and consistency with published document versions, and verifying that changes are properly recorded with descriptions and dates.
4. **Accessibility and Transparency:** Ensuring that the process for submitting feedback is clear and accessible, and that updated documents are promptly published and easily retrievable.

Key Control Indicators (KCIs):

1. **Review Cycle Compliance Rate:** Percentage of semi-annual review cycles completed on schedule for companion documents.
2. **Feedback Integration Rate:** Percentage of relevant community comments and identified necessary updates integrated into documents per review cycle.
3. **Comprehensive Review Progress:** Tracking milestones and projected completion of the comprehensive AI RMF review against the 2028 deadline.
4. **Versioning System Adherence Rate:** Percentage of AI RMF and companion documents that consistently follow the major.minor versioning scheme and have accurate, up-to-date Version Control Tables.
5. **Change Audit Trail Quality:** Number of discrepancies or missing entries identified in Version Control Tables during audits.
6. **Response Time to Critical Feedback:** Average time taken to address and integrate critical feedback or corrections into documents.

Requirement Statements:

1. NIST will review the content and usefulness of the Framework regularly to determine if an update is appro- priate; a review with formal input from the AI community is expected to take place no later than 2028. (Page: 3)

2. The Framework will employ a two-number versioning system to track and identify major and minor changes. The first number will represent the generation of the AI RMF and its companion documents (e.g., 1.0) and will change only with major revisions. Minor revisions will be tracked using “.n” after the generation number (e.g., 1.1). All changes will be tracked using a Version Control Table which identifies the history, including version number, date of change, and description of change. (Page: 3)

3. NIST plans to update the AI RMF Playbook frequently. (Page: 3)

4. Comments on the AI RMF Playbook may be sent via email to AIframework@nist.gov at any time and will be reviewed and integrated on a semi-annual basis. (Page: 3)

5. While risk management processes generally address negative impacts, this Framework offers approaches to minimize anticipated negative impacts of AI systems and identify opportunities to maximize positive impacts. (Page: 9)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

2. AI RMF Core Functions Implementation and Impact Management

Control Actor: Risk Management Team

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain comprehensive processes to implement the four core functions of the AI Risk Management Framework (GOVERN, MAP, MEASURE, and MANAGE) across the AI system lifecycle, ensuring these efforts are conducted continuously and in a timely manner, integrating them within organizational processes and procedures. These processes should be iterative, allowing for continuous refinement and improvement, and include mechanisms for cross-referencing between functions (GOVERN, MAP, MEASURE, MANAGE) as necessary to ensure holistic and integrated risk management. These processes are designed to effectively address, document, and manage the risks and potential negative impacts associated with AI systems, fostering trustworthy AI. These comprehensive risk management efforts should ideally commence from the Plan and Design function within the application context and continue throughout the entire AI system lifecycle, encompassing all dimensions. Specifically, this includes approaches to minimize anticipated negative impacts, identify opportunities for positive impacts, and enable AI developers and users to understand the impacts and account for the inherent limitations and uncertainties in AI models and systems, including accounting for potential human assumptions that AI systems work well in all settings, to improve overall system performance and trustworthiness and ensure beneficial uses of AI technologies. This comprehensive risk management process should cover the development, deployment, and use of AI systems, whether they function as standalone or integrated components. The GOVERN function applies to all stages of an organization's AI risk management, while the MAP, MEASURE, and MANAGE functions can be applied in AI system-specific contexts and at particular stages of the AI lifecycle.

Enhanced Implementation Guide:

Control Statement: Establish, implement, and continuously maintain a comprehensive and iterative AI Risk Management Framework (AI RMF) across the entire AI system lifecycle, from planning and design through development, deployment, and ongoing use. This framework must effectively leverage the GOVERN, MAP, MEASURE, and MANAGE core functions to proactively identify, assess, mitigate, monitor, and manage AI-related risks and potential negative impacts, while also identifying opportunities for positive impacts. The aim is to foster trustworthy AI by ensuring developers and users understand inherent limitations and uncertainties, account for potential human assumptions, and ensure beneficial uses of AI technologies, whether standalone or integrated.

High-Level Control Implementation Guide: The Risk Management Team, in collaboration with AI development, legal, and operational teams, shall lead the establishment and continuous refinement of the AI RMF. This involves:

1. **GOVERN:** Define and document organizational AI RMF policies, principles, risk appetite, roles, and responsibilities (including for third-party providers). Establish governance structures (e.g., AI ethics committee, AI risk council) and integrate AI risk management into existing enterprise risk management and GRC frameworks. Ensure clear accountability for AI system risks across the lifecycle.
2. **MAP:** Systematically identify and characterize AI risks and potential impacts (technical, societal, ethical, legal, commercial) from the initial plan and design phase, documenting AI system characteristics, data provenance, use cases, and inherent limitations. Conduct comprehensive AI impact assessments (e.g., bias assessments, privacy impact assessments) for all AI systems, cross-referencing findings with governance objectives.
3. **MEASURE:** Develop and apply qualitative and quantitative metrics and methodologies to assess AI system performance, trustworthiness attributes (e.g., fairness, explainability, robustness, security, privacy), and overall risk levels. Establish clear thresholds for acceptable performance and risk, and implement tools for continuous data collection and analysis.
4. **MANAGE:** Implement robust risk mitigation strategies, controls, and incident response plans for identified AI risks. This includes developing transparent communication protocols for AI system limitations and impacts, fostering stakeholder feedback loops, and ensuring continuous monitoring of AI system performance and risk profiles. Promote iterative improvement cycles based on lessons learned from risk events, audits, and performance data, cross-referencing insights gained from measurement and impact assessments.

These processes must be integrated into the AI system lifecycle, ensuring continuous application and timely adaptation to emerging risks and new AI initiatives, starting from the earliest conceptual stages.

Control Frequency: Continuous (as AI systems are developed, deployed, and used), with formal reviews and updates to the overall AI RMF and its integration mechanisms conducted at least bi-annually or upon significant organizational/technological changes. AI system-specific risk assessments and mitigation efforts should occur at key lifecycle gates (e.g., design, pre-deployment, post-deployment, significant model updates).

Monitoring Frequency: Quarterly by the Risk Management Team to review dashboards and reports on AI RMF performance and risk profiles, with an annual independent assurance review (e.g., by Internal Audit or an external third party) of the AI RMF's design and operating effectiveness.

Control Type: Semi-Automated. While strategic oversight, policy setting, risk acceptance decisions, and qualitative impact assessments are manual, many aspects of risk identification (e.g., automated vulnerability scanning for AI components), impact measurement (e.g., bias detection tools, performance monitoring systems, data drift detection), and continuous monitoring of deployed AI systems can and should leverage automated tools and platforms.

Guide for Evaluating Control Effectiveness: Control effectiveness is evaluated by verifying:
1. **Existence and Adoption:** Comprehensive AI RMF policies, procedures, and governance structures are formally documented, widely communicated, and demonstrably integrated into the organization's existing risk management framework and AI system development lifecycle.
2. **Completeness of Risk Identification:** Evidence that all new and significantly modified AI systems undergo a thorough AI impact assessment (MAP) and risk identification process from the planning phase, covering all relevant risk dimensions (technical, societal, ethical, legal) and considering provider organization contributions.
3. **Robustness of Measurement:** Defined and implemented metrics (MEASURE) for AI system trustworthiness (e.g., fairness, explainability, robustness) and performance are consistently applied, and thresholds for acceptable risk/performance are established and monitored.
4. **Effectiveness of Mitigation and Management:** Documented mitigation plans for identified critical risks are in place, actively managed (MANAGE), and demonstrate a reduction in residual risk to acceptable levels. Incident response procedures for AI failures or negative impacts are defined and regularly tested.
5. **Iterative Improvement and Integration:** Evidence of continuous feedback loops (e.g., lessons learned from incidents, audit findings, performance data) informing iterative refinements of the AI RMF and AI system designs, demonstrating cross-referencing between

GOVERN, MAP, MEASURE, and MANAGE functions for holistic risk management.
6. **Accountability and Transparency:** Clear lines of accountability for AI risks are established (GOVERN), and AI system limitations, uncertainties, and anticipated impacts are transparently communicated to relevant stakeholders.

Key Control Indicators (KCIs):
* **AI RMF Coverage:** Percentage of AI systems (in development or production) with completed AI RMF assessments (GOVERN, MAP, MEASURE, MANAGE functions applied and documented).
* **Risk Mitigation Efficacy:** Percentage of identified critical/high AI risks with documented mitigation plans, assigned owners, and tracked progress against agreed-upon timelines.
* **Trustworthiness Performance:** Average scores or compliance rates against defined metrics for AI system fairness, explainability, robustness, and privacy (e.g., bias detection rates below threshold, model drift within tolerance).
* **Incident Management:** Number of AI-related incidents or significant negative impacts identified, and the average time to resolution or containment.
* **Governance Integration:** Number of AI governance committee meetings held and proportion of AI initiatives formally reviewed and approved by the governance structure.
* **Audit/Assessment Findings:** Number and severity of AI RMF-related findings from internal or external audits, and the timely completion rate of corrective actions.
* **AI RMF Maturity Score:** A periodic assessment score reflecting the maturity level of the organization's AI RMF implementation based on established frameworks (e.g., NIST AI RMF).

Requirement Statements:

1. Part 2 comprises the “Core” of the Framework. It describes four specific functions to help organizations address the risks of AI systems in practice. These functions – GOVERN, MAP, MEASURE, and MANAGE – are broken down further into categories and subcate- gories. While GOVERN applies to all stages of organizations’ AI risk management pro- cesses and procedures, the MAP, MEASURE, and MANAGE functions can be applied in AI system-specific contexts and at specific stages of the AI lifecycle. (Page: 8)

2. Addressing, documenting, and managing AI risks and potential negative impacts effectively can lead to more trustworthy AI systems. (Page: 9)

3. While risk management processes generally address negative impacts, this Framework offers approaches to minimize anticipated negative impacts of AI systems and identify opportunities to maximize positive impacts. (Page: 9)

4. Risk management can enable AI developers and users to understand impacts and account for the inherent lim- itations and uncertainties in their models and systems, which in turn can improve overall system performance and trustworthiness and the likelihood that AI technologies will be used in ways that are beneficial. (Page: 9)

5. AI risk management efforts should consider that humans may assume that AI systems work – and work well – in all settings. (Page: 9)

6. Regardless, all parties and AI actors should manage risk in the AI systems they develop, deploy, or use as standalone or integrated components. (Page: 10)

7. Once tolerance is defined, this AI RMF can be used to manage risks and to document risk management processes. (Page: 12)

8. Ideally, risk management efforts start with the Plan and Design function in the application context and are performed throughout the AI system lifecycle. (Page: 15)

9. Risk management should be continuous, timely, and performed throughout the AI system lifecycle dimensions. (Page: 25)

10. However users integrate the functions, the process should be iterative, with cross-referencing between functions as necessary. (Page: 26)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

3. Emergent AI Risk Identification and Continuous Measurement

Control Actor: Risk Management Team, AI Operations Team, AI Governance Committee

Control Types: preventive, detective, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes to proactively identify, continuously track, and conduct ongoing measurement of emergent risks associated with AI systems throughout their lifecycle. Acknowledge that risks can evolve over time and that different stakeholders across the AI lifecycle may hold varying risk perspectives. Develop and evaluate appropriate techniques and methodologies for measuring these emergent risks, recognizing and addressing challenges such as the lack of consensus on robust and verifiable methods. This includes the development and utilization of human baseline metrics for comparison when managing risks of AI systems designed to augment or replace human activity. Strive to develop or adopt suitable internal measurement approaches that avoid pitfalls like oversimplification, bias, or inadvertent reflection of factors unrelated to the underlying impact, ensuring timely mitigation and enhancing the overall effectiveness of the organization's AI risk management efforts. This also includes implementing ongoing monitoring of AI risks in operational, real-world settings, supplementing pre-deployment laboratory measurements to account for potential differences in risk emergence between controlled and real-world environments.

Enhanced Implementation Guide:

Control Statement: Establish and maintain a comprehensive, cross-functional framework and scalable processes for the proactive identification, continuous tracking, and robust measurement of emergent risks inherent to AI systems throughout their entire lifecycle, from design to real-world operation. This includes acknowledging the evolving nature of risks, integrating diverse stakeholder perspectives, developing suitable quantitative and qualitative measurement methodologies (including human baseline metrics for AI systems augmenting or replacing human activity), and implementing ongoing monitoring in operational environments to complement pre-deployment assessments, thereby ensuring timely mitigation and enhancing the overall effectiveness of the organization's AI risk management posture.

High-Level Control Implementation Guide:

1. **Framework Establishment:** Define a formal AI Risk Management Framework outlining

roles, responsibilities, and an integrated workflow involving the Risk Management Team, AI Operations Team, AI Governance Committee, and other relevant stakeholders (e.g., legal, ethics, business units) for identifying, assessing, and mitigating emergent AI risks.
2. **Proactive Risk Scouting:** Implement mechanisms for continuous environmental scanning and intelligence gathering (e.g., monitoring academic research, industry trends, regulatory changes, adversarial techniques) to anticipate and identify novel or evolving AI-specific threats (e.g., model brittleness, data poisoning, emergent bias patterns).
3. **Methodology Development & Application:** Research, select, and develop robust, verifiable methodologies and metrics for measuring emergent AI risks. This includes defining specific indicators for attributes like fairness, transparency, robustness, performance drift, and potential societal impacts. Where AI systems augment or replace human activity, establish and continuously collect human baseline metrics (e.g., human error rates, decision quality, efficiency) to serve as a comparative standard for AI risk assessment.
4. **Continuous Operational Monitoring:** Deploy automated and manual monitoring tools and processes for AI systems in live production environments. This ensures real-time detection of emergent risks that may manifest differently or unexpectedly compared to controlled laboratory settings (e.g., unexpected user interactions, shifts in real-world data distributions, adversarial attacks).
5. **Stakeholder Integration:** Actively solicit and integrate diverse risk perspectives from technical experts, ethical advisors, legal counsel, business owners, and affected stakeholders throughout the risk identification and measurement process.
6. **Reporting & Feedback Loop:** Establish regular reporting mechanisms to the AI Governance Committee and relevant operational teams on identified emergent risks, their measured impact, and the effectiveness of mitigation strategies. Ensure a continuous feedback loop that informs improvements to AI development, deployment practices, and the overall AI risk management framework.

Control Frequency:
*   **Identification & Tracking:** Continuous (daily/weekly review of intelligence feeds, real-time monitoring of operational AI systems).
*   **Measurement Methodology Review & Updates:** Annually, or ad-hoc as new AI systems are deployed, new risk types emerge, or measurement science advances.
*   **Formal Risk Assessment & Reporting:** Quarterly to the AI Governance Committee; monthly for operational teams to review real-time monitoring outputs.
*   **Human Baseline Collection & Comparison:** Continuous for relevant AI systems.

Monitoring Frequency:
*   **Control Effectiveness Monitoring:** Quarterly by the AI Governance Committee or

delegated compliance function, and annually by Internal Audit.
* **Performance Monitoring of AI Systems (as part of the control):** Continuous.

Control Type: Semi-automated (Leverages automated tools for data collection and real-time monitoring, combined with manual processes for expert analysis, qualitative risk assessment, strategic decision-making, methodology development, and stakeholder engagement).

Guide for Evaluating Control Effectiveness:
Effectiveness is evaluated by assessing the proactive nature, comprehensiveness, and responsiveness of the emergent AI risk management process. This includes:

1. **Documentation Review:** Verification of a well-defined AI Risk Management Framework, documented methodologies for emergent risk identification and measurement, established human baseline metrics, and records of regular risk assessments, mitigation plans, and incident reports specifically tied to emergent risks.
2. **Process Walkthroughs & Interviews:** Confirming that control actors (Risk Management Team, AI Operations Team, AI Governance Committee) understand and consistently adhere to established processes for risk identification, measurement, and reporting.
3. **Sampling & Evidence of Application:** Selecting a sample of AI systems in production and verifying that emergent risks were identified, appropriately measured using established methodologies (including human baselines where applicable), continuously monitored in real-world settings, and effectively addressed through timely mitigation actions.
4. **Timeliness & Efficacy of Response:** Assessing the speed and effectiveness with which newly identified or evolving emergent risks are mitigated, evidenced by a reduction in actual adverse impacts or incident frequency/severity.
5. **Integration of Perspectives:** Confirming active engagement and integration of diverse stakeholder perspectives (technical, ethical, legal, business, user feedback) in the ongoing identification and evaluation of emergent risks.
6. **Framework Adaptability:** Evidence of periodic review and enhancement of the AI risk management framework itself, demonstrating its ability to adapt to new AI technologies, use cases, and evolving risk landscapes.

Key Control Indicators (KCIs):
* **Number of Unique Emergent AI Risks Identified:** Quantifies the proactive discovery of new or evolving risks over time.
* **Average Time to Identify Emergent Risks:** Measures the efficiency of the risk scouting and monitoring mechanisms (e.g., from initial manifestation to formal documentation).
* **Coverage of Operational Monitoring:** Percentage of in-production AI systems continuously monitored for real-world emergent risks.

*   **Accuracy of Risk Measurement Methodologies:** Qualitative assessment of the correlation between measured risk levels and actual observed impact/incident frequency (e.g., through post-incident analysis).
*   **Timeliness of Mitigation Actions:** Average time from the identification of an emergent risk to the implementation of initial mitigation controls.
*   **Effectiveness of Mitigation:** Reduction in the severity or frequency of incidents attributed to previously emergent risks after mitigation actions are implemented.
*   **Number of AI Risk Methodology Reviews/Updates:** Indicates commitment to continuous improvement and adaptation of measurement approaches.
*   **Utilization Rate of Human Baselines:** Percentage of AI systems designed to augment/replace human activity for which human baseline metrics are actively established, collected, and utilized in risk assessment.
*   **Stakeholder Engagement Index:** Number of unique departments or expert groups contributing to emergent risk identification workshops or review meetings.
*   **AI Incident Rate from Emergent Risks (Post-Mitigation):** Tracking the rate of incidents specifically stemming from previously unmanaged or underestimated emergent risks, demonstrating control effectiveness.

Requirement Statements:

1. Organizations’ risk management efforts will be enhanced by identifying and tracking emergent risks and considering techniques for measuring them. (Page: 10)

2. The current lack of consensus on robust and verifiable measurement methods for risk and trustworthiness, and applicability to different AI use cases, is an AI risk measurement challenge. Potential pitfalls when seeking to measure negative risk or harms include the reality that development of metrics is often an institu- tional endeavor and may inadvertently reflect factors unrelated to the underlying impact. In addition, measurement approaches can be oversimplified, gamed, lack critical nuance, be- come relied upon in unexpected ways, or fail to account for differences in affected groups and contexts. (Page: 11)

3. Measuring risk at an earlier stage in the AI lifecycle may yield different results than measuring risk at a later stage; some risks may be latent at a given point in time and may increase as AI systems adapt and evolve. Fur- thermore, different AI actors across the AI lifecycle can have different risk perspectives. (Page: 11)

4. While measuring AI risks in a laboratory or a controlled environment may yield important insights pre-deployment, these measurements may differ from risks that emerge in operational, real-world settings. (Page: 11)

5. Risk management of AI systems that are intended to augment or replace human activity, for example decision making, requires some form of baseline metrics for comparison. (Page: 11)

6. Processes for assessing emergent risks are in place... (Page: 36)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ______________________________

Evidence/Comments:

____

____

____

4. AI System Lifecycle Risk Communication, Metrics Alignment, and Transparency

Control Actor: Risk Management Team, AI Development Team, Legal Department, Compliance Department, AI Operations Team, Product Management

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes and communication mechanisms to ensure alignment, transparency, and exchange of information regarding AI risk metrics and methodologies used by the organization with those employed by other organizations involved in the AI system lifecycle, including developers, deployers, and operators. This includes establishing mechanisms for AI developers and deployers to communicate and exchange information on potential and use-case specific risks, ensuring such risks are recognized and assessed across the lifecycle. Furthermore, provide meaningful transparency regarding AI systems and their outputs by making appropriate information available to individuals interacting with such systems, with descriptions tailored to individual differences such as the user's role, knowledge, and skill level, and based on the stage of the AI lifecycle. This helps manage risk from lack of explainability and facilitates responsible decision-making to enhance safe operation, including explanations and documentation of AI system risks based on empirical evidence of incidents, and ensure negative residual risks to both downstream acquirers and end users are thoroughly documented. This control is crucial for preventing misalignments or lack of transparency in risk measurement and assessment approaches across organizational boundaries, fostering a common understanding of AI system risks, enhancing overall trustworthiness and accountability, and addressing diverse risk perspectives.

Enhanced Implementation Guide:

This control establishes and maintains robust processes and communication mechanisms to ensure continuous, transparent, and consistent alignment of AI system risk metrics, methodologies, and communication across all internal and external stakeholders involved in the AI system lifecycle, from developers and deployers to operators and end-users. This involves defining and implementing formalized protocols for exchanging use-case specific and potential AI risks, ensuring these risks are recognized, assessed, and proactively addressed throughout the entire lifecycle. Furthermore, it mandates providing meaningful and context-appropriate transparency regarding AI systems and their outputs, tailoring explanations and documentation of AI system risks (including empirical evidence of incidents) based on the user's role, knowledge, skill level, and the stage of the AI lifecycle. The ultimate goal is to prevent misalignments in risk understanding, foster a common and holistic view of AI system risks, enhance overall trustworthiness, accountability, and responsible decision-making, and ensure that negative residual risks are thoroughly documented and communicated to downstream acquirers and end-users.

High-level control implementation guide:

1. **Define Communication Frameworks:** Establish clear communication channels, protocols, and standard templates for AI risk identification, assessment, mitigation, and reporting across all lifecycle stages and stakeholder groups (internal and external).
2. **Harmonize Risk Methodologies:** Develop and mandate the use of unified AI risk taxonomies, metrics, and assessment methodologies (e.g., risk appetite, impact scales, likelihood scoring) that are consistent across the organization and shared with external partners.
3. **Mandate Cross-Functional Engagement:** Institute regular, structured meetings (e.g., AI Risk Working Groups) involving AI Development, Operations, Legal, Compliance, Risk Management, and Product Management, as well as designated points of contact from external AI providers/customers.
4. **Implement Formal Information Exchange Points:** Designate explicit hand-off points and required documentation (e.g., risk registers, residual risk reports, risk mitigation plans) for risk information sharing at each critical stage of the AI lifecycle (design, development, deployment, operation, decommissioning).
5. **Develop Tiered Transparency Guidelines:** Create clear guidelines for providing AI system transparency, differentiating information needs based on target audience (e.g., developers, end-users, regulators) and system criticality, ensuring explanations of AI functionality, outputs, limitations, and empirical risk incidents are readily accessible and comprehensible.
6. **Integrate into Contracts:** Incorporate explicit clauses in contracts with third-party AI developers, deployers, and operators, mandating adherence to the organization's risk communication, metrics alignment, and transparency standards.
7. **Maintain Centralized Documentation:** Ensure all identified risks, mitigation strategies, communication logs, transparency disclosures, and empirical incident data are thoroughly documented, version-controlled, and centrally accessible.

Control frequency: This control is performed **continuously** throughout the AI system lifecycle. Specific structured activities include:

*   Risk communication and documentation at each stage gate of the AI lifecycle (e.g., design review, pre-deployment, post-deployment review).
*   Formal stakeholder alignment meetings: **Quarterly**, or **ad-hoc** as new AI systems are introduced, significant updates occur, or emerging risks are identified.
*   Transparency information updates: **Upon AI system release, significant updates, or identified changes in risk profile.**

Monitoring frequency:
*   **Quarterly** by the AI Governance Committee or relevant management oversight body to review compliance with communication protocols, metric alignment, and transparency efforts.
*   **Annually** by the Internal Audit or Compliance department as part of routine control effectiveness testing.
*   **Continuous** spot checks on documentation and communication logs.

Control type: Semi-automated (Processes are manual but supported by automated tools for documentation management, collaboration, risk register tracking, and potentially automated generation of transparency statements based on predefined templates).

Guide for evaluating control effectiveness:
1. **Documentation Review:** Assess the completeness, accuracy, and consistency of AI risk registers, communication logs, stakeholder meeting minutes, and transparency statements across all relevant teams and with external partners. Verify that residual risks are clearly documented and communicated downstream.
2. **Stakeholder Interviews:** Conduct interviews with key personnel from AI Development, Operations, Legal, Compliance, Risk Management, Product Management, and select external partners to confirm their understanding of communication protocols, shared methodologies, and access to necessary risk information.
3. **Cross-Lifecycle Traceability:** Trace a sample of identified AI risks from initial identification through assessment, mitigation, and residual risk documentation across different lifecycle stages and organizational boundaries, verifying consistent tracking and communication.
4. **Transparency Audit:** Conduct simulated user inquiries or reviews to verify the accessibility, clarity, and comprehensiveness of AI system explanations and risk disclosures for target audiences.
5. **Metrics Alignment Verification:** Compare risk assessment methodologies and metrics used by different internal teams and external providers to confirm harmonization and absence of significant misalignments.
6. **Incident Root Cause Analysis:** Review past AI system incidents or near-misses to identify if communication gaps, lack of transparency, or misaligned risk understanding contributed to

the issue, and if corrective actions addressed these control weaknesses.

Key control indicators (metrics):
* **Percentage of AI systems with a formally documented AI Risk Communication and Transparency Plan.** (Target: 100%)
* **Number of documented instances of AI risk metrics misalignment between internal teams or external partners.** (Target: Zero)
* **Completion rate of mandatory AI risk documentation (e.g., lifecycle risk registers, residual risk reports) at defined stage gates.** (Target: 100%)
* **Average score from stakeholder feedback surveys on clarity and timeliness of AI risk communication.** (Target: High, e.g., >4/5)
* **Number of AI incidents or adverse events directly attributed to communication breakdowns or insufficient transparency.** (Target: Zero)
* **Percentage of AI system users or external stakeholders who rate AI transparency information as "clear" and "sufficient."** (Target: >85%)
* **Frequency of formal risk information exchange meetings/touchpoints with external AI lifecycle partners.** (Adherence to established schedule)

Requirement Statements:

1. Risk metrics or methodologies used by the organization developing the AI system may not align with the risk metrics or methodologies uses by the organization deploying or operating the system. Also, the organization developing the AI system may not be transparent about the risk metrics or methodologies it used. (Page: 10)

2. For example, an AI developer who makes AI software available, such as pre-trained models, can have a different risk perspective than an AI actor who is responsible for deploying that pre-trained model in a specific use case. Such deployers may not recognize that their particular uses could entail risks which differ from those perceived by the initial developer. (Page: 11)

3. Safe operation of AI systems is improved through: • clear information to deployers on responsible use of the system; (Page: 19)

4. Safe operation of AI systems is improved through: • responsible decision-making by deployers and end users; (Page: 19)

5. Safe operation of AI systems is improved through: • explanations and documentation of risks based on empirical evidence of incidents. (Page: 19)

6. Meaningful transparency provides access to appropriate levels of information based on the stage of the AI lifecycle and tailored to the role or knowledge of AI actors or individuals interacting with or using the AI system. (Page: 20)

7. Risk from lack of explainability may be managed by describing how AI systems function, with descriptions tailored to individual differences such as the user’s role, knowledge, and skill level. (Page: 21)

8. Negative residual risks (defined as the sum of all unmitigated risks) to both downstream acquirers of AI systems and end users are documented. (Page: 37)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

5. Third-Party AI Risk Governance, Safeguards, and Ongoing Management

Control Actor: Senior Management, Information Security Team, Risk Management Team, Legal Department, Third-Party Risk Management Team, Vendor Management

Control Types: administrative, technical, preventive, detective, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement robust internal governance structures, policies, procedures, and technical safeguards specifically designed to manage and mitigate risks and maximize benefits that arise from third-party software, data, and other supply chain issues, as well as when customers integrate or use third-party data or systems with the organization's AI products or services. This includes defining clear responsibilities and technical controls to ensure secure and trustworthy AI system operations, particularly in scenarios involving external data sources or systems, and specifically addressing legal and other compliance issues concerning the use of third-party software, hardware systems, and data within AI systems, including risks of infringement of a third-party’s intellectual property or other rights. Furthermore, regularly monitor AI risks and benefits stemming from these third-party resources, applying and documenting appropriate risk controls on an ongoing basis.

Enhanced Implementation Guide:

Control Statement: The organization must establish, implement, and continuously manage a robust third-party AI risk governance framework, including comprehensive policies, procedures, and technical safeguards, to proactively identify, assess, mitigate, and monitor risks (e.g., security, privacy, intellectual property, ethical, compliance) and optimize benefits arising from the integration or use of all third-party software, hardware, data, and services within its AI products and operations, as well as customer-driven integrations. This framework ensures secure, trustworthy, and compliant AI system operations across the entire third-party AI supply chain and usage lifecycle.

High-Level Control Implementation Guide:

1. **Define Governance Structure:** Establish a cross-functional Third-Party AI Risk Governance Committee (comprising Senior Management, InfoSec, Legal, Risk, Procurement, AI Product teams) with clear roles, responsibilities, and accountability for AI-related third-party risks.

2. **Policy & Procedure Development:** Develop and approve a comprehensive Third-Party AI Risk Management Policy and accompanying procedures that cover the entire lifecycle: due diligence, contracting, ongoing monitoring, and offboarding. These should specifically address AI-centric risks like data provenance, model bias, intellectual property rights (including potential infringement from training data or models), data security for AI workloads, ethical AI principles, and compliance with emerging AI regulations.
3. **Risk Assessment & Due Diligence:** Implement a structured process for conducting pre-engagement and periodic risk assessments of all third parties contributing to or interacting with AI systems. This includes assessing their security controls, data handling practices, intellectual property safeguards, ethical AI governance, and compliance with relevant laws and standards.
4. **Contractual Safeguards:** Ensure all contracts with third-party AI vendors, data providers, and system integrators include specific clauses addressing data ownership, intellectual property rights, data security requirements (e.g., encryption, access controls), audit rights, incident response protocols, ethical AI usage, model explainability/transparency requirements where applicable, and liability for AI-related risks.
5. **Technical Safeguards Implementation:** Deploy and enforce technical controls tailored for third-party AI integrations, such as secure API management, data ingress/egress filtering, network segmentation for AI environments, robust access controls, encryption for AI training data and models at rest and in transit, vulnerability management, and continuous security monitoring of integrated third-party components. Implement data provenance tracking for third-party data used in AI models.
6. **Ongoing Monitoring & Performance Management:** Establish mechanisms for continuous monitoring of third-party AI risks, including regular security posture reviews, performance against contractual SLAs, validation of compliance attestations, and tracking of AI-specific risk indicators. Implement a process for reassessing risks based on changes in third-party services, data usage, or regulatory landscapes.
7. **Incident Response & Remediation:** Integrate third-party AI-related incidents (e.g., data breaches, IP infringement claims, model failures due to third-party components) into the organization's broader incident management and remediation processes, ensuring clear communication and collaboration with third parties.

Control Frequency:

* **Policy/Procedure Review:** Annually, or upon significant changes in regulatory landscape, technology, or business model.
* **Risk Assessments (Pre-engagement):** Prior to onboarding any new third-party AI vendor/data source.
* **Risk Assessments (Periodic):** Annually for high-risk third parties; Biennially for moderate-risk.

*   **Contract Review/Updates:** Annually, or upon contract renewal/amendment.
*   **Technical Control Configuration/Review:** Continuous for automated controls; Quarterly for manual review and effectiveness checks.
*   **Ongoing Monitoring:** Continuous for automated data feeds; Quarterly for aggregated risk reporting and performance reviews.

Monitoring Frequency:
*   **Management Review of KCI/KRI:** Quarterly.
*   **Internal Audit/Compliance Review:** Annually.
*   **External Audit/Attestation Review (e.g., SOC 2 reports from third parties):** Annually.

Control Type: Semi-Automated. While governance framework development, policy writing, contract negotiation, and high-level risk assessments are manual, the implementation and monitoring of technical safeguards (e.g., API security, data encryption, network segmentation, automated vulnerability scanning of integrated components, continuous monitoring of third-party security postures via platforms) involve significant automation, with manual oversight and review.

Guide for Evaluating Control Effectiveness:
Effectiveness is evaluated by assessing the existence, implementation, and operating effectiveness of the stated control activities. This includes:
1. **Documentation Review:** Verify the approval and currency of the Third-Party AI Risk Management Policy, procedures, risk assessment templates, and contractual agreements. Confirm that specific AI-related clauses are present and appropriately tailored.
2. **Evidence of Due Diligence:** Review completed third-party AI risk assessments, security questionnaires, and audit reports (e.g., SOC 2, ISO 27001) for a representative sample of third parties. Verify that identified risks have corresponding mitigation plans.
3. **Technical Control Validation:** Review configurations of security tools (e.g., API gateways, network firewalls, data loss prevention systems) to confirm appropriate safeguards for third-party AI data flows. Examine logs for anomalous activities related to third-party integrations.
4. **Interview Stakeholders:** Conduct interviews with Control Actors (Senior Management, InfoSec, Legal, Risk, Vendor Management, AI Development teams) to confirm understanding of responsibilities, adherence to policies, and awareness of AI-specific third-party risks.
5. **Review Incident Reports:** Examine incident reports, particularly those involving third parties or AI-related issues (e.g., data breaches, IP claims, model integrity issues), to assess the effectiveness of preventive and corrective controls.
6. **Review Risk Register:** Verify that identified third-party AI risks are captured, tracked, and remediated in the organizational risk register.

Key Control Indicators (KCIs):

* **Percentage of critical/high-risk third-party AI vendors with completed risk assessments:** Target: 100%.

* **Number of identified third-party AI-related security incidents or data breaches:** Target: Zero.

* **Percentage of new third-party AI contracts including all mandatory AI-specific clauses:** Target: 100%.

* **Time to remediate identified critical/high-severity vulnerabilities in third-party AI integrations:** Target: Within defined SLAs (e.g., <7 days for critical).

* **Number of non-compliance findings from internal or external audits related to third-party AI risk management:** Target: Zero or trending down.

* **Percentage of third-party AI risk reviews completed on schedule:** Target: 95%.

* **Number of intellectual property infringement claims or legal disputes initiated due to third-party AI component usage:** Target: Zero.

Requirement Statements:

1. Risk measurement and management can be complicated by how customers use or integrate third- party data or systems into AI products or services, particularly without sufficient internal governance structures and technical safeguards. (Page: 10)

2. including legal and other issues concerning use of third-party software or hardware systems and data. (Page: 26)

3. Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues. (Page: 29)

4. Policies and procedures are in place that address AI risks associated with third-party entities, including risks of in- fringement of a third-party’s intellectual property or other rights. (Page: 29)

5. AI risks and benefits from third-party resources are regularly monitored, and risk controls are applied and documented. (Page: 37)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

6. AI System Impact Assessment Procedure and Communication

Control Actor: Risk Management Team

Control Types: preventive, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain a comprehensive procedure for conducting AI system impact assessments. This procedure should guide AI actors in understanding and documenting potential positive and negative impacts and harms associated with AI systems within specific contexts throughout their lifecycle, from design and development to deployment, evaluation, and ongoing use. The assessment process should proactively identify and manage AI risks by involving a broad set of diverse perspectives and actors across the AI lifecycle, ensuring that relevant stakeholders, varied affected groups (including non-users), and differential harms are recognized. This procedure must also evaluate risks and opportunities, inform risk mitigation strategies, and include impact measurement approaches that recognize diverse contexts, ensuring that practices are established to appropriately manage risks that emerge within social contexts, including recognizing and considering differences in perceptions of fairness among cultures and applications to enhance AI risk management efforts. Furthermore, the organization's Risk Management Team must ensure that the documented risks and potential impacts of the AI technology are communicated broadly to relevant stakeholders and end users to foster trustworthy AI development and deployment, particularly by informing them about potential negative impacts of interacting with the system. The organization's Risk Management Team must incorporate processes for assessing these potential impacts.

Enhanced Implementation Guide:

The organization, under the leadership of its Risk Management Team, must establish, maintain, and rigorously execute a comprehensive procedure for conducting AI system impact assessments, ensuring the systematic identification, documentation, and evaluation of potential positive and negative impacts and harms of AI systems across their entire lifecycle – from design and development to deployment, evaluation, and ongoing use. This procedure must guide AI actors in understanding and documenting these impacts within specific contexts, actively involving a broad set of diverse perspectives and actors, including relevant stakeholders, varied affected groups (notably non-users), and recognizing differential harms, while also considering differences in perceptions of fairness among cultures to enhance AI risk management efforts. It must proactively identify and manage AI risks and opportunities, inform tailored risk mitigation strategies, and include effective impact measurement approaches. Crucially, the Risk Management Team is responsible for ensuring that all documented risks and potential impacts, particularly negative ones, are broadly and clearly communicated to relevant stakeholders and end-users to foster trustworthy AI development and deployment. High-level implementation involves the Risk Management Team formally documenting a multi-stage AI system impact assessment procedure encompassing scoping, stakeholder identification and engagement, data collection, in-depth impact and risk analysis (including social and cultural dimensions), development of specific mitigation strategies, and a defined communication protocol for both internal and external audiences; this procedure must also embed mechanisms for ongoing assessment of AI systems post-deployment. Control frequency dictates that an initial impact assessment be completed for every new or significantly modified AI system prior to design/development commencement, re-assessments conducted before initial deployment, upon significant contextual changes, or at least annually for deployed systems, and ad-hoc upon detection of new risks or incidents. Monitoring frequency for the control's adherence and effectiveness should be conducted quarterly by an independent compliance function or internal audit, culminating in a comprehensive annual review. This control is primarily manual, relying heavily on expert judgment, collaborative stakeholder engagement, and administrative processes, though semi-automated tools may be leveraged to support data gathering, analysis, and document management aspects. Control effectiveness is evaluated by verifying the existence of an approved, up-to-date AI impact assessment procedure; reviewing completed impact assessment reports for all in-scope AI systems, ensuring they demonstrate comprehensive risk and opportunity identification, robust stakeholder engagement, derivation of actionable mitigation plans, and consideration of diverse contexts; examining communication records to confirm broad and clear dissemination of relevant risk information to stakeholders and end-users; and confirming that identified risks translate into effectively implemented mitigation strategies. Key control indicators include: the percentage of new AI systems with a completed impact assessment prior to development/deployment; the average number of unique stakeholder groups consulted per assessment; the closure rate of identified AI risk mitigation actions; the frequency of AI impact assessment procedure reviews and updates; and feedback scores from stakeholders/users regarding the clarity and usefulness of communicated AI risk information.

Requirement Statements:

1. AI system impact assessment approaches can help AI actors understand potential impacts or harms within specific contexts. (Page: 11)

2. Approaches for measuring impacts on a population work best if they recognize that contexts matter, that harms may affect varied groups or sub-groups differently, and that communities or other sub-groups who may be harmed are not always direct users of a system. (Page: 11)

3. Documenting residual risks will call for the system provider to fully consider the risks of deploying the AI product and will inform end users about potential negative impacts of interacting with the system. (Page: 13)

4. Identifying and managing AI risks and potential impacts – both positive and negative – requires a broad set of perspectives and actors across the AI lifecycle. Ideally, AI actors will represent a diversity of experience, expertise, and backgrounds and comprise demographically and disciplinarily diverse teams. (Page: 14)

5. These actors can: • assist in providing context and understanding potential and actual impacts; (Page: 15)

6. These practices can increase the likelihood that risks arising in social contexts are managed appropriately. (Page: 18)

7. Organizations’ risk management efforts will be enhanced by recognizing and considering these differences. (Page: 22)

8. incorporates processes to assess potential impacts; (Page: 26)

9. Organizational teams document the risks and po- tential impacts of the AI technology they design, develop, deploy, evaluate, and use, and they communicate about the impacts more broadly. (Page: 29)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

7. AI System Accountability and Responsibility Framework

Control Actor: Senior Management

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain a clear framework defining shared responsibilities, accountability mechanisms, roles, and incentive structures across all involved AI actors (including internal departments, external partners, and third-party providers) for the entire lifecycle of AI systems, from design and development to deployment and ongoing operation. This framework, fostered by a strong organizational culture, ensures that trustworthy AI systems are developed and deployed in a manner that is fit for purpose, with clear ownership and oversight of AI-related risks and impacts. Senior management must ensure these responsibilities are clearly communicated, understood, and adhered to by all stakeholders, and that AI actors involved in the design, development, deployment, evaluation, and use of AI systems, across all relevant AI dimensions (Application Context, Data and Input, AI Model, Task and Output), are actively engaged in driving AI risk management efforts. This demonstrates organizational commitment at senior levels to realize effective AI risk management and facilitate necessary cultural change, including fostering collaboration among all AI actors to collectively manage risks and achieve the goals of trustworthy and responsible AI. When seeking accountability for the outcomes of AI systems, the roles of all relevant AI actors should be carefully considered as part of this framework. This framework specifically includes establishing and maintaining accountability for AI systems to foster their trustworthiness, including establishing policies, processes, practices, and procedures for improving organizational accountability efforts related to AI system risks.

Enhanced Implementation Guide:

Senior Management shall establish, maintain, and enforce a comprehensive AI System Accountability and Responsibility Framework, clearly defining roles, shared responsibilities, accountability mechanisms, and incentive structures for all involved AI actors (internal departments, external partners, and third-party providers) throughout the entire AI lifecycle, from design and development to deployment and ongoing operation. This framework must ensure clear ownership and oversight of AI-related risks and impacts across all relevant AI dimensions (Application Context, Data and Input, AI Model, Task and Output), fostering the development and deployment of trustworthy AI systems fit for purpose, and demonstrating organizational commitment at senior levels to realize effective AI risk management. Implementation requires Senior Management to formally approve and communicate the framework via documented policies, procedures, and RACI matrices, integrating it into existing risk management and project methodologies, and implementing mandatory training programs for all AI actors to ensure understanding and adherence. Incentive structures should be developed to promote adherence to responsible AI practices and foster collaboration in risk management. This control should be maintained and formally reviewed by Senior Management annually, or more frequently (e.g., quarterly) if significant changes in AI technology, regulation, or organizational structure occur. Monitoring of the framework's effectiveness should occur on an ongoing basis through internal audits, risk assessments, and project reviews, with formal review by Senior Management or an AI Governance Committee quarterly. This is predominantly a **Manual** control, though elements like training completion tracking or policy dissemination can be **Semi-automated**. Control effectiveness is evaluated by verifying formal Senior Management approval of the framework, assessing the comprehensiveness and clarity of documented policies and RACI matrices, reviewing evidence of widespread communication and mandatory training completion rates, conducting interviews with AI actors to confirm understanding of their roles and accountability, analyzing AI-related incident reports for clear accountability assignment, and examining AI project documentation for consistent application of framework principles. Key Control Indicators (KCIs) include the Policy Adoption Rate (percentage of departments/teams integrating the framework), Training Completion Rate for targeted AI actors, RACI Matrix Coverage (percentage of AI initiatives with documented matrices), Accountability Assignment Rate for AI-related incidents, compliance with scheduled framework reviews, number of audit findings related to AI accountability, and stakeholder survey scores on framework clarity and utility.

Requirement Statements:

1. All involved AI actors share responsibilities for designing, developing, and deploying a trustworthy AI system that is fit for purpose. (Page: 11)

2. Organizations need to establish and maintain the appropriate accountability mechanisms, roles and responsibilities, culture, and incentive structures for risk management to be effective. (Page: 14)

3. Effective risk management is realized through organizational commitment at senior levels and may require cultural change within an organization or industry. (Page: 14)

4. AI actors involved in these dimensions who perform or manage the design, development, deployment, evaluation, and use of AI systems and drive AI risk management efforts are the primary AI RMF audience. (Page: 14)

5. Within the AI RMF, all AI actors work together to manage risks and achieve the goals of trustworthy and responsible AI. (Page: 14)

6. Successful risk management depends upon a sense of collective responsibility among AI actors shown in Figure 3. (Page: 15)

7. Trustworthy AI depends upon accountability. (Page: 20)

8. The role of AI actors should be considered when seeking accountability for the outcomes of AI systems. (Page: 21)

9. established policies, processes, practices, and procedures for improving organiza- tional accountability efforts related to AI system risks; (Page: 24)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

8. AI System Inscrutability Mitigation and Contextual Awareness

Control Actor: AI Development Team

Control Types: preventive, technical, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement processes to identify and address factors contributing to AI system inscrutability, such as limited explainability, interpretability, transparency, or inadequate documentation, throughout the AI system lifecycle. This includes addressing the opaque nature of some AI systems and ensuring sufficient transparency and documentation during development and deployment, covering aspects like design decisions, training data, model structure, intended use, and deployment decisions. The goal is to improve the accuracy and completeness of AI risk measurement by ensuring that the internal workings and outputs of AI systems can be sufficiently understood and assessed, facilitating actionable redress for negative impacts while considering human-AI interaction. This also includes implementing approaches for enhancing contextual awareness throughout the AI lifecycle to further improve understanding of system behavior and impacts, and communicating descriptions of why an AI system made a particular prediction or recommendation to address risks to interpretability.

Enhanced Implementation Guide:

The organization must establish and maintain robust processes to ensure all AI systems demonstrate sufficient explainability, interpretability, and transparency across their entire lifecycle (design, development, testing, deployment, and post-deployment). This includes comprehensive documentation of design decisions, training data characteristics, model architecture, intended use, and deployment parameters. Furthermore, mechanisms must be in place to enhance contextual awareness regarding AI system behavior and impacts, and to effectively communicate the rationale behind AI predictions or recommendations, thereby enabling thorough risk assessment, effective mitigation of negative impacts, and informed human-AI interaction.

Implementation requires the AI Development Team to integrate explainability, interpretability, and transparency considerations into the AI system development lifecycle (AIDLC) from inception. This includes: (a) Defining and documenting the required levels of explainability and interpretability for each AI system based on its risk profile and criticality. (b) Implementing appropriate technical methodologies, such as Explainable AI (XAI) techniques (e.g., LIME, SHAP, causal inference models) and interpretability tools, to enable understanding of model decisions and behavior. (c) Developing and maintaining comprehensive documentation, including AI system design documents, training data provenance and characteristics, model cards detailing purpose and performance, AI impact assessments, and deployment decision logs, ensuring they are accessible and kept current. (d) Establishing clear, understandable processes for communicating the rationale behind AI predictions or recommendations to relevant stakeholders (e.g., end-users, affected individuals, auditors, regulators). (e) Incorporating mechanisms for ongoing contextual awareness throughout the AI lifecycle, such as real-time monitoring of system performance in diverse operational environments, continuous feedback loops from users, and regular re-evaluation of system behavior against expected outcomes and evolving contexts. (f) Ensuring traceability of AI system modifications and their potential impact on inscrutability factors through robust version control and change management. (g) Integrating transparency and explainability requirements into standard development gates, including design reviews, code reviews, testing protocols, and deployment checklists.

Control activities are continuous throughout the AI system lifecycle, with formal reviews, documentation updates, and implementation of new insights occurring at each major lifecycle gate (e.g., design approval, pre-deployment, post-deployment reviews) and upon significant model retraining or updates. Monitoring of this control's effectiveness should be performed at least semi-annually by an independent function (e.g., Internal Audit, AI Governance Committee, or Risk Management), with ad-hoc reviews triggered by significant AI system incidents, material regulatory or policy changes, or critical model updates. This control is primarily semi-automated, leveraging technical tools and frameworks for explainability and contextual monitoring, complemented by essential manual processes for documentation, stakeholder reviews, and communication.

Control effectiveness is evaluated by assessing: (a) The availability, accuracy, and completeness of AI system documentation (e.g., model cards, data sheets, design specifications, intended use, deployment logs) for all in-scope AI systems. (b) Observable evidence of explainability and interpretability mechanisms being implemented and utilized, demonstrated through accessible XAI outputs, interpretability reports, and clear feature importance analyses. (c) Demonstrated understanding of AI system behavior and rationale by relevant stakeholders (e.g., confirmed through internal audit findings, risk assessment outcomes, and structured user feedback sessions). (d) Measurable improvements in the accuracy, completeness, and actionability of AI risk assessments directly attributable to enhanced transparency and understanding. (e) The demonstrated capability to effectively identify, analyze, and implement actionable redress mechanisms for negative impacts arising from AI system behavior, supported by granular insights into the system's decisions. (f) Consistent application and documented review of contextual awareness methods, including monitoring logs and performance metrics across varying operational environments. Key control indicators include: (a) Percentage of high-risk AI systems with comprehensive, up-to-date documentation. (b) Quantitative explainability metrics (e.g., fidelity scores of XAI models, consistency of explanations across similar inputs, or a defined 'explainability index'). (c) Stakeholder comprehension scores (e.g., average scores from structured surveys or interviews gauging understanding of AI decisions by auditors, risk managers, and affected individuals). (d) Mean time to identify the root cause of AI-related incidents or anomalies where inscrutability was a contributing factor. (e) Number or percentage of AI system behaviors that are deemed insufficiently explained or without a clear, attributable cause during routine monitoring or incident response. (f) Rate of compliance with internal policies and standards regarding AI system explainability, transparency, and documentation throughout the AIDLC. (g) Number of actionable insights derived from contextual awareness monitoring leading to model improvements or risk mitigations.

Requirement Statements:

1. Inscrutability: Inscrutable AI systems can complicate risk measurement. Inscrutability can be a result of the opaque nature of AI systems (limited explainability or interpretabil- ity), lack of transparency or documentation in AI system development or deployment, or inherent uncertainties in AI systems. (Page: 11)

2. There are multiple approaches for enhancing contextual awareness in the AI lifecycle. (Page: 18)

3. This characteristic’s scope spans from design decisions and training data to model training, the structure of the model, its intended use cases, and how and when deployment, post-deployment, or end user decisions were made and by whom. Transparency is often necessary for actionable redress related to AI system outputs that are incorrect or otherwise lead to negative impacts. Transparency should consider human-AI interaction. (Page: 20)

4. Risks to interpretability often can be addressed by communicating a description of why an AI system made a particular prediction or recommendation. (Page: 22)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

9. AI Risk Governance: Compliance, Stakeholder Engagement, and Operational Boundaries

Control Actor: AI Governance Committee

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain comprehensive processes for AI risk governance, ensuring the organization's AI risk management practices are continuously aligned with all applicable national and international laws, regulations, industry standards, and ethical norms. This includes regularly reviewing legal and regulatory updates, assessing their impact, and making necessary adjustments. Proactively engage with a broad and diverse range of external AI actors and interested parties (e.g., trade associations, standards developing organizations, researchers, advocacy groups) throughout the AI lifecycle to inform and enhance AI risk management efforts, facilitate contextually sensitive evaluations, and identify both AI system benefits and positive impacts. Furthermore, collaborate with internal and external stakeholders to designate and enforce clear operational boundaries for AI systems, encompassing technical, societal, legal, and ethical aspects, ensuring ongoing compliance and minimizing legal and reputational risks associated with AI system development, deployment, and and use.

Enhanced Implementation Guide:

The AI Risk Governance control ensures the organization's AI risk management is proactively aligned with evolving legal, regulatory, and ethical landscapes through structured governance and stakeholder engagement, while clearly defining and enforcing operational boundaries for AI systems.

Control Statement: Establish, maintain, and continuously enhance a comprehensive AI risk governance framework that systematically incorporates all applicable national and international laws, regulations, industry standards, and ethical norms; proactively engage with diverse external and internal stakeholders throughout the AI lifecycle to inform and strengthen AI risk management efforts; and collaboratively define, document, and enforce clear technical, societal, legal, and ethical operational boundaries for all AI systems to ensure ongoing compliance and mitigate associated legal and reputational risks.

High-Level Control Implementation Guide: The AI Governance Committee, as the control actor, is responsible for this control. Implementation involves establishing a dedicated AI regulatory and ethical intelligence function to continuously monitor and assess new and updated AI-related laws, regulations, industry standards, and ethical guidelines. This includes conducting regular impact analyses on existing AI systems and practices, followed by necessary adjustments to internal policies, procedures, and the overall AI risk management framework. Concurrently, develop and execute a comprehensive stakeholder engagement strategy that identifies, categorizes, and proactively involves a broad spectrum of external AI actors (e.g., trade associations, standards bodies, researchers, advocacy groups) and internal departments (e.g., Legal, Ethics, IT, Business Units) throughout the AI lifecycle. This engagement should facilitate contextually sensitive evaluations, identify both benefits and potential harms, and inform the refinement of AI risk management practices. Furthermore, the Committee, in collaboration with relevant technical and business teams, must define, document, and embed clear operational boundaries for AI systems at the design, development, and deployment stages. These boundaries must encompass technical (e.g., performance thresholds, data privacy, security), societal (e.g., fairness, bias, accessibility), legal (e.g., data residency, IP), and ethical (e.g., explainability, human oversight) aspects. Mechanisms for pre-deployment boundary verification and continuous post-deployment monitoring, including automated checks and manual reviews, must be established and enforced, with clear escalation and remediation pathways for any boundary deviations or breaches.

Control Frequency: Regulatory and ethical monitoring, as well as ongoing stakeholder engagement, should be continuous processes. Formal reviews and updates to the AI governance framework, policies, and operational boundaries by the AI Governance Committee should occur at least quarterly, or immediately upon significant regulatory updates, material changes in AI system deployments, or identified high-risk incidents. AI system-specific boundary definition and enforcement occur throughout the AI system lifecycle, from design to deployment.

Monitoring Frequency: The effectiveness of this control should be monitored monthly by reviewing AI Governance Committee meeting minutes, regulatory impact assessment reports, stakeholder engagement logs, and AI system boundary adherence reports. A comprehensive internal audit or external assessment of the overall AI risk governance framework and its alignment with emerging standards should be conducted at least annually.

Control Type (Manual/Automated/Semi-Automated): Semi-automated. While regulatory scanning and certain technical boundary checks (e.g., performance, data usage) can leverage automated tools, the interpretation of legal/ethical requirements, stakeholder engagement, impact assessment, policy formulation, and enforcement of societal/ethical boundaries remain predominantly manual, requiring expert judgment and committee deliberation.

Guide for Evaluating Control Effectiveness: To evaluate effectiveness, verify the existence, approval, and regular updates of the AI Governance Framework, policies, and procedures. Confirm that a defined process for monitoring legal, regulatory, and ethical updates exists, and audit logs to ensure impact assessments are conducted and policies are adjusted in a timely manner. Review records of stakeholder engagement activities, including meeting minutes, feedback mechanisms, and evidence of how stakeholder input has informed AI risk management decisions. Sample AI system documentation (e.g., design documents, risk assessments) to confirm the clear definition and incorporation of technical, societal, legal, and ethical operational boundaries. Assess incident reports and audit trails related to AI system performance and compliance to verify that boundary violations are detected, escalated, and remediated effectively. Interview key personnel from the AI Governance Committee, Legal, and technical teams to confirm understanding and adherence to the governance framework.

Key Control Indicators:
* **# of regulatory/ethical updates identified, assessed, and incorporated into policies/procedures per quarter.**
* **% of critical AI systems with documented, approved, and clearly defined operational boundaries (technical, societal, legal, ethical).**
* **# of unique external stakeholders engaged per quarter, and evidence of their feedback informing AI risk management.**
* **% of AI risk assessments that incorporate stakeholder feedback and address defined operational boundaries.**
* **Average time from identification of a new AI regulation/ethical norm to policy update completion.**
* **# of AI system incidents or near-misses directly attributed to undefined or unenforced operational boundaries.**
* **Audit finding rate related to AI governance framework compliance or boundary enforcement.**

Requirement Statements:

1. The Framework is intended to be flexible and to augment existing risk practices which should align with applicable laws, regulations, and norms. (Page: 12)

2. The AI actors in this dimension comprise a separate AI RMF audience who informs the primary audience. These AI actors may in- clude trade associations, standards developing organizations, researchers, advocacy groups, Page 9 (Page: 14)

3. These actors can: • designate boundaries for AI operation (technical, societal, legal, and ethical); (Page: 15)

4. When properly resourced, increasing the breadth and diversity of input from interested parties and relevant AI actors throughout the AI lifecycle can en- hance opportunities for informing contextually sensitive evaluations, and for identifying AI system benefits and positive impacts. (Page: 18)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

10. AI Risk Governance: Compliance with External Criteria

Control Actor: Risk Management Team

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement processes to ensure that AI risk criteria, tolerance levels, and response strategies are defined and consistently applied in accordance with established organizational, domain-specific, discipline-specific, sector-specific, or professional requirements and guidelines. This includes utilizing external stakeholders as a source of formal or quasi-formal norms and guidance for AI risk management, and specifically, developing AI safety risk management approaches by drawing inspiration from safety guidelines in established fields (e.g., transportation, healthcare) to align with existing sector- or application-specific guidelines or standards.

Enhanced Implementation Guide:

Enhanced Control Description:

This control, "AI Risk Governance: Compliance with External Criteria," mandates the establishment and ongoing operation of robust processes to define, apply, and continually align AI risk criteria, tolerance levels, and response strategies with authoritative external requirements. This includes drawing upon organizational, domain-specific, discipline-specific, sector-specific, and professional guidelines and standards (e.g., ISO/IEC 42001, NIST AI RMF, sector-specific safety standards from healthcare or transportation) to ensure comprehensive and contextualized AI risk management.

*   **Control Statement:** Ensure all AI risk criteria, tolerance levels, and response strategies are explicitly defined and consistently applied in alignment with relevant internal and external industry standards, regulatory requirements, and established professional guidelines for AI safety and governance.
*   **High-Level Control Implementation Guide:** The Risk Management Team, in collaboration with legal, compliance, and relevant business units, must first identify and catalog all applicable external AI-related standards, guidelines, and regulatory requirements (e.g., data protection laws, industry-specific safety norms, ethical AI principles). Subsequently, internal AI

risk management policies, frameworks, and procedures (including risk assessment methodologies, tolerance definitions, and incident response plans) must be developed or updated to explicitly reference and incorporate these external criteria. A formal process for ongoing monitoring of new or updated external guidelines must be established, along with a structured review cycle for internal documentation to maintain alignment. Training on these policies and the external requirements should be provided to relevant personnel.

* **Control Frequency:** Ongoing application, with a formalized review and update cycle for policies and frameworks performed at least Annually, or more frequently if triggered by significant changes in external regulations, organizational strategy, or AI system deployments.
* **Monitoring Frequency:** Quarterly, through internal audits and management reviews of AI risk governance documentation and activities.
* **Control Type:** Predominantly Manual, involving policy development, strategic reviews, and expert judgment, with Semi-automated elements possible for regulatory scanning and document version control.
* **Guide for Evaluating Control Effectiveness:** Effectiveness is evaluated by verifying the existence and currency of documented AI risk policies and frameworks that explicitly reference and integrate external standards. This includes reviewing AI system risk assessments and documentation to confirm that defined risk criteria, tolerance levels, and response strategies align with these external guidelines. Evidence of regular policy reviews, updates, and approval by relevant governance bodies (e.g., AI Governance Committee, Risk Committee) must be available. Furthermore, the effectiveness of the control is demonstrated by the absence of non-compliance findings related to external AI risk criteria and by successful resolution of any identified AI safety or governance incidents.
* **Key Control Indicators (KCIs):**
    * Percentage of AI risk policies and frameworks updated to reflect new/amended external guidelines within 90 days of publication.
    * Number of external AI governance standards explicitly referenced and integrated into internal AI risk management documentation.
    * Completion rate of mandatory AI risk governance training for relevant stakeholders.
    * Number of significant AI risk incidents or compliance deviations directly attributable to a failure in aligning with external criteria.
    * Maturity score progression against recognized AI governance frameworks (e.g., NIST AI RMF, ISO 42001) as assessed through internal audits or external benchmarks.
    * Percentage of new AI initiatives undergoing a risk assessment that clearly articulates alignment with sector-specific safety guidelines.

Requirement Statements:

1. Organizations should follow existing regulations and guidelines for risk criteria, tolerance, and response established by organizational, domain, discipline, sector, or professional requirements. (Page: 12)

2. These actors can: • be a source of formal or quasi-formal norms and guidance for AI risk management; (Page: 15)

3. AI safety risk management approaches should take cues from efforts and guidelines for safety in fields such as transportation and healthcare, and align with existing sector- or application-specific guidelines or standards. (Page: 20)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

11. Strategic AI System Commissioning and Deployment Decision Framework

Control Actor: Senior Management, AI Governance Committee, AI Project Management

Control Types: administrative, preventive, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish a comprehensive process for Senior Management and the AI Governance Committee to strategically assess and determine whether AI technology is an appropriate or necessary tool for a given context or purpose, including explicit processes for making go/no-go decisions regarding AI system commissioning and deployment, leveraging the contextual knowledge gained from the completed MAP function. This decision to commission or deploy an AI system must be based on a contextual assessment of its trustworthiness characteristics, relative risks, impacts, costs, and benefits, informed by a broad set of interested parties. This assessment must also account for resources required to manage AI risks, including viable non-AI alternative systems, approaches, or methods, to reduce the magnitude or likelihood of potential impacts. This process must also define how AI systems will be used responsibly, including establishing reasonable AI risk tolerance levels, especially where established organizational or sector-specific guidelines are absent. Furthermore, this process must include clear criteria and procedures for Senior Management to mandate the safe cessation of AI system development and deployment when unacceptable negative risk levels are identified, such as imminent significant negative impacts, severe harms occurring, or catastrophic risks. This ensures the organization has clear boundaries for acceptable AI risk, predefined critical response mechanisms, and a foundational strategic assessment to manage and mitigate severe AI-related harms, fostering responsible AI system development and use from inception. This process specifically involves carrying out the MAP function to gather essential information that enables negative risk prevention and informs critical decisions, including those related to model management and the initial determination of an AI solution's appropriateness or necessity. This framework also includes a determination by AI Project Management and Senior Management as to whether the AI system achieves its intended purposes and stated objectives before its development or deployment proceeds.

Enhanced Implementation Guide:

This control mandates a robust, auditable strategic decision-making framework, spearheaded by Senior Management and the AI Governance Committee, to determine the appropriateness and necessity of commissioning or deploying any AI system. It requires a comprehensive pre-development/pre-deployment assessment rooted in the contextual knowledge gained from the completed MAP (Measurement, Assessment, and Prevention) function, ensuring an informed evaluation of the AI system's trustworthiness characteristics, relative risks (including ethical, privacy, security, operational, reputational), potential impacts, associated costs, and anticipated benefits. A critical component is the explicit consideration of viable non-AI alternative systems, approaches, or methods to achieve the intended objectives, along with a thorough analysis of resources required to manage AI risks. The framework must define clear organizational AI risk tolerance levels, especially where sector-specific guidelines are absent, and establish explicit guidelines for responsible AI system use. Go/no-go decisions must be based on a broad set of interested parties' input and formally documented with clear rationale. Crucially, this control establishes clear criteria and procedures for Senior Management to mandate the safe cessation of AI system development or deployment when unacceptable negative risk levels are identified, such as imminent significant negative impacts, severe harms occurring, or catastrophic risks, ensuring predefined critical response mechanisms. Furthermore, AI Project Management and Senior Management must formally determine whether the AI system achieves its intended purposes and stated objectives before development or deployment proceeds.

High-level Control Implementation Guide: Establish a formal AI Decision Framework policy and associated procedures, clearly outlining decision points, required documentation, and roles/responsibilities for Senior Management, AI Governance Committee, and AI Project Management. Mandate the completion and review of the MAP function outputs as a prerequisite for all strategic AI decisions. Develop standardized templates and criteria for comprehensive assessments of AI trustworthiness, risks, impacts, costs, benefits, and the evaluation of non-AI alternatives. Define and publish organizational AI risk tolerance levels and principles for responsible AI use. Integrate formal sign-off gates at key stages (e.g., concept approval, design approval, deployment approval) requiring explicit go/no-go decisions based on the conducted assessments. Develop and communicate detailed procedures for the safe cessation of AI development/deployment, including triggers, escalation paths, and operational shutdown protocols. Ensure the framework includes a formal checkpoint where AI Project Management and Senior Management verify alignment with intended purposes and objectives before proceeding.

Control Frequency: This control must be performed for every new AI system proposed for commissioning or deployment, and for significant changes or re-evaluations of existing AI systems that alter their risk profile or intended use. The cessation criteria assessment is continuous and triggered ad-hoc whenever unacceptable risk levels are identified.

Monitoring Frequency: The effectiveness of the overall AI Decision Framework (policies, procedures, criteria, roles) should be reviewed by the AI Governance Committee or Internal Audit on a quarterly or bi-annual basis. Monitoring for unacceptable risk levels that could trigger cessation should be continuous as part of ongoing AI risk management and performance oversight.

Control Type: Manual with Semi-Automated Elements. The core strategic decision-making and assessment require human judgment and deliberation. However, supporting processes like data collection (e.g., MAP function outputs), documentation, workflow management, and tracking of decisions can be facilitated by semi-automated tools (e.g., governance platforms, risk management systems).

Guide for Evaluating Control Effectiveness: Verify through documentation review that all AI commissioning/deployment requests have gone through the established framework, with explicit go/no-go decisions, rationale, and approvals. Confirm the consistent utilization of MAP function insights in decision-making. Assess the quality and completeness of risk, impact, cost, and benefit assessments, including the rigor of non-AI alternative evaluations. Validate that approved AI systems adhere to defined AI risk tolerance levels. Audit whether relevant stakeholders (legal, compliance, ethics, business units) were consistently engaged. Review the clarity and readiness of cessation procedures and, if applicable, evaluate the effectiveness and promptness of any triggered cessation events. Confirm that commissioned/deployed AI systems demonstrably align with their stated purposes and objectives.

Key Control Indicators:
*   **Percentage of AI Initiatives Subjected to the Decision Framework:** (Target: 100%)
*   **Number of Documented AI Go/No-Go Decisions per Reporting Period:** (Metric to track process adherence)
*   **Average Time from AI Proposal Submission to Final Strategic Decision:** (Metric for efficiency, target TBD based on complexity)
*   **Number of AI Initiatives Redesigned or Rejected Due to Risk Assessment Findings:** (Indicates effectiveness of risk gating)
*   **Number of Instances Where AI System Cessation Protocol Was Triggered or Considered:** (Indicates proactive risk management)
*   **Percentage of AI Initiatives with Documented Non-AI Alternative Analysis:** (Target: 100%)

*   **Number of Post-Deployment AI Incidents Directly Attributable to Inadequate Pre-Commissioning/Deployment Assessment:** (Target: Low, indicates predictive power)
*   **Percentage of Deployed AI Systems Achieving Stated Objectives (as verified post-deployment):** (Measures accuracy of initial purpose alignment check)

Requirement Statements:

1. Where established guidelines do not exist, organizations should define reasonable risk tolerance. (Page: 12)

2. In cases where an AI system presents unacceptable negative risk levels – such as where significant negative impacts are imminent, severe harms are actually occurring, or catastrophic risks are present – development and deployment should cease in a safe manner until risks can be sufficiently managed. (Page: 13)

3. It is the joint responsibility of all AI ac- tors to determine whether AI technology is an appropriate or necessary tool for a given context or purpose, and how to use it responsibly. (Page: 18)

4. The decision to commission or deploy an AI system should be based on a contextual assessment of trustworthi- ness characteristics and the relative risks, impacts, costs, and benefits, and informed by a broad set of interested parties. (Page: 18)

5. explicit processes for making go/no-go system commissioning and deployment deci- sions; (Page: 24)

6. The information gathered while carrying out the MAP function enables negative risk prevention and informs decisions for processes such as model management, as well as an initial decision about appropriateness or the need for an AI solution. (Page: 30)

7. After completing the MAP function, Framework users should have sufficient contextual knowledge about AI system impacts to inform an initial go/no-go decision about whether to design, develop, or deploy an AI system. (Page: 30)

8. A determination is made as to whether the AI system achieves its intended purposes and stated objectives and whether its development or deployment should proceed. (Page: 37)

9. Resources required to manage AI risks are taken into account – along with viable non-AI alternative systems, ap- proaches, or methods – to reduce the magnitude or likelihood of potential impacts. (Page: 37)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

12. AI System Trustworthiness Assessment and Application Guidelines

Control Actor: Risk Management Team

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain clear, actionable guidelines and procedures for comprehensively assessing and applying the characteristics of trustworthy AI systems developed or deployed by the organization. These guidelines should provide a structured approach for identifying, evaluating, and embedding various trustworthiness characteristics—including validity and reliability, safety, security and resilience, accountability and transparency, explainability and interpretability, privacy enhancement, and fairness with harmful bias managed—throughout the AI system lifecycle. This specifically includes identifying and managing new risks to privacy presented by AI systems, such as those allowing inference to identify individuals or previously private information. When managing AI risks, ensure these guidelines enable the Risk Management Team to make difficult decisions by carefully considering and balancing these various trustworthiness characteristics, acknowledging and addressing potential tradeoffs (e.g., between interpretability and privacy, or predictive accuracy and interpretability). The aim is to ensure that systems meet defined organizational standards and regulatory requirements, thereby reducing negative AI risks and potential negative consequences before and during operational use. This comprehensive approach to risk management specifically includes balancing tradeoffs among the trustworthiness characteristics (e.g., security, fairness, accuracy, interpretability, privacy).

Enhanced Implementation Guide:

The organization shall establish, implement, and continuously maintain comprehensive, actionable, and auditable guidelines and procedures for assessing, embedding, and iteratively managing the characteristics of trustworthy AI systems throughout their entire lifecycle, encompassing development, deployment, and operational use. This control mandates a structured approach to integrate validity and reliability, safety, security and resilience, accountability and transparency, explainability and interpretability, privacy enhancement (specifically identifying and mitigating new risks like inference to identify individuals or private information), and fairness with harmful bias managed. A critical aspect is to empower the Risk Management Team to navigate and document difficult decisions by carefully considering and balancing potential tradeoffs among these trustworthiness characteristics (e.g., between interpretability and privacy, or predictive accuracy and interpretability), ensuring AI systems consistently meet defined organizational standards and evolving regulatory requirements, thereby proactively mitigating negative AI risks and potential adverse consequences.

High-level Control Implementation Guide:

1. **Policy & Guideline Development:** Formalize and document the 'AI System Trustworthiness Guidelines' covering each characteristic (validity, safety, security, transparency, explainability, privacy, fairness) and a structured process for risk assessment and tradeoff management.
2. **Lifecycle Integration:** Mandate specific checkpoints and required evidence (e.g., Trustworthiness Impact Assessments, AI Privacy Impact Assessments, Bias Audits, Security Reviews) at each stage of the AI system lifecycle (concept, design, development, testing, deployment, monitoring, decommission).
3. **Risk Management Framework Integration:** Embed AI-specific risk assessment methodologies and a risk appetite framework within the existing enterprise risk management process, ensuring identified AI risks are evaluated, prioritized, and mitigated.
4. **Tradeoff Decision Protocol:** Establish a clear protocol for documenting and approving decisions where tradeoffs between trustworthiness characteristics are necessary, ensuring justification, stakeholder consultation, and residual risk acceptance are formally recorded.
5. **Training & Awareness:** Implement mandatory training programs for all relevant personnel (AI developers, product owners, legal, privacy, risk management, and operational teams) on the AI trustworthiness guidelines and their specific roles in their application.
6. **Tooling & Automation:** Identify and leverage appropriate tools (e.g., for bias detection, explainability, security scanning, privacy-enhancing technologies) to support the systematic application of the guidelines.
7. **Documentation & Evidence:** Require comprehensive documentation for all AI system assessments, design decisions, risk mitigations, and tradeoff resolutions, establishing an auditable trail.
8. **Continuous Improvement:** Define a mechanism for periodic review and update of the guidelines based on internal lessons learned, technological advancements, and evolving regulatory landscape.

Control Frequency:

- **Guidelines Establishment/Review:** Annually, or upon significant regulatory changes or major AI strategy shifts.
- **Application:** Per AI system, at defined gates throughout its lifecycle (e.g., pre-design, post-development, pre-deployment, ongoing monitoring).

Monitoring Frequency:
* **Internal Audit/Compliance Review:** Bi-annually or Annually.
* **Risk Management Oversight:** Continuous for critical AI systems; Quarterly for portfolio review of all AI systems.

Control Type: Predominantly **Manual** with increasing elements of **Semi-automated** support.
* **Manual:** Policy drafting, risk assessment decisions, tradeoff approvals, interpretative analysis, stakeholder consultations, training delivery, incident response planning.
* **Semi-automated:** The use of specialized tools for bias detection, explainability generation, data privacy assessment, and security vulnerability scanning provides inputs and supports the manual decision-making and assessment processes.

Guide for Evaluating Control Effectiveness:
1. **Documentation Completeness & Quality:** Verify the existence, comprehensiveness, and regular update of the 'AI System Trustworthiness Guidelines.' Assess the quality and rigor of documented trustworthiness assessments, AI-PIAs, bias assessments, and tradeoff justifications for a representative sample of AI systems.
2. **Process Adherence:** Sample AI systems across different lifecycle stages to confirm that the defined guidelines and procedures were consistently followed, including evidence of checkpoints, approvals, and required artifacts (e.g., risk registers, mitigation plans).
3. **Stakeholder Understanding & Application:** Conduct interviews with key personnel (AI developers, product owners, risk managers) to assess their understanding of the guidelines and their practical application in their roles.
4. **Issue & Incident Analysis:** Review AI-related incidents or near-misses (e.g., privacy breaches, fairness concerns, model failures) to determine if control weaknesses contributed and if the guidelines were effective in prevention or response.
5. **Regulatory Alignment:** Confirm the guidelines and their application align with current and anticipated AI-related regulations (e.g., GDPR, NIST AI RMF, upcoming AI Acts).
6. **Effectiveness of Tradeoff Management:** Evaluate whether the documented tradeoff decisions demonstrably balance competing objectives effectively and if subsequent monitoring confirms the accepted risk levels.

Key Control Indicators (KCIs):
* **Percentage of New AI Systems with Completed Trustworthiness Impact Assessments (TIAs) and AI-PIAs:** Target 100%.
* **Number of Documented and Approved Tradeoff Decisions:** Indicating proactive management of complex AI characteristics.

* **Guideline Review and Update Timeliness:** (e.g., last update within 12 months for foundational guidelines).
* **Mean Time to Remediate Identified AI Trustworthiness Risks:** Measuring the efficiency of risk mitigation.
* **Number of AI-related Incidents or Regulatory Non-compliance Findings:** Specifically those preventable by adherence to trustworthiness guidelines.
* **Training Completion Rate for Relevant AI Stakeholders:** Target >95%.
* **Audit Findings per AI System Review:** Number of material deficiencies related to AI trustworthiness controls.
* **Coverage of Automated/Semi-Automated Trustworthiness Tools:** Percentage of AI systems using approved tools for bias detection, explainability, or privacy risk assessment.

Requirement Statements:

1. Actionable risk management efforts lay out clear guidelines for assessing trustworthiness of each AI system an organization develops or deploys. (Page: 12)

2. For AI systems to be trustworthy, they often need to be responsive to a multiplicity of criteria that are of value to interested parties. Approaches which enhance AI trustworthiness can reduce negative AI risks. This Framework articulates the following characteristics of trustworthy AI and offers guidance for addressing them. Characteristics of trustworthy AI systems include: valid and reliable, safe, secure and resilient, accountable and trans- parent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed. Neglecting these characteristics can increase the probability and magnitude of negative consequences. (Page: 17)

3. Human judgment should be employed when deciding on the specific metrics related to AI trustworthiness characteristics and the precise threshold values for those metrics. (Page: 17)

4. When managing AI risks, organizations can face difficult decisions in balancing these characteristics. For example, in certain scenarios tradeoffs may emerge between optimizing for interpretability and achieving privacy. In other cases, organizations might face a tradeoff between predictive accuracy and interpretability. Or, under certain conditions such as data sparsity, privacy-enhancing techniques can result in a loss in accuracy, affecting decisions (Page: 17)

5. A comprehensive approach to risk management calls for balancing tradeoffs among the trustworthiness characteristics. (Page: 18)

6. AI systems can also present new risks to privacy by allowing inference to identify individuals or previously private information about individuals. (Page: 22)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

13. Senior Management AI Risk Governance, Core RMF Process Enhancement, and Documentation

Control Actor: Senior Management, AI Governance Committee, Risk Management Team

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for Senior Management to prioritize organizational policies, resources, and investments based on the assessed risk level and potential impact of AI systems, and to urgently apply thorough risk management processes for safety risks. Senior Management must also maintain overarching organizational practices and governing structures for harm reduction, encompassing comprehensive risk management strategies. Crucially, AI risk governance must be established as a cross-cutting function, designed to inform and be infused throughout the other AI Risk Management Framework functions (MAP, MEASURE, and MANAGE). These processes should be continuously enhanced, and mechanisms established for their continual improvement, for governing, mapping, measuring, and managing AI risk, including developing action plans to address identified gaps and prioritizing their mitigation based on organizational needs and risk management processes, with clear documentation of outcomes, ensuring flexibility and adaptation of the AI RMF functions based on available resources and capabilities. This includes ensuring a robust governance structure is in place to oversee and manage AI risks across the organization. Furthermore, as part of the MANAGE function, allocate risk resources to mapped and measured risks on a regular basis as defined by the GOVERN function.

Enhanced Implementation Guide:

The enhanced control description for "Senior Management AI Risk Governance, Core RMF Process Enhancement, and Documentation" is as follows:

Control Statement: Senior Management, supported by the AI Governance Committee and Risk Management Team, must establish, maintain, and continuously enhance a robust, cross-cutting AI Risk Governance framework. This framework will prioritize organizational policies, resources, and investments based on assessed AI risk levels and potential impact, with an urgent focus on AI safety risks. It must ensure comprehensive harm reduction strategies, inform and integrate with all AI Risk Management Framework (RMF) functions (MAP, MEASURE, MANAGE), and mandate the regular allocation of resources to address identified AI risks, all thoroughly documented and adaptable to organizational capabilities.

High-Level Control Implementation Guide: Establish a formally chartered AI Governance Committee comprising senior leadership. Define and regularly communicate the organization's AI risk appetite and develop a comprehensive AI risk management strategy, explicitly including safety risk protocols. Implement processes that ensure the AI Governance framework consistently informs and integrates with the mapping (identification), measuring (assessment), and managing (mitigation) of AI risks throughout the AI system lifecycle. Develop and enforce clear policies, procedures, and guidelines for AI risk assessment, prioritization, resource allocation, and incident response. Crucially, establish a structured mechanism for Senior Management to review mapped and measured AI risks, and formally allocate appropriate financial, human, and technological resources for their mitigation and ongoing management, including detailed action plans for identified gaps. Mandate consistent documentation of all AI risk assessments, mitigation strategies, decisions, resource allocations, and outcomes. Implement a continuous improvement cycle through regular reviews (e.g., annual strategic reviews, post-incident analyses) to assess the AI RMF's effectiveness and adapt it based on emerging AI technologies, evolving threats, and organizational learning.

Control Frequency: The core governance framework is ongoing; strategic reviews of the AI RMF and risk appetite occur annually; resource allocation reviews are performed quarterly or bi-annually, depending on AI initiative criticality; policy and procedure updates occur as needed, but at least annually.

Monitoring Frequency: The AI Governance Committee and Risk Management Team perform quarterly reviews of implementation progress, policy adherence, and mitigation action effectiveness; an independent review or internal audit of the overall AI RMF and governance structure is conducted annually.

Control Type: This control is primarily **Manual** due to the inherent human judgment required for strategic decision-making, policy formulation, and oversight, but it incorporates **Semi-automated** elements for risk tracking, reporting, and workflow management (e.g., using GRC platforms or dedicated risk management software).

Guide for Evaluating Control Effectiveness: Evaluate control effectiveness by verifying demonstrable adherence of AI-related projects to established AI risk governance policies. Assess if documented risk assessments consistently lead to prioritized mitigation actions and appropriate resource allocation, particularly for safety risks, with clear evidence of action plan implementation and gap closure. Confirm that the GOVERN function demonstrably influences and integrates with the MAP, MEASURE, and MANAGE functions, evidenced by consistent risk identification, assessment, and management practices across the organization. Review the completeness, accuracy, and accessibility of documentation related to AI risk governance decisions, risk registers, mitigation plans, and resource allocations. Evaluate the active participation and documented contributions of Senior Management, the AI Governance Committee, and the Risk Management Team in AI risk oversight. Look for clear evidence of regular reviews of the AI RMF, identified improvement opportunities, and implemented enhancements. Ultimately, assess whether the governance framework contributes to a demonstrable reduction in AI-related harms or incidents.

Key Control Indicators:
* **Number of AI-related Policies/Procedures Reviewed/Updated Annually:** To reflect proactivity in governance.
* **Percentage of New AI Systems with Documented Risk Assessments and Mitigation Plans:** To indicate coverage and adherence.
* **Average Time to Remediate High-Priority AI Risks:** To measure responsiveness.
* **Number of AI Safety Incidents/Near Misses:** To directly measure harm reduction (expected to decrease or remain low).
* **Compliance Score against Internal AI Governance Framework:** Based on internal audit or self-assessment.
* **Resource Allocation vs. Identified High-Priority Risks:** The percentage of allocated resources against total high-priority risk mitigation needs.
* **Frequency of AI Governance Committee Meetings and Average Attendance Rate:** To indicate engagement and oversight consistency.
* **Percentage of AI RMF Gaps Identified and Closed:** To measure continuous improvement.
* **Evidence of Cross-Functional RMF Integration:** Traceability of identified risks from mapping to mitigation actions.

Requirement Statements:

1. Policies and resources should be prioritized based on the assessed risk level and potential impact of an AI system. (Page: 12)

2. Safety risks that pose a potential risk of serious injury or death call for the most urgent prioritization and most thorough risk management process. (Page: 19)

3. Maintaining organizational practices and governing structures for harm reduction, like risk management, can help lead to more accountable systems. (Page: 21)

4. enhanced processes for governing, mapping, measuring, and managing AI risk, and clearly documenting outcomes; (Page: 24)

5. Governance is designed to be a cross-cutting function to inform and be infused throughout the other three functions. (Page: 25)

6. Framework users may apply these functions as best suits their needs for managing AI risks based on their resources and capabilities. (Page: 26)

7. Assuming a governance structure is in place, functions may be performed in any order across the AI lifecycle as deemed to add value by a user of the framework. (Page: 26)

8. The MANAGE function entails allocating risk resources to mapped and measured risks on a regular basis and as defined by the GOVERN function. (Page: 36)

9. ...along with mechanisms for continual improvement. (Page: 36)

10. Action plans can be developed to address these gaps to fulfill outcomes in a given category or subcategory. Prioritization of gap mitigation is driven by the user’s needs and risk management processes. (Page: 38)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

14. High-Priority AI Risk Management, Prioritization, and Response Planning

Control Actor: Risk Management Team, AI Governance Body

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement processes to identify, regularly assess, prioritize, tailor, and develop planned responses to AI risks based on their context of use. This includes specifically addressing high-priority risks by considering options such as mitigating, transferring, avoiding, or accepting them. This ensures that the most significant AI risks receive urgent attention and customized, comprehensive management efforts, aligning resources with critical areas of concern within the organization's AI risk management framework. The planning and documentation of these responses are crucial for effective risk management.

Enhanced Implementation Guide:

This control mandates the establishment and ongoing operation of robust processes to identify, assess, prioritize, and manage high-priority AI risks across all organizational AI systems and those provided by third parties. The core objective is to ensure that the most significant AI risks receive immediate and tailored attention, aligning organizational resources to either mitigate, transfer, avoid, or formally accept these risks through comprehensive, documented response plans.

Control Statement: The organization must maintain an effective and auditable framework for identifying, continuously assessing, prioritizing, and promptly developing and executing documented response plans for all high-priority AI risks, ensuring timely and appropriate risk treatment aligned with organizational risk appetite.

High-Level Control Implementation Guide:

1. **Define High-Priority AI Risks:** Establish clear, quantifiable criteria for classifying AI risks as "high-priority" based on potential impact (e.g., financial, reputational, ethical, safety, regulatory non-compliance) and likelihood.
2. **Formalize AI Risk Assessment Process:** Implement a structured methodology for regular and ad-hoc identification and assessment of AI risks across all AI systems/use cases (internal

and third-party), leveraging a standardized risk register.

3. **Implement Prioritization Mechanism:** Utilize a robust scoring or ranking system to consistently prioritize identified AI risks, flagging those that meet the defined high-priority criteria for immediate attention.
4. **Develop Tailored Response Plans:** For each identified high-priority AI risk, formulate specific, actionable, and documented response plans outlining the chosen treatment strategy (mitigate, transfer, avoid, accept), specific actions, responsible parties, timelines, and required resources.
5. **Secure Approvals and Resource Allocation:** Ensure all high-priority AI risk response plans are reviewed and approved by the AI Governance Body or delegated authority, with confirmation of necessary resource allocation.
6. **Monitor and Report Progress:** Establish mechanisms to track the execution status of response plans, ensuring timely completion and providing regular updates to relevant stakeholders and governance bodies.
7. **Maintain Audit Trail:** Keep a comprehensive, centralized, and auditable record of all identified AI risks, their prioritization, associated response plans, execution status, and any related documentation.

Control Frequency:

* **Risk Identification & Assessment:** At least semi-annually, or immediately upon significant changes to AI systems, new AI deployments, or evolving regulatory/threat landscapes.
* **High-Priority Risk Response Planning:** Within 10 business days of a risk being identified and prioritized as high-priority.
* **Response Plan Review & Update:** Quarterly for active high-priority risks, or as dictated by the risk's velocity or severity.

Monitoring Frequency:

* **AI Governance Body Oversight:** Quarterly review of the AI risk register, high-priority risk status, and effectiveness of response plans.
* **Internal Audit/Compliance Review:** Annually.
* **Operational Management Review:** Monthly review of AI risk dashboards by the Risk Management Team.

Control Type: Semi-automated (manual judgment and decision-making supported by automated tools for risk register management, workflow approvals, and potentially data aggregation/reporting).

Guide for Evaluating Control Effectiveness:

1. **Documentation Review:** Verify the existence and adherence to documented policies, procedures, and a formal framework for AI risk management, specifically for high-priority risks.
2. **Risk Register Completeness:** Review a sample of AI systems and use cases to confirm that all significant AI risks, particularly high-priority ones, have been identified, assessed, and recorded in the risk register.
3. **Adequacy of Response Plans:** Assess a sample of high-priority AI risk entries to confirm that corresponding response plans are specific, actionable, assigned, budgeted, and align with the chosen risk treatment strategy.
4. **Evidence of Execution & Follow-up:** Examine evidence of action execution, timelines adherence, and documented outcomes for a selection of high-priority risk response plans. Verify that follow-up activities (e.g., post-implementation reviews) occur as planned.
5. **Timeliness:** Confirm that high-priority risks are identified, prioritized, and have response plans initiated within defined timeframes.
6. **Governance & Oversight:** Validate that the AI Governance Body or delegated authority regularly reviews and challenges AI risk assessments, prioritization decisions, and the status of high-priority risk response plans.

Key Control Indicators (Metrics):

- **Percentage of identified high-priority AI risks with approved, documented response plans:** Target: 100%.
- **Average time to develop and approve a response plan for a newly identified high-priority AI risk:** Target: < 10 business days.
- **Percentage of high-priority AI risk response actions completed on time:** Target: > 95%.
- **Number of overdue high-priority AI risk response actions:** Target: 0.
- **Number of critical AI incidents directly attributable to unmanaged or mismanaged high-priority risks:** Target: 0.
- **Frequency of AI Governance Body review of high-priority AI risk register:** Target: Quarterly, minimum.
- **Maturity level score of the AI risk management process:** Based on internal or external assessment.

Requirement Statements:

1. When applying the AI RMF, risks which the organization determines to be highest for the AI systems within a given context of use call for the most urgent prioritization and most thorough risk management process. (Page: 13)

2. Nonethe- less, regularly assessing and prioritizing risk based on context remains important because non-human-facing AI systems can have downstream safety or social implications. (Page: 13)

3. Different types of safety risks may require tailored AI risk management approaches based on context and the severity of potential risks presented. (Page: 19)

4. Responses to the AI risks deemed high priority, as identified by the MAP function, are developed, planned, and doc- umented. Risk response options can include mitigating, transferring, avoiding, or accepting. (Page: 37)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

15. Integration of AI Risk Management into Enterprise Risk Framework

Control Actor: Risk Management Team

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain comprehensive processes to integrate AI risk management effectively into the organization's broader enterprise risk management (ERM) strategies and processes. This ensures a holistic and consistent approach to identifying, assessing, mitigating, and monitoring AI-related risks alongside other organizational risks, aligning AI risk governance with overall corporate risk appetite and strategic objectives.

Enhanced Implementation Guide:

The organization shall establish, maintain, and continually integrate AI-specific risk identification, assessment, mitigation, and monitoring processes into its overarching Enterprise Risk Management (ERM) framework, ensuring alignment with corporate risk appetite and strategic objectives. This is achieved by: developing and codifying specific policies, standards, and procedures for AI risk management, clearly defining roles, responsibilities, and methodologies for AI risk identification, assessment (e.g., impact, likelihood, inherent/residual), treatment, monitoring, and reporting; embedding AI risk considerations into existing ERM processes, including risk registers, risk appetite statements, risk committees, and reporting structures, thereby updating ERM methodologies to include AI-specific risk categories (e.g., bias, privacy, security, explainability, performance degradation, ethical risks, regulatory non-compliance); articulating the organization's specific risk appetite for AI technologies, aligning it with the broader corporate risk appetite; creating or adapting tools and templates for conducting AI-specific risk assessments (e.g., AI impact assessments, ethical reviews); providing training to relevant stakeholders (AI development teams, legal, compliance, business units, senior management) on AI risks and the integrated risk management processes; implementing a schedule for regular review and update of AI risk profiles, policies, and the integration effectiveness within the ERM framework; and fostering cross-functional collaboration between AI development teams, legal, compliance, privacy, cybersecurity, and the ERM team to ensure comprehensive risk identification and mitigation. This control is performed on an ongoing basis, with formal reviews of the integrated processes at least Annually, and ad-hoc reviews triggered by new AI system deployments, significant changes to existing AI systems, or emerging AI risk events. The effectiveness of this control is monitored Quarterly, with a comprehensive review by the Risk Management Committee or relevant governance body at least Annually. This control is primarily Manual, leveraging human judgment, policy development, and committee oversight, with aspects utilizing Semi-Automated GRC tools for documentation, tracking, and reporting. Control effectiveness is evaluated by: reviewing documented AI risk management policies, procedures, and their integration within the ERM framework, including verifying that AI risk categories are included in the organization's risk register and risk appetite statements; examining the enterprise risk register to confirm that AI-specific risks are identified, assessed, and assigned appropriate ownership and mitigation plans consistent with the ERM methodology; reviewing minutes from ERM committee meetings, risk council meetings, or relevant governance forums to confirm regular discussion and oversight of AI risks; verifying that AI systems undergo formal risk assessments (e.g., AI impact assessments, ethical reviews) before deployment and periodically thereafter; interviewing key stakeholders (e.g., AI development leads, legal, compliance, ERM team members) to confirm awareness and adherence to integrated AI risk management processes; assessing adherence to established AI risk management policies and procedures across relevant business units; and reviewing any internal or external audit findings related to AI risk management and ERM integration to identify gaps. Key control indicators include: the percentage of AI systems with completed risk assessments; the number of AI risks formally integrated into the ERM register; the frequency of AI risk reporting to governance bodies; the compliance rate with AI risk policy review and update schedules; the percentage of relevant staff trained on the integrated AI risk framework; the number of AI-related incidents categorized and managed within the ERM incident response process; and the average timeliness of AI risk mitigation actions.

Requirement Statements:

1. AI risk management should be integrated and incorporated into broader enterprise risk management strategies and processes. (Page: 13)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

16. AI System Lifecycle Test, Evaluation, Verification, and Validation (TEVV) Capacity and Processes

Control Actor: Risk Management Team, Operations Team, Technical Management

Control Types: preventive, detective, corrective, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish, maintain, and continuously augment capacity for comprehensive Test, Evaluation, Verification, and Validation (TEVV) processes throughout the entire AI system lifecycle. These processes are crucial to ensure the trustworthiness, performance, and reliability of AI systems in various operational contexts, minimizing risks and ensuring compliance with organizational standards and regulatory requirements. Regularly performed, these TEVV tasks provide insights into technical, societal, legal, and ethical standards or norms, assist with anticipating impacts and assessing and tracking emergent risks, and enable both mid-course remediation and post-hoc risk management. This also includes ensuring the reliability of AI system operation, confirming its ability to perform as required, without failure, for a given time interval, under given conditions, including its entire lifetime.

Enhanced Implementation Guide:

The organization shall establish, maintain, and continuously enhance a comprehensive Test, Evaluation, Verification, and Validation (TEVV) framework, including adequate capacity, skilled personnel, and robust processes, throughout the entire AI system lifecycle to ensure the trustworthiness, performance, reliability, and regulatory compliance of all AI systems. This encompasses defining roles, responsibilities, and methodologies for various TEVV types (e.g., functional, security, performance, bias/fairness, robustness, ethical alignment, regulatory compliance), integrating TEVV seamlessly into the AI System Development Lifecycle (AI-SDLC), allocating dedicated resources (tools, infrastructure, subject matter expertise), and ensuring rigorous documentation of all TEVV activities and outcomes to facilitate traceability and auditability. Control performance is continuous, occurring at key AI system lifecycle milestones (e.g., design, development, integration, pre-deployment, post-deployment), upon significant model or system updates, and periodically during operational phases to proactively identify and mitigate risks. Monitoring of the TEVV process effectiveness is conducted quarterly by the Risk Management Team and Operations Team, with a comprehensive annual review led by Technical Management and Internal Audit to assess adherence to the framework and identify areas for enhancement. This control is semi-automated, leveraging automated testing platforms, continuous integration/continuous delivery (CI/CD) pipelines, and monitoring tools for efficiency, complemented by expert-driven manual review, ethical assessments, stakeholder validation, and human-in-the-loop validation for complex or qualitative aspects. Control effectiveness is evaluated by assessing the completeness and rigor of TEVV documentation for all AI systems, reviewing independent audit findings related to AI system assurance, analyzing the trend of post-deployment incidents and their root causes (specifically those attributable to inadequate TEVV), verifying the adequacy of allocated TEVV resources and tooling, and confirming the integration of TEVV requirements into AI system design and development gates. Key Control Indicators (KCIs) include: the percentage of AI systems undergoing independent TEVV prior to deployment; the number of critical vulnerabilities or bias incidents identified and remediated during TEVV phases versus post-deployment; the coverage of TEVV activities across defined AI system lifecycle stages; the average time to complete TEVV cycles; and the percentage of AI systems with complete, auditable TEVV documentation.

Requirement Statements:

1. The NIST modification high- lights the importance of test, evaluation, verification, and validation (TEVV) processes throughout an AI lifecycle and generalizes the operational context of an AI system. (Page: 14)

2. Performed regularly, TEVV tasks can provide insights relative to technical, societal, legal, and ethical standards or norms, and can assist with anticipating impacts and assessing and tracking emergent risks. As a regular process within an AI lifecycle, TEVV allows for both mid-course remediation and post-hoc risk management. (Page: 14)

3. Reliability is a goal for overall correctness of AI system operation under the conditions of expected use and over a given period of time, including the entire lifetime of the system. (Page: 18)

4. augmented capacity for TEVV of AI systems and associated risks. (Page: 24)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

17. AI Societal Values and Tradeoffs External Engagement

Control Actor: Senior Management

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for Senior Management to promote discussions with external stakeholders regarding the tradeoffs necessary to balance diverse societal values and priorities. These discussions should specifically address considerations related to civil liberties and rights, equity, the environment, the planet, and the economy, in the context of AI system development and deployment. The goal is to ensure a comprehensive understanding of societal impacts and to inform responsible AI practices by considering broader community perspectives.

Enhanced Implementation Guide:

Senior Management is mandated to proactively establish and maintain structured, auditable processes for ongoing external engagement with diverse stakeholders to discuss, understand, and integrate societal values, tradeoffs, and potential impacts (including civil liberties, rights, equity, environmental, and economic considerations) into AI system development and deployment. This includes defining a stakeholder engagement framework, identifying key groups (e.g., civil society, academia, advocacy, affected communities), designing accessible engagement mechanisms (e.g., workshops, advisory boards, forums), and ensuring discussions specifically address ethical, social, economic, and environmental tradeoffs. A robust system must be in place to capture, analyze, and systematically integrate insights from these discussions into AI governance frameworks, policy development, risk assessments, and product design decisions, with outcomes communicated transparently. This control is ongoing, with formal engagements conducted at least semi-annually and continuous informal channels maintained, triggered by significant AI milestones or emerging societal concerns. It is a Manual control, relying heavily on human interaction, judgment, and documentation. Control effectiveness is evaluated quarterly by the Risk & Compliance function or an independent assurance team, by assessing the presence, quality, and documented impact of external engagement processes; verifying adherence to the framework, diversity of stakeholder participation, systematic integration of feedback into AI design principles and risk mitigation strategies, active Senior Management involvement, and communication of outcomes. Key

Control Indicators include: the number of unique external stakeholder engagement events held per period; the diversity index of stakeholder groups engaged; the number of documented insights/recommendations derived from external engagements; the percentage of AI projects/policies where external stakeholder feedback has been formally integrated; Senior Management attendance rates at key engagement events; and the number of internal policy/guidance updates directly referencing external stakeholder input.

Requirement Statements:

1. These actors can: • promote discussion of the tradeoffs needed to balance societal values and priorities related to civil liberties and rights, equity, the environment and the planet, and the economy. (Page: 15)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

18. AI Risk Management Team Diversity and Inclusion

Control Actor: Human Resources

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes to ensure that teams performing AI Risk Management Framework functions integrate diverse perspectives, disciplines, professions, and experiences. This initiative, typically led by Human Resources, is critical for fostering open sharing of ideas, identifying implicit assumptions about AI system purposes and functions, and proactively surfacing existing and emergent risks throughout the AI system lifecycle.

Enhanced Implementation Guide:

The organization must establish and maintain auditable processes to ensure that all teams involved in AI Risk Management Framework (AI RMF) functions are composed of individuals integrating diverse perspectives, disciplines, professions, and experiences, thereby promoting comprehensive risk identification and mitigation. Human Resources, in collaboration with AI Governance, Legal, and relevant business units, shall develop and implement formal policies and procedures for staffing AI Risk Management Framework (AI RMF) teams. This includes defining key diversity dimensions relevant to AI (e.g., technical expertise across AI domains, ethical, legal, sociological, human-centered design, demographic diversity, and varied experiential backgrounds), establishing recruitment strategies that proactively target diverse candidate pools (e.g., utilizing specialized job boards, partnering with diversity-focused professional organizations, and structuring interview panels to minimize bias), incorporating mandatory unconscious bias and cultural competency training for all personnel involved in the hiring and team formation process for AI RMF roles, and developing structured onboarding programs that explicitly emphasize the value of diverse perspectives in AI risk identification and management. Regular reviews of AI RMF team compositions against defined diversity targets should be conducted to identify and address any gaps, with remediation plans promptly developed and executed. This control is ongoing, applied at the inception of new AI RMF teams and during significant restructuring or expansion of existing teams. Policies and procedures governing this control should be reviewed and updated at least annually, or more frequently if triggered by changes in regulatory requirements or organizational structure. Monitoring of this control should occur quarterly, or semi-annually at a minimum, depending on the volume of AI initiatives and team changes, with ad-hoc monitoring instituted for critical AI projects or upon identification of new high-risk AI applications. This control is semi-automated, leveraging HR Information Systems (HRIS) for demographic data analysis, diversity dashboards, and applicant tracking systems, while the core decision-making, policy enforcement, and cultural integration aspects remain manual. Effectiveness is evaluated by assessing the presence and adherence to established diversity and inclusion policies specific to AI RMF teams. This involves reviewing HR policies, recruitment guidelines, training records for unconscious bias, and documented evidence of diverse candidate slates for AI RMF positions. Audit evidence includes analysis of AI RMF team rosters against established diversity criteria and participation rates in relevant Diversity & Inclusion training. Furthermore, qualitative assessments through interviews with team members and leaders can ascertain whether diverse perspectives are actively encouraged, integrated into risk identification, and contribute to more robust mitigation strategies, demonstrating the control's proactive impact on surfacing existing and emergent risks. Key control indicators (KCIs) include: the percentage of AI RMF teams consistently meeting defined diversity targets (e.g., representation across relevant technical disciplines, ethical backgrounds, and demographic attributes); the average number of unique professional disciplines or expertise areas represented per AI RMF team; the completion rate of unconscious bias training among hiring managers and AI RMF team leads; the ratio of diverse candidates interviewed to total candidates interviewed for AI RMF roles; employee sentiment scores related to psychological safety and inclusivity within AI RMF teams (e.g., via anonymous surveys); and the number of identified implicit assumptions or emergent risks directly attributed to the integration of diverse perspectives in AI RMF activities.

Requirement Statements:

1. The AI RMF functions, described in Section 5, require diverse perspectives, disciplines, professions, and experiences. Diverse teams contribute to more open sharing of ideas and assumptions about the purposes and functions of technology – making these implicit aspects more explicit. This broader collective perspective creates opportunities for surfacing problems and identifying existing and emergent risks. (Page: 15)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

19. Senior Management Oversight of AI Trustworthiness Tradeoffs and Practical Implementation

Control Actor: Senior Management

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for Senior Management to strategically balance and prioritize various AI trustworthiness characteristics (e.g., validity, safety, fairness, privacy, security, transparency) based on the specific context of use and criticality of the AI system. This includes acknowledging and managing inherent tradeoffs between characteristics, understanding that not all characteristics may apply or be equally important in every situation, and ensuring that decisions reflect the organization's risk tolerance and ethical guidelines. Furthermore, when enhancing transparency and accountability measures, Senior Management must consider the impact on the implementing entity, including the level of necessary resources and the need to safeguard proprietary information.

Enhanced Implementation Guide:

Enhanced Description:

Control Statement: Senior Management must establish, maintain, and actively utilize a formalized process for strategically balancing and prioritizing AI trustworthiness characteristics (e.g., validity, safety, fairness, privacy, security, transparency) in a context-specific manner, acknowledging inherent tradeoffs, aligning with the organization's risk tolerance and ethical guidelines, and pragmatically considering the resource and proprietary information impacts of transparency and accountability measures.

Implementation Guide: To implement this control, Senior Management shall define clear organizational policies and frameworks for AI trustworthiness. This involves establishing a structured approach for identifying and assessing relevant trustworthiness characteristics for each AI system based on its criticality and specific use-case. A robust decision-making process must be developed to facilitate explicit discussions and documented approvals of tradeoffs between characteristics (e.g., prioritizing performance over explainability in a safety-critical system, or balancing privacy with data utility). Furthermore, Senior Management must ensure that all decisions on AI trustworthiness tradeoffs are clearly communicated to relevant stakeholders and that the practical implications, including necessary resource allocation and the safeguarding of proprietary information, are thoroughly evaluated and factored into implementation plans for transparency and accountability mechanisms.

Control Frequency: This control should be performed proactively for all new AI systems during their design and pre-deployment phases, and re-evaluated periodically (e.g., annually or biennially) for existing critical AI systems, or upon significant changes in system functionality, use-case, risk profile, or relevant regulatory requirements.

Monitoring Frequency: The effectiveness of this control should be monitored on a semi-annual or annual basis by an independent function such as Internal Audit, AI Governance Office, or Compliance, to ensure ongoing adherence and effectiveness.

Control Type: Manual (involving strategic discussions, expert judgment, and documented decision-making processes by Senior Management, potentially supported by semi-automated tools for data aggregation or documentation management).

Evaluating Control Effectiveness: Effectiveness is evaluated by reviewing evidence of Senior Management's active engagement in tradeoff discussions, including meeting minutes, decision logs, risk assessments, and system design documents. Verification includes confirming that explicit justifications for prioritized characteristics and acknowledged tradeoffs are documented, that decisions align with the organization's stated risk appetite and ethical principles, and that resource and proprietary information considerations for transparency and accountability are explicitly addressed in implementation plans. Interviews with key stakeholders (e.g., AI development teams, legal, risk management) can further confirm the integration and communication of these strategic decisions.

Key Control Indicators (KCIs):
1. Percentage of new or critical AI systems with documented Senior Management-approved trustworthiness tradeoff decisions.
2. Number of instances where resource allocation and proprietary information impacts related to AI transparency/accountability were explicitly considered and documented by Senior Management.
3. Timeliness of Senior Management review and approval for new/significant AI systems against predefined service level agreements (SLAs).
4. Results from internal audits or external assessments indicating the robustness and adherence to AI trustworthiness governance principles.

5. Evidence of clear communication channels for Senior Management's AI trustworthiness decisions to relevant development and operational teams.

Requirement Statements:

1. Creating trustworthy AI requires balancing each of these characteristics based on the AI system’s context of use. Addressing AI trustworthiness characteristics individually will not ensure AI system trust- worthiness; tradeoffs are usually involved, rarely do all characteristics apply in every set- ting, and some will be more or less important in any given situation. (Page: 17)

2. Measures to enhance transparency and accountability should also consider the impact of these efforts on the implementing entity, including the level of necessary resources and the need to safeguard proprietary information. (Page: 21)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ___________________________________

Evidence/Comments:

____

____

____

20. AI Governance Committee Value Tradeoff Resolution

Control Actor: AI Governance Committee

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for the AI Governance Committee to transparently and justifiably resolve tradeoffs between different measures and values in the context of AI system design, development, and deployment. This includes documenting the decision-making context, the values at play, and the rationale for chosen resolutions to ensure accountability and align with organizational ethical guidelines and risk tolerance. Such processes facilitate responsible AI development and deployment by providing clear guidance on navigating complex ethical and operational dilemmas, including considering potential trade-offs between privacy-enhancing techniques and system accuracy (e.g., in conditions of data sparsity) when making decisions about fairness and other values.

Enhanced Implementation Guide:

The AI Governance Committee (AIGC) must establish, document, and consistently apply a structured process for transparently and justifiably resolving inherent trade-offs between competing values (e.g., accuracy vs. privacy, fairness vs. performance) in AI system design, development, and deployment, ensuring decisions are aligned with organizational ethical guidelines, risk appetite, and regulatory requirements. To implement this, the AIGC should first define a formal decision-making framework, outlining criteria for identifying, analyzing, and resolving value trade-offs, integrating ethical principles, legal obligations, business objectives, and stakeholder impact assessments. Secondly, comprehensive documentation standards must be mandated for all trade-off resolution decisions, detailing the AI system context, competing values, potential impacts, alternative solutions, the rationale for the chosen resolution linked to organizational guidelines, and any residual risks or mitigation strategies. This process should be embedded into critical stages of the AI system lifecycle (e.g., concept, design, development, testing, deployment), and regular training on the framework and ethical AI principles must be provided to AIGC members and relevant AI development teams. Clear communication channels for escalating trade-offs and communicating resolutions are also essential. This control is primarily **Manual** in its decision-making core, but can be **Semi-automated** for documentation and tracking using GRC or project management tools. The

Control Frequency is **Ad-hoc/Event-driven**, performed whenever a significant value trade-off is identified in an AI system's lifecycle, and **Annually** for reviewing and updating the underlying process framework. **Monitoring Frequency** should be **Quarterly** by internal audit or compliance, reviewing a sample of decisions, and **Annually** for a comprehensive effectiveness review. Evaluating control effectiveness involves verifying adherence to the defined process, assessing the completeness and quality of documentation, scrutinizing the justification of decisions against ethical guidelines and risk tolerance, confirming that stakeholder impact assessments and mitigation strategies were considered, and ensuring adequate training and a clear audit trail. Key Control Indicators (KCIs) include: the percentage of AI projects where identified value trade-offs were escalated to the AIGC (target 100% of material trade-offs); the average time-to-resolution for escalated trade-offs (e.g., within X business days SLA); the number of documented trade-off resolutions lacking required documentation (target 0); the number of internal audit findings related to AIGC trade-off resolutions (target 0 critical findings); the percentage of AIGC members and relevant AI teams completing trade-off resolution training (target 100%); and the number of post-deployment incidents or ethical concerns directly attributable to unresolved or poorly resolved value trade-offs (target 0).

Requirement Statements:

1. Those depend on the values at play in the relevant context and should be resolved in a manner that is both transparent and appropriately justifiable. (Page: 18)

2. Under certain conditions such as data sparsity, privacy- enhancing techniques can result in a loss in accuracy, affecting decisions about fairness and other values in certain domains. (Page: 22)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

21. AI TEVV Findings Evaluation and Parameter Alignment

Control Actor: Subject Matter Experts

Control Types: administrative, detective, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for Subject Matter Experts to assist in the comprehensive evaluation of Test, Evaluation, Verification, and Validation (TEVV) findings. This includes collaborating with product and deployment teams to align TEVV parameters with specific AI system requirements and real-world deployment conditions. The goal is to ensure that TEVV outcomes accurately reflect system performance and trustworthiness in its intended operational context, facilitating necessary adjustments and improvements.

Enhanced Implementation Guide:

Subject Matter Experts (SMEs) must comprehensively evaluate all Test, Evaluation, Verification, and Validation (TEVV) findings, collaborating with product and deployment teams to ensure TEVV parameters are precisely aligned with AI system requirements and real-world deployment conditions, thereby validating system performance and trustworthiness for its intended operational context. To implement this control, organizations should establish formal processes for SME engagement in TEVV evaluation, including clear roles, responsibilities, and defined collaboration mechanisms (e.g., joint reviews, feedback loops) with product and deployment teams. A dedicated system or repository for TEVV findings and SME evaluations should be maintained, ensuring all parameter alignment rationales and subsequent adjustments are meticulously documented. This control should be performed upon the completion of any TEVV phase, prior to significant AI system deployment or updates, and during periodic reviews of operational AI systems (e.g., annually or upon material changes). Monitoring of this control's effectiveness should occur at least semi-annually by the AI Governance Committee or Internal Audit function. This control is primarily manual, relying on expert judgment and collaborative human interaction, though aspects like finding tracking and notification workflows may be semi-automated. Control effectiveness can be evaluated by reviewing documented TEVV findings, SME evaluation reports, and evidence of parameter alignment discussions; assessing the completeness and timeliness of SME sign-offs; tracing identified TEVV discrepancies to subsequent system adjustments; and interviewing involved SMEs, product, and deployment team members to confirm process adherence and efficacy.

Key control indicators include: the percentage of TEVV findings comprehensively reviewed and endorsed by SMEs, the average resolution time for TEVV-identified discrepancies, the documented alignment score between TEVV parameters and actual deployment conditions, the reduction in post-deployment issues directly attributable to pre-vetted TEVV findings, and the feedback satisfaction scores from product/deployment teams regarding SME collaboration quality.

Requirement Statements:

1. subject matter experts can assist in the evaluation of TEVV findings and work with product and deployment teams to align TEVV parameters to requirements and de- ployment conditions. (Page: 18)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

22. AI System Validation and Ongoing Performance Monitoring

Control Actor: Validation Engineers, Quality Assurance Personnel

Control Types: preventive, detective, administrative

Related Entities: Organization, Provider Organization

Control Description:

Conduct thorough validation of AI systems to objectively confirm that all requirements for a specific intended use or application have been fulfilled. This process involves providing objective evidence that the AI system meets its design specifications and operational objectives, aligning with established validation standards (e.g., ISO 9000:2015) and ensuring fitness for purpose prior to or during deployment. This also includes ongoing testing and monitoring of deployed AI systems to measure their validity, accuracy, robustness, and reliability, confirming intended performance and taking into consideration that certain types of failures can cause greater harm. Specifically, processes developed or adopted in the MEASURE function should include rigorous software testing and performance assessment methodologies with associated measures of uncertainty, comparisons to performance benchmarks, and formalized reporting and documentation of results.

Enhanced Implementation Guide:

This control, 'AI System Validation and Ongoing Performance Monitoring', mandates the rigorous and objective confirmation that all AI system requirements for specific intended uses or applications are fulfilled prior to or during deployment, coupled with continuous performance monitoring to ensure ongoing validity, accuracy, robustness, and reliability, especially where failures could cause significant harm. Implementation necessitates a multi-faceted approach: **pre-deployment validation** involves defining detailed scope, establishing performance benchmarks (e.g., for accuracy, bias, robustness, fairness), applying rigorous software testing and performance assessment methodologies with associated measures of uncertainty, comparing outcomes to benchmarks, and formally documenting results and stakeholder sign-offs, aligning with established standards like ISO 9000:2015; **ongoing monitoring** requires defining continuous metrics (e.g., drift, accuracy decay, fairness), implementing automated tools with defined alert thresholds, and establishing a formalized process for investigation, remediation (e.g., retraining, recalibration), and reporting of anomalies or performance degradation. The control frequency for pre-deployment validation is **prior to initial deployment or significant modification** of any AI system, while ongoing monitoring is

continuous (real-time or near real-time) for all deployed systems, with periodic formal reviews (e.g., monthly/quarterly) of monitoring reports. The monitoring frequency for the control's effectiveness by independent oversight functions (e.g., Internal Audit, Compliance) should be **quarterly or semi-annually**. This control is **semi-automated** in its validation phase (leveraging automated testing alongside manual analysis and documentation) and largely **automated** for ongoing performance monitoring (via specialized tools with manual intervention for investigations). Control effectiveness is evaluated by assessing the completeness and adherence of validation documentation to internal policies and external standards, verifying that all identified pre-deployment issues were resolved, confirming appropriate stakeholder sign-offs, and reviewing the responsiveness and efficacy of ongoing monitoring alerts and remediation processes; ultimately, success is evidenced by the absence of critical AI system failures attributable to inadequate validation or monitoring. Key control indicators (KCIs) include: the **percentage of new/modified AI systems with completed and signed-off validation reports**, the **number of critical validation issues identified and resolved pre-deployment**, the **percentage of deployed AI systems under continuous monitoring**, the **rate of performance degradation alerts and their average resolution time**, the **degree of deviation in key fairness or accuracy metrics from established benchmarks**, and the **frequency of model retraining or recalibration cycles triggered by performance monitoring data**.

Requirement Statements:

1. Validation is the “confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled” (Source: ISO 9000:2015). (Page: 18)

2. Validity and reliability for deployed AI systems are often assessed by ongoing testing or monitoring that confirms a system is performing as intended. Measurement of validity, accuracy, robustness, and reliability contribute to trustworthiness and should take into consideration that certain types of failures can cause greater harm. (Page: 19)

3. Processes developed or adopted in the MEASURE function should include rigorous software testing and performance assessment methodologies with associated measures of uncertainty, comparisons to performance benchmarks, and formal- ized reporting and documentation of results. (Page: 33)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ______________________________

Evidence/Comments:

____

____

____

23. AI System Accuracy, Reliability, Safety, and Bias Mitigation Assurance for Deployment

Control Actor: AI Development Team

Control Types: preventive, technical, administrative, detective, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement rigorous processes for the AI Development Team to ensure that AI systems are accurate, reliable, adequately generalized to data and settings beyond their training data, and operate safely. These processes must include comprehensive testing, evaluation, and quality assurance measures throughout the development lifecycle to identify and prevent the deployment of systems that exhibit inaccuracies, unreliability, or unsafe behavior. This specifically includes implementing responsible design, development, and deployment practices to improve the safe operation of AI systems, such as rigorous simulation, in-domain testing, real-time monitoring, and integrating mechanisms for human intervention (e.g., shutdown, modification) when AI systems deviate from intended or expected functionality. AI System Developers must conduct accuracy measurements with clearly defined and realistic test sets that are representative of conditions of expected use, ensuring details about test methodology are included in associated documentation, and consider disaggregating results for different data segments to identify potential biases or performance discrepancies. Furthermore, actively implement strategies and techniques to mitigate identified harmful biases in AI systems, acknowledging that bias mitigation is a continuous process contributing to system fairness and trustworthiness, but does not inherently guarantee absolute fairness. The objective is to mitigate negative AI risks and uphold trustworthiness by ensuring that only systems meeting defined performance, quality, and safety standards are deemed ready for deployment.

Enhanced Implementation Guide:

This control mandates the establishment and rigorous adherence to processes within the AI Development Team to ensure that all AI systems deployed meet stringent standards for accuracy, reliability, generalizability, and safety, coupled with proactive mitigation of harmful biases. The **control statement** is: To ensure that all AI systems deployed meet defined standards for accuracy, reliability, safety, generalizability, and mitigated bias, through rigorous testing, evaluation, and responsible development practices throughout their lifecycle. The

high-level control implementation guide requires formalizing an AI System Development Lifecycle (AI-SDLC) that incorporates mandatory gates for comprehensive testing, evaluation, and approval prior to deployment. This includes defining and mandating the use of diverse, representative test sets for accuracy and reliability measurements, conducting rigorous simulations and in-domain testing for safety assurance, and implementing real-time monitoring capabilities for post-deployment performance and deviation detection. Crucially, the AI-SDLC must integrate responsible design principles that allow for human intervention mechanisms (e.g., shutdown, override, modification) in critical AI systems. A dedicated bias detection and mitigation framework must be established, requiring disaggregated performance analysis across different data segments to identify and address potential biases. Comprehensive documentation of all testing procedures, results, identified biases, mitigation strategies, and human intervention protocols is mandatory for each AI system. Final deployment authorization must be contingent on review and approval by a designated cross-functional body, such as an AI Governance Committee, confirming adherence to all performance, quality, and safety criteria. The **control frequency** is continuous throughout the AI-SDLC (development, testing, pre-deployment phases), performed prior to each major model version deployment or significant change, and then regularly post-deployment as part of continuous model monitoring and maintenance (e.g., triggered by drift detection or scheduled reviews). The **monitoring frequency** for internal oversight by AI Governance or Compliance functions should be quarterly, with automated technical monitoring being continuous post-deployment, and an independent audit performed annually or bi-annually by Internal Audit. This control is primarily **semi-automated**, as it leverages automated testing tools and real-time monitoring but requires significant manual oversight, expert analysis for bias and safety assessments, human decision-making for approvals, and manual intervention capabilities when systems deviate. **To evaluate control effectiveness**, compliance teams should conduct documentation reviews to verify the existence and completeness of AI-SDLC policies, test plans, evaluation reports, bias assessment reports, and human intervention protocols for all deployed AI systems. Process walkthroughs and interviews with the AI Development Team should confirm adherence to defined procedures for testing, evaluation, bias mitigation, and safety assurance. Independent sample testing of recently deployed or updated AI systems is crucial to verify test methodologies, results, and the efficacy of mitigation actions taken, ensuring deployment approvals were based on established criteria. Effectiveness also hinges on assessing if identified inaccuracies, unreliable behaviors, safety risks, or biases were effectively addressed prior to deployment or through post-deployment corrective actions. Key **control indicators** include the percentage of AI models undergoing rigorous pre-deployment testing and evaluation, the number of critical safety/accuracy/bias issues identified and remediated pre-deployment, the test coverage percentage for accuracy, reliability, and generalizability metrics, quantitative bias metrics (e.g., disparate impact, equal opportunity difference) with target thresholds for sensitive attributes, the trend in documented safety incidents or unintended behaviors post-deployment, the Mean Time To Recovery (MTTR) for AI system failures or deviations requiring human intervention, and the adherence rate to AI-SDLC gate requirements (e.g., percentage of systems approved by the review board).

Requirement Statements:

1. Deployment of AI systems which are inaccurate, unreliable, or poorly gener- alized to data and settings beyond their training creates and increases negative AI risks and reduces trustworthiness. (Page: 18)

2. Accuracy measurements should always be paired with clearly defined and realistic test sets – that are representative of conditions of expected use – and details about test methodology; these should be included in associated documentation. (Page: 19)

3. Accuracy measurements may include disaggregation of results for different data segments. (Page: 19)

4. Safe operation of AI systems is improved through: • responsible design, development, and deployment practices; (Page: 19)

5. Employing safety considerations during the lifecycle and starting as early as possible with planning and design can prevent failures or conditions that can render a system dangerous. (Page: 20)

6. Other practical approaches for AI safety often relate to rigorous simulation and in-domain testing, real-time monitoring, and the ability to shut down, modify, or have human inter- vention into systems that deviate from intended or expected functionality. (Page: 20)

7. Systems in which harmful biases are mitigated are not necessarily fair. (Page: 22)

8. AI systems should be tested before their deployment and regu- larly while in operation. (Page: 33)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

24. AI System Error Remediation and Human Oversight Protocol

Control Actor: Risk Management Team

Control Types: preventive, detective, corrective, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain a protocol for prioritizing the minimization of potential negative impacts arising from AI systems, including specific provisions for human intervention in cases where the AI system cannot detect or correct errors independently. This protocol ensures that mechanisms are in place for human oversight and remedial action when AI system failures or critical uncertainties occur, safeguarding against unintended consequences and ensuring responsible AI operation.

Enhanced Implementation Guide:

This enhanced description details the "AI System Error Remediation and Human Oversight Protocol" control, a critical component for responsible AI operation.

Control Statement: A formal, documented, and regularly reviewed protocol is established and maintained to ensure prompt human oversight and timely remedial action for AI system errors, failures, or critical uncertainties that the AI cannot self-correct, thereby prioritizing the proactive minimization of potential negative impacts and safeguarding against unintended consequences.

High-Level Control Implementation Guide:
The Risk Management Team, in collaboration with AI development, operations, legal, and subject matter experts, shall:
1. **Develop the Protocol:** Create a comprehensive AI Incident Response and Remediation Protocol outlining clear definitions for AI system errors, failures, and critical uncertainties. This protocol must define:
 * **Trigger Conditions:** Specific thresholds or indicators (e.g., performance degradation, anomalous outputs, safety-critical deviations, ethical concerns, compliance breaches) that necessitate human intervention.
 * **Roles and Responsibilities:** Delineate clear roles for incident detection, triage, escalation, analysis, decision-making (e.g., human-in-the-loop validation, temporary

deactivation), and remediation, including cross-functional team responsibilities.
    * **Communication & Escalation Paths:** Establish clear communication channels and defined escalation matrices for AI incidents, including reporting to relevant stakeholders (e.g., senior management, legal, regulatory bodies if required).
    * **Remediation Procedures:** Outline prescribed steps for addressing different types of AI failures, including data correction, model retraining, algorithm adjustments, temporary deactivation, or full system shutdown. This should also cover procedures for documenting root causes and lessons learned.
    * **Documentation & Logging:** Mandate comprehensive logging of all AI system incidents, human interventions, remediation actions, and their outcomes.
2. **Integrate with Existing Frameworks:** Ensure the protocol integrates seamlessly with the organization's broader enterprise risk management, incident response, and business continuity frameworks.
3. **Training & Awareness:** Conduct regular training for all relevant personnel (AI developers, operations, risk, legal, and business stakeholders) on the protocol's content, their specific roles, and incident management procedures.
4. **Provider Organization Engagement:** For AI systems provided by external entities, the protocol must include provisions for contractual requirements ensuring the Provider Organization adheres to similar robust error remediation and human oversight standards, and provides necessary transparency and collaboration during incidents.
5. **Review and Update:** Establish a schedule for periodic review and update of the protocol, particularly after significant AI system changes, new AI deployments, or major incidents, to incorporate lessons learned and adapt to evolving risks or regulatory requirements.

Control Frequency: The protocol is *established* initially and *maintained* (reviewed and updated) at least annually, or whenever significant changes to AI systems occur, new AI systems are deployed, or a major AI-related incident takes place. Remedial *actions* and human interventions are performed on an ad-hoc basis as incidents arise.

Monitoring Frequency: Quarterly, or immediately following any significant AI incident or near-miss event, to assess the protocol's effectiveness and adherence.

Control Type: Semi-Automated with a strong **Manual** oversight component. While initial error detection may involve automated monitoring tools, the core decision-making, root cause analysis, human intervention, and remediation steps are inherently manual processes requiring expert judgment. The protocol itself is an **Administrative** control, acting as a **Preventive** measure by establishing guidelines, a **Detective** measure by outlining incident identification, and a **Corrective** measure by defining remediation.

Guide for Evaluating Control Effectiveness:
Effectiveness is evaluated by assessing the protocol's existence, adherence, and actual impact on minimizing negative outcomes. This includes:

1. **Protocol Completeness:** Verify the protocol is formally documented, comprehensive, and addresses all critical aspects outlined in the implementation guide.
2. **Adherence Review:** Audit a sample of recorded AI incidents to confirm that the protocol was followed, including timely detection, appropriate escalation, human intervention where necessary, and documented remediation.
3. **Effectiveness of Remediation:** Evaluate the success rate and timeliness of remedial actions in mitigating negative impacts and preventing recurrence.
4. **Training Compliance:** Confirm that relevant personnel have completed required training on the protocol.
5. **Lessons Learned Integration:** Verify that findings from past incidents (root causes, identified gaps) have led to meaningful updates to the protocol or AI system design.
6. **Simulation/Tabletop Exercises:** Conduct periodic tabletop exercises or simulations of AI failure scenarios to test the protocol's operational readiness and identify potential weaknesses before real-world incidents occur.
7. **Provider Oversight:** For Provider Organizations, review contractual agreements and audit their incident response logs and human oversight mechanisms to ensure alignment with organizational standards.

Key Control Indicators (KCIs):

* **Number of critical AI incidents with documented human intervention:** Measures the application of oversight.
* **Average time to detect critical AI errors:** Shorter times indicate more effective detection mechanisms.
* **Average time to initiate human intervention for critical AI errors:** Reflects response readiness.
* **Percentage of critical AI incidents resolved within defined service level objectives (SLOs) or remediation targets:** Measures efficiency and effectiveness of resolution.
* **Number of recurrences of previously remediated AI errors:** Indicates the effectiveness of root cause analysis and permanent fixes.
* **Number of critical AI incidents resulting in significant negative business, reputational, or ethical impact:** Ideally zero, indicating successful prevention/mitigation.
* **Frequency of protocol reviews and updates:** Ensures the protocol remains current and relevant.

*   **Completion rate of required training for control actors:** Measures personnel preparedness.

Requirement Statements:

1. AI risk management efforts should prioritize the minimization of potential negative impacts, and may need to include human intervention in cases where the AI system cannot detect or correct errors. (Page: 19)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

25. AI System Resilience Design and Safe Degradation

Control Actor: AI Development Team

Control Types: preventive, technical, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement processes for the AI Development Team to design and build AI systems with inherent resilience, enabling them to withstand unexpected adverse events or environmental changes, maintain core functionality amidst such changes, and degrade safely and gracefully when necessary. This ensures the AI system's ability to operate reliably under varying conditions and to minimize negative impacts during critical failures, contributing to overall trustworthiness.

Enhanced Implementation Guide:

This control mandates that AI systems are designed and built with inherent resilience, enabling them to withstand unexpected adverse events or environmental changes, maintain core functionality amidst such changes, and degrade safely and gracefully when necessary, thereby ensuring reliable operation and minimizing negative impacts during critical failures. To achieve this, the AI Development Team must integrate resilience principles (e.g., redundancy, fault tolerance, circuit breakers, rate limiting) into the AI system architecture from the initial design phase. This includes conducting AI-specific threat modeling and Failure Modes and Effects Analysis (FMEA) to identify potential failure points and define clear, documented safe degradation pathways, such as fallback mechanisms (e.g., reverting to simpler models, human-in-the-loop intervention, pre-computed safe outputs). Robust error handling and comprehensive monitoring with automated alerting for AI system health, performance, data quality, and model drift are essential. Resilience testing, including chaos engineering, stress testing, and adversarial robustness testing, must be performed to validate the system's ability to withstand various adverse conditions and verify safe degradation. This control is primarily performed during the AI system design, development, and significant architectural changes, with resilience testing conducted periodically (e.g., quarterly or upon major changes). Monitoring of system health and operational metrics is continuous, complemented by periodic review of incident reports (e.g., monthly/quarterly). The control is largely **Semi-Automated**, encompassing manual design and analysis activities, automated technical implementations like failover and monitoring, and requiring manual intervention for complex incident response.

Control effectiveness is evaluated by reviewing design documentation for explicit resilience principles and degradation strategies, conducting architectural reviews to confirm integration of resilience patterns, analyzing results from resilience tests to confirm safe degradation under simulated conditions, assessing post-incident reports for the proper functioning of degradation mechanisms and minimized impact, and verifying implementation of error handling and fallback logic through code reviews. Key control indicators include the percentage of critical AI system failure modes with documented degradation strategies, the pass rate of resilience/chaos engineering tests, the Mean Time To Recovery (MTTR) for AI system-related incidents, the number of critical incidents where safe degradation mechanisms failed, and the uptime/availability of core AI functionalities under stress or adverse conditions.

Requirement Statements:

1. AI systems, as well as the ecosystems in which they are deployed, may be said to be resilient if they can withstand unexpected adverse events or unexpected changes in their environment or use – or if they can maintain their functions and structure in the face of internal and external change and degrade safely and gracefully when this is necessary (Adapted from: ISO/IEC TS 5723:2022). (Page: 20)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

26. AI System Cybersecurity and Data Protection

Control Actor: Information Security Officer

Control Types: preventive, technical, administrative

Related Entities: Organization, Provider Organization

Control Description:

Implement and maintain robust cybersecurity measures and protection mechanisms, applying guidelines from established frameworks such as the NIST Cybersecurity Framework and Risk Management Framework, to ensure the confidentiality, integrity, availability, and resilience of AI systems, their models, training data, and associated intellectual property throughout their lifecycle. This includes maintaining the provenance of training data and supporting attribution of AI system decisions to subsets of training data to enhance transparency and accountability. These measures must prevent unauthorized access and use, specifically addressing common and emerging AI-specific security concerns such as adversarial examples, data poisoning, and the exfiltration of sensitive information. This control ensures that AI systems are secured against malicious attacks and vulnerabilities, preserving their trustworthiness and operational integrity and enabling them to withstand unexpected adverse events.

Enhanced Implementation Guide:

Establish, implement, and continuously maintain a robust cybersecurity and data protection program for all AI systems, models, training data, and associated intellectual property throughout their lifecycle, aligning strictly with established frameworks such as the NIST Cybersecurity Framework (CSF) and Risk Management Framework (RMF). This control must ensure the confidentiality, integrity, availability, and resilience of these AI assets by implementing comprehensive preventive and detective technical and administrative controls. Key implementation steps include conducting regular AI-specific threat modeling and risk assessments (e.g., for adversarial examples, data poisoning, model inversion, and data exfiltration), enforcing stringent access controls (including least privilege and multi-factor authentication) for AI development and production environments, encrypting all AI data (training data, models, inference data) at rest and in transit, and securing the AI model development and deployment pipeline. Furthermore, mechanisms must be in place to ensure the provenance of training data and support the attribution of AI system decisions to subsets of training data to enhance transparency and accountability. The control requires continuous monitoring of AI system logs and security events for anomalies, regular vulnerability scanning and penetration testing of AI infrastructure and applications, and proactive adversarial robustness testing of AI models. Incident response plans must be specifically tailored to address AI-specific security breaches and attacks. This control is primarily **semi-automated**, leveraging automated tools for monitoring, scanning, and enforcement, but requiring significant manual effort for policy definition, risk assessment, incident response coordination, and human oversight. **Control frequency** is ongoing for technical enforcement and continuous monitoring, with periodic reviews (e.g., annually for policies and risk assessments, quarterly for access controls) and event-driven activities (e.g., post-major system changes, new AI model deployments). **Monitoring frequency** is continuous for security logs and performance, with weekly/monthly reviews of vulnerability reports and quarterly/bi-annual effectiveness assessments. To evaluate control effectiveness, regular audits should verify adherence to AI security policies, examine the results of penetration tests and adversarial robustness assessments, review incident response drill outcomes, assess the completeness and accuracy of data provenance records, and confirm the timely remediation of identified vulnerabilities. Key control indicators (KCIs) include: the percentage of AI systems and critical data covered by comprehensive security assessments, the average time to detect (MTTD) and respond (MTTR) to AI-specific security incidents, the number of successful adversarial attacks mitigated during testing, the percentage of AI decisions for which data provenance and attribution can be successfully demonstrated, and the rate of compliance with mandatory AI security training for relevant personnel.

Requirement Statements:

1. Common security concerns relate to adversarial examples, data poisoning, and the exfiltration of models, training data, or other intellectual property through AI system endpoints. AI systems that can maintain confidentiality, integrity, and availability through protection mechanisms that prevent unauthorized access and use may be said to be secure. (Page: 20)

2. Guidelines in the NIST Cybersecurity Framework and Risk Manage- ment Framework are among those which are applicable here. (Page: 20)

3. Maintaining the provenance of training data and supporting attribution of the AI system’s decisions to subsets of training data can assist with both transparency and accountability. (Page: 21)

Implementation Status:

☐ Implemented

☐ Partially Implemented

□ Not Implemented

□ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

27. Proportional Transparency and Accountability for Severe AI Consequences

Control Actor: Senior Management

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for Senior Management to proactively and proportionally adjust transparency and accountability practices for AI systems. This is especially critical when the consequences of AI system operation are severe, such as those potentially impacting life and liberty. This ensures that the level of openness and responsibility is dynamically adapted to the criticality of potential harms, fostering responsible AI development and deployment in high-stakes contexts.

Enhanced Implementation Guide:

Control Statement: Senior Management must establish, formalize, and continuously adapt a robust framework to ensure that the levels of transparency and accountability applied to AI systems are directly proportionate to the severity of their potential consequences, particularly for those impacting life, liberty, or fundamental rights.

High-Level Implementation Guide: Senior Management, in conjunction with relevant AI Governance bodies (e.g., AI Ethics Committee, Risk Management), must first define clear criteria for categorizing AI systems by the severity of their potential consequences, including a risk matrix or impact assessment methodology. Subsequently, for each severity category, they must establish corresponding minimum and enhanced requirements for transparency (e.g., model explainability, data provenance, decision rationale disclosure) and accountability (e.g., clear ownership of outcomes, established remediation pathways, human oversight mechanisms, auditability). This involves creating a formal review and approval process for all high-consequence AI systems prior to and during deployment, ensuring that the selected transparency and accountability measures align proportionally with the assessed risk. Documentation of these decisions, including the rationale for chosen levels of transparency and accountability, is mandatory. Furthermore, Senior Management must ensure regular communication of these practices to internal stakeholders (development, legal, product teams) and relevant external parties (e.g., users, regulators) as appropriate, alongside establishing mechanisms for ongoing feedback and redress.

Control Frequency: This control is performed on an ad-hoc basis whenever a new high-consequence AI system is introduced, an existing high-consequence system undergoes significant modification, or new severe consequence potentials are identified. Additionally, a comprehensive review of the framework's effectiveness and application across the organization's AI portfolio must be conducted at least Annually as part of the overall AI governance strategy.

Monitoring Frequency: The operational effectiveness of this control should be monitored Quarterly by the Compliance and Risk Management functions, assessing adherence to established processes and documentation requirements. A comprehensive independent review by Internal Audit must be conducted Annually to evaluate the framework's design effectiveness and operating effectiveness against organizational policies and best practices.

Control Type: This control is primarily **Manual**, involving strategic decision-making, policy formulation, and oversight by Senior Management. However, it can be supported by **Semi-automated** tools for risk assessment, documentation management, and tracking of AI system inventories and their associated consequence levels.

Guide for Evaluating Control Effectiveness: To evaluate effectiveness, auditors should examine formal records of Senior Management's or AI Governance bodies' decisions concerning high-consequence AI systems, verifying that a proportional approach to transparency and accountability has been systematically applied and documented. This includes reviewing AI system risk assessments, associated transparency/accountability plans, and approval records. Effectiveness is also evidenced by the consistent application of defined criteria for proportionality, the presence of clear rationales for chosen transparency/accountability levels, and verifiable communication of these practices to relevant internal and external stakeholders. Furthermore, the control's ability to adapt to new risks or changes in AI system characteristics should be assessed by reviewing change management procedures and incident response logs related to severe AI consequences. Interviews with key stakeholders, including Senior Management, AI development leads, and Legal/Compliance personnel, will provide insights into the practical understanding and implementation of the framework.

Key Control Indicators (KCIs):

1. **Percentage of high-consequence AI systems (as defined by severity criteria) with formally documented and approved proportional transparency and accountability**

frameworks. (Target: 100%)
2. **Number of instances where the proportional transparency and accountability framework was not applied or was applied inadequately to a high-consequence AI system.** (Target: 0)
3. **Average time to adjust transparency and accountability practices in response to identified severe consequence changes or new high-consequence AI system deployments.** (Target: TBD, e.g., within 30 days of identification/deployment)
4. **Completion rate of scheduled annual reviews of the proportional transparency and accountability framework by Senior Management/AI Governance bodies.** (Target: 100%)
5. **Score from internal and external stakeholder feedback surveys regarding the clarity, fairness, and responsiveness of the organization's AI transparency and accountability practices for high-consequence systems.** (Target: >85% satisfaction)

Requirement Statements:

1. When consequences are severe, such as when life and liberty are at stake, AI developers and deployers should consider proportionally and proactively adjusting their transparency and accountability practices. (Page: 21)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

28. AI Training Data Intellectual Property Compliance

Control Actor: Legal Counsel

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes to ensure that all data used for training AI systems complies with applicable intellectual property rights laws, including copyright. This involves legal review, obtaining necessary licenses or permissions, and documenting data provenance to demonstrate compliance with relevant regulations and contractual obligations, thereby mitigating legal risks associated with AI system development and deployment.

Enhanced Implementation Guide:

This control ensures that all data utilized for training Artificial Intelligence (AI) systems adheres strictly to applicable intellectual property (IP) rights laws, notably copyright, thereby mitigating legal and reputational risks.

Control Statement: Systematically review, license, and document all AI training data to confirm compliance with intellectual property rights, including copyright, prior to its use in AI model development and deployment.

High-Level Control Implementation Guide:

1. **Data Ingestion and Classification:** Establish a process to identify, categorize, and inventory all potential data sources intended for AI training (e.g., internal proprietary data, licensed third-party data, publicly available data, web-scraped data).
2. **Legal Due Diligence and Review:** Conduct thorough legal reviews of all external data sources to ascertain IP ownership, usage restrictions, and licensing obligations. This includes scrutinizing terms of service, end-user license agreements (EULAs), and direct contractual agreements. For internal data, verify compliance with organizational IP policies.
3. **Licensing and Permission Acquisition:** Proactively obtain explicit licenses, permissions, or contractual agreements for any third-party data requiring them, ensuring that the scope of use explicitly permits AI training and commercial deployment where applicable. Maintain a centralized repository of all licenses and agreements, including renewal dates.
4. **Comprehensive Data Provenance Documentation:** Implement a robust system to

meticulously record the origin, legal status, specific usage rights, and any transformations applied to each dataset used for AI training. This documentation should include metadata, references to legal agreements, and records of internal approvals.
5. **Usage Restriction Enforcement:** Integrate technical and procedural controls within data pipelines and AI development environments to ensure that data is used strictly within the confines of its granted permissions or licenses, preventing unauthorized access, reuse, or distribution.
6. **Periodic Compliance Audits:** Conduct regular internal audits of AI training data inventories and their associated IP documentation to verify ongoing compliance with policies and legal requirements.
7. **Training and Awareness:** Provide mandatory and recurring training to all personnel involved in AI data procurement, preparation, development, and deployment (e.g., data scientists, AI engineers, legal, procurement teams) on IP compliance requirements specific to AI training data.

Control Frequency: Continuously for new data ingestion and prior to initial use of any dataset for AI model training. Periodically (e.g., annually or upon significant model retraining/release) for re-verification of existing datasets.

Monitoring Frequency: Ongoing by the Legal Counsel and Data Governance teams, with formal review and reporting to senior management on a quarterly basis. Internal Audit performs independent monitoring and testing annually.

Control Type: Semi-automated. Legal review, interpretation of licenses, and negotiation are manual processes. However, data provenance tracking systems, license management tools, and technical access controls contribute automated or semi-automated support to enforce compliance.

Guide for Evaluating Control Effectiveness:
Control effectiveness is evaluated based on:
* **Absence of IP Infringement Claims:** Zero valid legal claims, disputes, or cease-and-desist orders related to intellectual property infringement concerning AI training data.
* **Completeness and Accuracy of Documentation:** All AI training datasets possess complete, accurate, and readily auditable provenance documentation, including detailed licensing information and defined usage rights.
* **Adherence to Policy and Procedure:** Consistent and demonstrable adherence by AI development and data teams to established IP compliance policies and procedures for data acquisition and usage.

*   **Audit Findings:** Internal and external compliance audits consistently report no material weaknesses or significant deficiencies related to AI training data intellectual property compliance.
*   **License Management Maturity:** All required licenses and permissions are current, correctly scoped for AI training and deployment, and easily verifiable through a centralized management system.

Key Control Indicators (KCIs):
*   **Percentage of AI training datasets with complete IP provenance documentation:** Target: 100%.
*   **Number of identified IP non-compliance incidents (e.g., unauthorized data use, expired licenses discovered):** Target: 0.
*   **Average time to obtain necessary licenses/permissions for new external datasets:** Target: X business days (e.g., < 30 days).
*   **Percentage of AI development and data procurement personnel completing mandatory IP compliance training:** Target: 100%.
*   **Number of legal intellectual property infringement claims received related to AI training data:** Target: 0.

Requirement Statements:

1. Training data may also be subject to copyright and should follow applicable intellectual property rights laws. (Page: 21)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

29. AI System Transparency Tool Testing and Intended Use Verification

Control Actor: Information Security Officer

Control Types: preventive, detective, technical, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for the Information Security Officer to conduct testing of various transparency tools for AI systems. This includes cooperating with AI deployers to verify that AI systems are being used as intended and that transparency mechanisms effectively communicate relevant information to users and stakeholders, thereby enhancing trust and accountability in AI operations.

Enhanced Implementation Guide:

The organization shall establish and maintain robust processes, led by the Information Security Officer (ISO), to regularly test the effectiveness of AI system transparency tools and verify that AI systems are deployed and operated strictly for their intended purposes, ensuring all relevant information is clearly communicated to users and stakeholders to foster trust and accountability, in compliance with organizational policies and applicable AI governance frameworks. **Control Implementation Guide:** The ISO, in collaboration with AI deployers and business owners, will maintain a comprehensive inventory of AI systems and their associated transparency mechanisms (e.g., model cards, explainability interfaces, audit trails, user disclosures). For each system, a formal "intended use" document will be established, outlining scope, limitations, and ethical considerations. The ISO will develop and execute structured test plans to evaluate transparency tools for accuracy, completeness, clarity, understandability, accessibility, and timeliness of information for diverse audiences. Concurrently, regular verification will be conducted to confirm actual AI system usage aligns strictly with documented intended uses, reviewing usage logs, process documentation, and stakeholder feedback. A continuous feedback loop with users and stakeholders will be maintained to assess the practical effectiveness and perceived clarity of transparency mechanisms. All testing activities, verification outcomes, identified deficiencies, and remediation efforts will be meticulously documented and reported to relevant governance bodies (e.g., AI Governance Committee). This control integrates into the AI system lifecycle, from development to post-deployment operation. **Control Frequency:** Transparency tool testing will occur bi-annually for high-risk AI systems, annually for medium-risk systems, and upon significant model updates or changes to underlying data for all systems. Intended use verification will be performed quarterly for high-risk, semi-annually for medium-risk, and annually for low-risk AI systems, and additionally, whenever a suspected deviation or change in business context arises. **Monitoring Frequency:** The AI Governance Committee or a designated oversight body will review the ISO's testing and verification reports, remediation progress, and overall compliance posture related to AI transparency and intended use on a quarterly basis. Internal Audit will perform independent monitoring annually. **Control Type:** Semi-automated. While the ISO's oversight, analysis, and stakeholder engagement are manual, the testing process can incorporate automated tools for data validation, log analysis, and consistency checks of explanation outputs, enhancing efficiency and scalability. **Guide for Evaluating Control Effectiveness:** Effectiveness is evaluated by reviewing the completeness and quality of documented test plans, testing results, verification reports, identified non-conformities, and documented remediation actions. Independent sample testing of selected AI systems will be conducted, verifying transparency tool efficacy through user surveys, direct inspection of tool outputs, and comparison with actual model behavior/data sources. Adherence to intended use will be assessed by reviewing usage logs, business process flows, and interviewing relevant personnel. Analysis of stakeholder feedback, including user complaints related to AI system understanding, will provide qualitative input. Finally, audit trails will be reviewed to confirm that non-compliance incidents or deviations are properly logged, investigated, and resolved in a timely manner. **Key Control Indicators (Metrics):**

1. Percentage of AI systems with formally documented intended use and associated transparency tools.
2. Number/Percentage of identified transparency tool deficiencies (e.g., inaccurate, unclear information) and their remediation rate.
3. Number/Percentage of AI systems where usage deviations from intended purpose were identified, and their remediation rate.
4. Average user/stakeholder satisfaction scores regarding AI system transparency (e.g., from surveys).
5. Average time to remediate identified transparency tool issues or intended use deviations.
6. Frequency and comprehensiveness of ISO reporting to governance bodies on transparency and intended use compliance.
7. Number of formal complaints or escalations directly related to AI system understanding or perceived misuse.

This control directly addresses requirements for AI trustworthiness, accountability, and responsible deployment by ensuring systems are understandable and used within their defined boundaries, aligning with principles from frameworks like NIST AI RMF and emerging regulations such as the EU AI Act.

Requirement Statements:

1. As transparency tools for AI systems and related documentation continue to evolve, developers of AI systems are encouraged to test different types of transparency tools in cooperation with AI deployers to ensure that AI systems are used as intended. (Page: 21)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

30. AI System Privacy Values Integration and Guidance

Control Actor: Data Protection Officer

Control Types: preventive, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for the Data Protection Officer to guide choices for AI system design, development, and deployment, ensuring that privacy values such as anonymity, confidentiality, and control are embedded from inception. This includes integrating privacy considerations throughout the AI system lifecycle to protect individual data and promote responsible AI practices.

Enhanced Implementation Guide:

AI System Privacy Values Integration and Guidance

Control Statement: The Data Protection Officer (DPO) must ensure that privacy values, specifically anonymity, confidentiality, and individual control, are proactively and systematically integrated into all phases of AI system design, development, and deployment within both the Organization and its Provider Organizations, thereby embedding privacy considerations throughout the entire AI system lifecycle to safeguard individual data and foster responsible AI practices from inception.

High-Level Control Implementation Guide: The DPO will establish and maintain a robust framework for integrating privacy into AI systems. This includes: 1) Developing and mandating the use of comprehensive internal policies, standards, and guidelines for Privacy-by-Design and Privacy-by-Default principles tailored for AI systems, aligning with relevant data protection regulations (e.g., GDPR, CCPA). 2) Participating actively in the initial conceptualization, design reviews, data sourcing, model training, testing, and deployment phases of all new or significantly modified AI systems. 3) Conducting or overseeing mandatory Data Protection Impact Assessments (DPIAs) or Privacy Impact Assessments (PIAs) for each AI system, ensuring that privacy risks are thoroughly identified, assessed, and mitigated prior to system launch. 4) Providing expert guidance, consultation, and regular training sessions to AI development, engineering, product, and procurement teams on AI privacy principles, best practices, and compliance requirements, including secure data handling, anonymization techniques, and user consent mechanisms. 5) Ensuring contractual agreements with Provider Organizations explicitly define privacy responsibilities, data processing limitations, and audit rights concerning AI system development and data handling. 6) Establishing mechanisms for ongoing privacy monitoring and a feedback loop to address emerging privacy concerns related to deployed AI systems.

Control Frequency: Continuous and Event-Driven. This control must be performed at the inception of any new AI system project, upon significant changes to existing AI systems (e.g., new data sources, model architectures, use cases, or processing activities), during key design and development milestones (e.g., requirements sign-off, architectural review, pre-deployment), and reactively upon identification of new privacy risks or incidents.

Monitoring Frequency: Annually, as part of the internal audit or compliance program review cycle. Additionally, event-driven monitoring for high-risk AI initiatives, material changes, or in response to privacy incidents.

Control Type: Manual. While supporting tools may assist in documentation or risk assessment, the core activities of policy formulation, expert guidance, impact assessments, and decision-making for privacy integration are inherently human-driven and require DPO discretion.

Guide for Evaluating Control Effectiveness: To evaluate effectiveness, an auditor should: 1) Review documented evidence of DPO involvement in AI system lifecycle phases, including meeting minutes, design review sign-offs, and formal approvals for new or modified AI systems. 2) Examine a representative sample of completed DPIAs/PIAs for AI systems, verifying their thoroughness, the identification of relevant privacy risks (anonymity, confidentiality, control), and the implementation status of documented mitigation plans. 3) Assess the completeness, currency, and dissemination of internal policies, standards, and guidelines related to AI privacy, and interview relevant personnel to confirm their understanding and adherence. 4) Verify records of privacy training provided to AI development and product teams, including content relevance and attendance. 5) Review contractual agreements with Provider Organizations to confirm the inclusion of robust privacy clauses for AI data processing. 6) Analyze the root causes of any privacy incidents or breaches involving AI systems to identify potential control weaknesses.

Key Control Indicators (KCIs):
* **Percentage of new/modified AI systems with documented DPO sign-off/approval prior to deployment.**

* **Percentage of new/modified AI systems for which a comprehensive DPIA/PIA was completed and reviewed by the DPO *prior* to commencing significant development or deployment.**
* **Number of documented AI privacy-by-design guidelines or standards developed, reviewed, and disseminated by the DPO annually.**
* **Completion rate of identified privacy risk mitigation actions stemming from AI-specific DPIAs/PIAs.**
* **Number of privacy-related findings or recommendations from internal/external audits specific to AI systems (trend analysis).**
* **Percentage of relevant AI development and product staff who have completed mandatory AI privacy training facilitated or approved by the DPO.**

Requirement Statements:

1. Privacy values such as anonymity, confidentiality, and control generally should guide choices for AI system design, development, and deployment. (Page: 22)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

31. AI System Privacy-Enhanced Design and Data Minimization

Control Actor: AI Development Team

Control Types: preventive, technical, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement processes for AI Development Teams to design and develop AI systems to be privacy-enhanced from inception. This includes proactively integrating privacy-enhancing technologies (PETs) and utilizing data minimizing methods, such as de-identification and aggregation, specifically for model outputs. These measures ensure that privacy values, including anonymity, confidentiality, and control, are embedded throughout the AI system lifecycle, thereby protecting individual data and promoting responsible and trustworthy AI practices.

Enhanced Implementation Guide:

Control Statement: AI systems must be designed and developed with privacy embedded from inception, employing privacy-enhancing technologies (PETs) and rigorous data minimization techniques across the entire AI lifecycle, ensuring user data protection and alignment with responsible AI principles.

High-Level Control Implementation Guide: The AI Development Team, in collaboration with Privacy and Legal functions, shall mandate the application of privacy-by-design principles throughout the AI system development lifecycle (SDLC), from requirements gathering through deployment and maintenance. This includes conducting comprehensive Data Protection Impact Assessments (DPIAs) or Privacy Impact Assessments (PIAs) for all new or significantly modified AI systems to proactively identify and mitigate privacy risks. Furthermore, teams must implement robust data minimization strategies by designing data flows and processing to collect, use, and store only the absolute minimum amount of personal data necessary, applying techniques such as anonymization, pseudonymization, aggregation, de-identification, and synthetic data generation *before* data ingestion, during processing, and for model outputs. Proactive integration of appropriate Privacy-Enhancing Technologies (PETs), such as differential privacy for training data, homomorphic encryption, secure multi-party computation, or federated learning, is required where feasible and effective to protect data privacy while enabling AI functionality. This also necessitates establishing secure data handling practices, including robust access controls, encryption (at rest and in transit), data retention policies, and secure deletion protocols for all data used by or generated from AI systems. Finally, continuous training for AI Development Teams on privacy-by-design principles, relevant data protection regulations, and secure coding practices for AI systems is mandatory, alongside maintaining comprehensive documentation of all design decisions, PETs implemented, data minimization techniques applied, DPIA/PIA outcomes, and privacy control configurations.

Control Frequency: This control is **continuous** during the design and development phases of any new or significantly modified AI system, with specific activities (like DPIAs) performed at defined project milestones.

Monitoring Frequency: Quarterly reviews of AI development projects and **annually** for established AI systems, supplemented by ad-hoc reviews triggered by significant changes or incidents, will be conducted to assess adherence.

Control Type: This control is **Semi-Automated**, encompassing administrative processes (DPIAs, documentation, training), manual design considerations and review, and technical implementations (PETs integration, secure coding practices, automated data minimization scripts, and configuration of security controls).

Guide for Evaluating Control Effectiveness: Control effectiveness will be evaluated through: 1) Review of design artifacts (e.g., architecture diagrams, data flow maps, technical specifications) to verify explicit incorporation of privacy-by-design and data minimization; 2) Assessment of the completion rate and quality of DPIAs/PIAs, ensuring identified risks have appropriate, tracked mitigation plans; 3) Technical verification via audits, code reviews, and penetration tests to confirm the proper implementation and effectiveness of PETs and data minimization techniques (e.g., confirming de-identified outputs meet defined thresholds); 4) Confirmation of adherence to internal policies related to privacy-by-design and data handling in AI; 5) Verification of AI Development Teams' completion of required privacy and secure AI development training; and 6) Analysis of privacy-related incidents or data breaches originating from AI systems to identify control weaknesses and areas for improvement.

Key Control Indicators (KCIs):

1. **Percentage of New/Modified AI Systems with Completed DPIA/PIA:** Target 100%.
2. **Percentage of AI Systems Demonstrating Application of at least one PET or Advanced Data Minimization Technique:** (e.g., Differential Privacy, Federated Learning, Homomorphic Encryption, K-anonymity, L-diversity).
3. **Number of Privacy-Related Findings/Vulnerabilities in AI System Audits/Penetration

Tests: Aim for a decreasing trend or zero findings related to data privacy.
4. **AI Development Team Privacy Training Completion Rate:** Target >95%.
5. **Ratio of Anonymized/Pseudonymized Data to Raw Personal Data Used in AI Systems:** Track the proportion of data processed through minimization techniques versus original PII to demonstrate effective data minimization.

Requirement Statements:

1. Privacy-enhancing technologies ("PETs") for AI, as well as data minimizing methods such as de-identification and aggregation for certain model outputs, can support design for privacy-enhanced AI systems. (Page: 22)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

32. AI Bias Management Program and Lifecycle Integration

Control Actor: AI Governance Team

Control Types: preventive, administrative, detective, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement a comprehensive program for identifying, assessing, and managing AI bias throughout the entire AI system lifecycle. This program shall specifically address systemic, computational and statistical, and human-cognitive categories of bias, ensuring processes are in place for design, implementation, operation, maintenance, and system use. The objective is to mitigate the increased speed and scale of biases and prevent the perpetuation and amplification of harms to individuals, groups, communities, organizations, and society, even in the absence of discriminatory intent. This includes establishing organizational norms and practices to address bias in datasets and algorithmic processes, and accounting for human-cognitive biases in decision-making processes.

Enhanced Implementation Guide:

The AI Bias Management Program and Lifecycle Integration control mandates the establishment, implementation, and continuous maintenance of a comprehensive program designed to systematically identify, assess, mitigate, and monitor systemic, computational, statistical, and human-cognitive biases throughout the entire AI system lifecycle, encompassing design, implementation, operation, maintenance, and system use. This program's objective is to proactively prevent the perpetuation and amplification of harms to individuals, groups, communities, organizations, and society, even in the absence of discriminatory intent, by establishing organizational norms and practices that effectively address bias in datasets, algorithmic processes, and human-cognitive decision-making.

High-Level Control Implementation Guide: The AI Governance Team shall (1) develop and approve a formal AI Bias Management Policy and Framework, clearly defining roles, responsibilities, and accountability across all relevant teams; (2) integrate mandatory AI bias risk assessments and mitigation strategies into each phase of the AI system development lifecycle (SDLC), from initial concept to retirement; (3) implement and enforce the use of established bias detection methodologies and tools (e.g., fairness metrics, explainability techniques) for datasets and algorithmic outputs; (4) establish clear protocols for the selection and application of bias mitigation techniques (e.g., data re-balancing, model recalibration, human-in-the-loop interventions); (5) implement continuous monitoring mechanisms for deployed AI systems to detect emergent biases; (6) provide mandatory, regular training on AI bias awareness, identification, and mitigation techniques to all personnel involved in AI development, deployment, and oversight; (7) ensure comprehensive documentation of all bias assessments, mitigation efforts, monitoring results, and decision-making processes; and (8) establish a formal incident response process for reported AI bias events, including investigation, remediation, and root cause analysis.

Control Frequency: The comprehensive AI Bias Management Program itself shall be formally reviewed and updated at least Annually, or more frequently upon significant changes in AI technologies, organizational risk profiles, regulatory requirements, or observed bias trends.

Monitoring Frequency: The effectiveness of the AI Bias Management Program shall be monitored Continuously by the AI Governance Team through operational metrics and performance reviews, with formal independent monitoring and assurance activities conducted at least Quarterly by Internal Audit or Compliance functions.

Control Type: Semi-automated, as it involves a structured administrative framework and policies (manual) leveraging automated tools for bias detection, measurement, and continuous monitoring within AI systems.

Guide for Evaluating Control Effectiveness: To evaluate effectiveness, examine: (1) evidence of a formally approved and disseminated AI Bias Management Policy and Framework; (2) documented integration of bias risk assessments and mitigation steps into the AI SDLC for all new and existing AI systems; (3) records of bias assessments and mitigation actions taken for a representative sample of AI models, demonstrating thoroughness and adherence to established protocols; (4) completion rates and comprehension levels of mandatory AI bias training for relevant personnel; (5) analysis of incident reports related to AI bias, assessing the completeness, timeliness, and effectiveness of corrective actions; (6) results of independent fairness audits or red-teaming exercises conducted on AI systems; and (7) feedback from affected stakeholders regarding perceived fairness and non-discrimination.

Key Control Indicators (KCIs):
* **Percentage of AI Systems with Completed Bias Assessments:** Ratio of AI systems that have undergone formal bias assessments at relevant lifecycle stages.
* **Number of Identified Bias Incidents:** Count of confirmed bias incidents or near-misses reported per quarter/year.

*   **Average Time to Resolution for Bias Incidents:** The mean duration from bias identification to verified remediation.
*   **Percentage of Relevant Personnel Trained:** Proportion of AI developers, operators, and decision-makers who have completed mandatory AI bias training.
*   **Trend in Fairness Metric Performance:** Longitudinal tracking of key fairness metrics (e.g., statistical parity, equalized odds, demographic parity) for critical AI systems, indicating improvement or stability.
*   **Number of Policy/Procedure Updates:** Count of substantive revisions to AI bias policies or related procedures, reflecting continuous improvement.
*   **Audit Findings Rate:** Number of internal/external audit findings related to AI bias control deficiencies per audit cycle.
*   **Stakeholder Feedback on Fairness:** Qualitative and quantitative measures from user surveys or feedback channels regarding perceived fairness and non-discriminatory outcomes from AI systems.

Requirement Statements:

1. NIST has identified three major categories of AI bias to be considered and managed: systemic, computational and statistical, and human-cognitive. Each of these can occur in the absence of prejudice, partiality, or discriminatory intent. Systemic bias can be present in AI datasets, the organizational norms, practices, and processes across the AI lifecycle, and the broader society that uses AI systems. Computational and statistical biases can be present in AI datasets and algorithmic processes, and often stem from systematic errors due to non-representative samples. Human-cognitive biases relate to how an individual or group perceives AI sys- tem information to make a decision or fill in missing information, or how humans think about purposes and functions of an AI system. Human-cognitive biases are omnipresent in decision-making processes across the AI lifecycle and system use, including the design, implementation, operation, and maintenance of AI. While bias is not always a negative phenomenon, AI sys- tems can potentially increase the speed and scale of biases and perpetuate and amplify harms to individuals, groups, communities, organizations, and society. (Page: 23)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ______________________________

Evidence/Comments:

____

____

____

33. AI RMF Effectiveness and Trustworthiness Impact Evaluation

Control Actor: NIST Leadership

Control Types: administrative, detective

Related Entities: NIST

Control Description:

Establish and maintain processes for NIST to evaluate the overall effectiveness of the AI Risk Management Framework (AI RMF). This includes developing and applying metrics, methodologies, and goals to measure bottom-line improvements in the trustworthiness of AI systems as a direct result of AI RMF implementation. These evaluations should be conducted in conjunction with the broader AI community, with results and supporting information broadly shared, to ensure comprehensive assessment and foster continuous improvement of the framework's impact on trustworthy AI development and deployment.

Enhanced Implementation Guide:

This control requires NIST to establish, maintain, and periodically execute a structured process to evaluate the overall effectiveness and impact of the AI Risk Management Framework (AI RMF) on the trustworthiness of AI systems, ensuring broad community engagement and transparent dissemination of findings. **Implementation Guide:** NIST Leadership will designate a responsible team to define comprehensive evaluation methodologies, including the identification of key performance indicators (KPIs) for AI system trustworthiness (e.g., fairness, robustness, transparency, privacy, security) and the setting of baseline metrics. A strategic plan for continuous engagement with the broader AI community (industry, academia, government, civil society) must be developed and executed, involving surveys, workshops, public calls for information, and collaborative partnerships. Periodic evaluations (e.g., Annually or Bi-annually) will be conducted, collecting data on AI system trustworthiness attributes, assessing the adoption and perceived utility of the AI RMF by stakeholders, and measuring its direct impact on mitigating AI risks. Comprehensive evaluation reports, summarizing findings, successes, and areas for improvement, will be prepared and broadly shared with the AI community via official channels to foster transparency and continuous improvement of the framework. **Control Frequency:** Annually or Bi-annually, with ad-hoc evaluations as needed for significant framework updates or major industry shifts. **Monitoring Frequency:** Quarterly by NIST's internal governance bodies to track progress on planned evaluations, stakeholder engagement, and dissemination efforts. **Control Type:** Semi-automated (data collection and preliminary analysis may leverage automated tools, but strategic decision-making, stakeholder engagement, and final reporting are manual processes). **Evaluating Control Effectiveness:** Effectiveness is evaluated by verifying the existence and adherence to formally documented evaluation methodologies, metrics, and goals. This includes assessing the completeness and quality of collected evaluation data against defined KPIs and stakeholder feedback. Confirmation of active community engagement mechanisms (e.g., documented workshops, public comment analysis, survey participation rates) is crucial. Furthermore, the quality, clarity, and comprehensive dissemination of evaluation reports must be reviewed, ensuring public accessibility. Finally, the extent to which evaluation findings lead to documented recommendations for AI RMF enhancements or related guidance updates will be assessed. **Key Control Indicators (KCIs):** Number of formal AI RMF evaluations completed per period; Percentage of defined AI RMF effectiveness KPIs measured and reported; Number of unique external stakeholders engaged in evaluations (e.g., organizations, individuals); Volume and diversity of feedback received from the AI community on the AI RMF's effectiveness; Number of AI RMF revisions or supplementary guidance documents directly informed by evaluation findings; Public accessibility index of evaluation reports and supporting data; and trends in reported AI system trustworthiness attributes from organizations adopting the AI RMF, where measurable.

Requirement Statements:

1. Effectiveness of the AI RMF Evaluations of AI RMF effectiveness – including ways to measure bottom-line improve- ments in the trustworthiness of AI systems – will be part of future NIST activities, in conjunction with the AI community. (Page: 24)

2. NIST intends to work collaboratively with others to develop met- rics, methodologies, and goals for evaluating the AI RMF’s effectiveness, and to broadly share results and supporting information. (Page: 24)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

34. Organizational AI RMF Risk Management Improvement Evaluation

Control Actor: Senior Management

Control Types: administrative, detective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for Senior Management to periodically evaluate whether the AI Risk Management Framework (AI RMF) has effectively improved the organization's ability to manage AI risks. This evaluation should encompass an assessment of their policies, processes, practices, implementation plans, indicators, measurements, and expected outcomes, ensuring continuous improvement in AI risk management capabilities.

Enhanced Implementation Guide:

Control Statement: Senior Management must establish and maintain a structured, periodic process to thoroughly evaluate the overall effectiveness of the AI Risk Management Framework (AI RMF) in enhancing the organization's capability to identify, assess, mitigate, monitor, and report AI-related risks. This evaluation must systematically encompass all foundational elements of the AI RMF, including policies, processes, practices, implementation plans, performance indicators, measurement methodologies, and achieved outcomes, to ensure demonstrable continuous improvement in AI risk management.

Control Implementation Guide: Senior Management, potentially through a designated AI Governance Committee or equivalent steering body, shall implement this control by:
1. **Defining Scope and Criteria:** Clearly outlining the specific policies, processes, practices, implementation plans, indicators, measurements, and expected outcomes of the AI RMF to be evaluated. This includes assessing their adequacy, operationalization, and effectiveness.
2. **Establishing a Review Cadence:** Scheduling formal review sessions at predetermined intervals (e.g., bi-annually, annually for formal comprehensive review).
3. **Information Gathering:** Collating relevant information prior to the evaluation, including:
    * AI risk assessments and profiles.
    * AI incident reports, near misses, and post-mortems.
    * Results of AI RMF internal audits, compliance reviews, and independent assessments.
    * Feedback from AI development teams, risk owners, and business units.
    * AI RMF performance metrics and Key Risk Indicators (KRIs).

    * Updates on regulatory requirements and industry best practices related to AI risk.
4. **Conducting Structured Evaluation:** Performing a systematic review and discussion of the gathered information against predefined effectiveness criteria. This should include:
    * Assessing the clarity and comprehensiveness of AI risk policies.
    * Evaluating the efficiency and effectiveness of AI risk processes (e.g., risk identification, assessment, mitigation, monitoring).
    * Reviewing the implementation status and adoption of AI risk management practices.
    * Analyzing the utility and accuracy of performance indicators and measurements.
    * Comparing actual outcomes against expected outcomes to identify variances and root causes.
5. **Documenting Findings and Action Plans:** Recording all evaluation findings, identified gaps, weaknesses, and improvement opportunities. Developing concrete, prioritized action plans with clear ownership, timelines, and measurable objectives for addressing identified issues and enhancing the AI RMF.
6. **Reporting and Communication:** Formally communicating evaluation results and action plans to relevant stakeholders, including the Board of Directors, executive leadership, and AI risk owners.
7. **Follow-up and Verification:** Ensuring that agreed-upon action plans are implemented and their effectiveness is subsequently verified in future evaluation cycles.

Control Frequency: Bi-Annually, with a comprehensive formal evaluation conducted at least Annually.

Monitoring Frequency: Annually, typically by Internal Audit or a dedicated GRC function, to confirm Senior Management's adherence to the established evaluation process and documentation requirements.

Control Type (Manual/Automated/Semi-Automated): Manual. While data inputs for the evaluation (e.g., performance metrics, incident reports) may be automated or semi-automated, the core evaluation, judgmental assessment, decision-making, and strategic formulation of improvement plans are inherently manual activities performed by Senior Management.

Guide for Evaluating Control Effectiveness: To evaluate the effectiveness of this control, auditors or reviewers should:
1. **Verify Execution:** Confirm that evaluation sessions have been held at the prescribed frequency and that comprehensive meeting minutes, evaluation reports, or formal reviews are documented and approved by Senior Management.
2. **Assess Scope Coverage:** Validate that the evaluations encompass all required elements

of the AI RMF (policies, processes, practices, implementation plans, indicators, measurements, and outcomes) as per the control description.

3. **Review Quality of Assessment:** Assess the depth, rigor, and objectivity of the evaluations. Look for evidence of critical analysis, root cause identification, and a forward-looking perspective on continuous improvement.

4. **Trace Actionable Outcomes:** Confirm that the evaluations lead to specific, measurable, achievable, relevant, and time-bound (SMART) action plans for addressing identified deficiencies or enhancing the AI RMF.

5. **Validate Implementation and Impact:** Track the implementation status of previous action plans and assess their impact on the organization's AI risk posture and management capabilities. Look for demonstrable improvements (e.g., reduced AI risk exposure, fewer critical incidents, enhanced compliance).

6. **Review Stakeholder Engagement:** Confirm appropriate engagement of key stakeholders (e.g., AI development teams, legal, compliance, business units) in providing input to the evaluation.

Key Control Indicators (KCIs):

* **% of scheduled AI RMF effectiveness evaluations completed within the defined frequency.** (Process adherence)
* **Number of high-priority AI RMF gaps or improvement opportunities identified per evaluation cycle.** (Effectiveness of identification)
* **% of high-priority AI RMF improvement action plans closed within target timelines.** (Timeliness of remediation)
* **Trend in the organization's AI Risk Maturity Score (if a formal maturity model is adopted and measured).** (Overall framework effectiveness)
* **Number of critical or high-severity AI incidents attributed to gaps in the AI RMF, tracked over time.** (Effectiveness in preventing adverse events)
* **Trend in internal audit findings or regulatory observations related to AI risk management.** (External validation of effectiveness)
* **Number of substantive updates or revisions to AI policies, processes, or guidelines directly resulting from evaluation findings.** (Responsiveness and continuous improvement)

Requirement Statements:

1. Organizations and other users of the Framework are encouraged to periodically evaluate whether the AI RMF has improved their ability to manage AI risks, including but not lim- ited to their policies, processes, practices, implementation plans, indicators, measurements, and expected outcomes. (Page: 24)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ______________________________

Evidence/Comments:

____

____

____

35. Organizational Culture for AI Risk Prioritization and Management

Control Actor: Senior Management

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish, continuously enhance, and robustly implement an organizational culture that deeply prioritizes the identification, assessment, and management of AI system risks and their potential impacts on individuals, communities, organizations, and society. This involves fostering a collective mindset and practices across all levels of the organization to proactively address AI-related harms and ensure responsible AI system development and deployment.

Enhanced Implementation Guide:

Establish and sustain an organizational culture that embeds the proactive identification, comprehensive assessment, and effective management of AI system risks as a fundamental principle across all organizational levels, ensuring responsible AI development and deployment and mitigating potential adverse impacts on individuals, communities, and society. **Implementation Guide:** Senior Management must visibly champion responsible AI, articulating its strategic importance and leading by example. This involves developing and disseminating clear, actionable AI ethics and risk management policies and procedures, integrating them into existing governance frameworks. Implement mandatory and recurring training programs for all relevant employees on AI ethics, responsible innovation, risk identification, and mitigation strategies. Establish clear, confidential, and non-retaliatory reporting channels for AI-related concerns or incidents. Integrate AI risk management responsibilities and accountability into performance reviews and incentivize adherence to responsible AI practices. Foster cross-functional collaboration between engineering, product, legal, ethics, and business units to ensure holistic risk identification and mitigation. Regularly communicate the organization's commitment to responsible AI through internal channels and external reporting. **Control Frequency:** This control is continuous in its nature, requiring ongoing reinforcement and adaptation. Core foundational activities such as policy establishment and initial training rollouts are typically annual or biennial, while cultural reinforcement, communication, and risk identification processes are ongoing. **Monitoring Frequency:** Senior Management should conduct a formal review of cultural effectiveness and AI risk posture quarterly, with an independent cultural assessment or audit performed annually to provide an objective evaluation. **Control Type:** This is predominantly a **manual** control, driven by human leadership, policy enforcement, training, and behavioral norms, with **semi-automated** elements supporting its execution (e.g., online training platforms, policy management systems, incident reporting tools). **Evaluating Control Effectiveness:** Effectiveness can be evaluated through a combination of qualitative and quantitative measures. Qualitatively, this involves conducting employee surveys to gauge awareness of AI ethics policies, perception of psychological safety for reporting concerns, and confidence in the organization's commitment to responsible AI. Focus groups and interviews with diverse employee groups can provide deeper insights. Quantitatively, evaluate the completion rates of mandatory AI ethics and risk training; track the percentage of new or materially modified AI systems undergoing formal risk assessments; monitor the volume and resolution rates of reported AI-related concerns/incidents; assess the integration of AI risk gates into project management lifecycles; and review findings from internal audits related to AI governance and ethics. **Key Control Indicators (KCIs):**

1. **AI Ethics Training Completion Rate:** Percentage of employees completing mandatory AI ethics and risk training within specified deadlines.
2. **AI Risk Assessment Coverage:** Percentage of new or significantly updated AI systems that undergo a formal, documented risk assessment prior to deployment or material change.
3. **Employee AI Ethics Awareness Score:** Average score derived from periodic internal surveys measuring employee understanding of responsible AI principles and risk reporting mechanisms.
4. **AI-Related Concern Reporting Trend:** The number and qualitative nature of AI risk or ethics concerns reported through official channels over time, indicating a healthy reporting culture.
5. **Leadership Communication Cadence:** Frequency and consistency of C-suite communications on responsible AI strategy and values.
6. **Responsible AI Policy Adherence Rate:** Score or percentage based on internal audit findings assessing compliance with established AI governance and ethics policies.

Requirement Statements:

1. enhanced organizational culture which prioritizes the identification and management of AI system risks and potential impacts to individuals, communities, organizations, and society; (Page: 24)

2. cultivates and implements a culture of risk management within organizations design- ing, developing, deploying, evaluating, or acquiring AI systems; (Page: 26)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

36. Senior Management AI Risk Information Sharing and Collaboration

Control Actor: Senior Management

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for Senior Management to strengthen engagement with interested parties and relevant AI actors, to improve information sharing within and across organizations about AI risks, decision-making processes, responsibilities, common pitfalls, Test, Evaluation, Verification, and Validation (TEVV) practices, and approaches for continuous improvement, and to actively contribute suggestions related to AI risk management for sharing with the broader community. This ensures transparent communication and facilitates collective learning and adaptation in AI risk management, fostering a collaborative approach to responsible AI development and deployment.

Enhanced Implementation Guide:

This control mandates that Senior Management shall establish, maintain, and actively participate in formal processes for transparent and collaborative information sharing on AI risks, decision-making, responsibilities, common pitfalls, Test, Evaluation, Verification, and Validation (TEVV) practices, and approaches for continuous improvement. This includes fostering engagement and sharing insights both internally across organizational functions and externally with interested parties and relevant AI actors (e.g., regulators, industry bodies, academic institutions, technology providers). The objective is to facilitate collective learning, enhance adaptive risk management strategies, and actively contribute to the broader community's understanding and development of responsible AI.

Control Implementation Guide:

1. **Establish Governance Frameworks:** Define and document a clear charter for AI risk governance, including establishing dedicated AI risk committees, cross-functional working groups, or senior leadership forums with specific mandates for AI risk oversight and information exchange.
2. **Identify and Engage Stakeholders:** Map key internal stakeholders (e.g., Legal, Compliance, Product, Engineering, Data Science, Ethics Office) and external AI actors (e.g., industry consortia, regulatory bodies, academic researchers, strategic partners). Develop a

strategy for regular engagement with each group.
3. **Define Information Sharing Mechanisms:** Implement structured channels and tools for sharing information, such as recurring meeting schedules (e.g., quarterly AI risk council meetings, monthly technical deep-dives), secure collaborative platforms, standardized reporting templates for AI incidents, TEVV outcomes, and lessons learned.
4. **Promote Proactive Communication:** Encourage Senior Management to actively participate in discussions, share insights from their respective domains, and foster an environment of psychological safety where risks, challenges, and pitfalls related to AI development and deployment can be openly discussed without fear of reprisal.
5. **Facilitate External Contributions:** Create a process for consolidating internal expertise and insights into actionable suggestions or best practices for the broader AI risk management community (e.g., contributing to industry standards, white papers, policy recommendations, or open-source initiatives).
6. **Continuous Improvement Loop:** Establish a feedback mechanism to regularly review the effectiveness of information sharing processes, adapt to emerging AI risks, and integrate new learnings into organizational practices.

Control Frequency: Ongoing, with formal internal Senior Management reviews occurring at least Quarterly, and formal external engagement strategy reviews and contributions occurring at least Annually, supplemented by ad-hoc engagements as new risks or opportunities emerge.

Monitoring Frequency: Annually by the Internal Audit or Compliance function, with Quarterly reviews by the designated AI Governance Committee or Risk Committee to ensure adherence to established processes and effectiveness of outcomes.

Control Type: Manual. While digital tools may support communication and documentation, the core activities of strategic engagement, collaborative decision-making, and interpretive information sharing by Senior Management are inherently manual.

Guide for Evaluating Control Effectiveness:
1. **Documentation Review:** Verify the existence and regularity of documented processes, charters, meeting minutes, action items, and records of internal/external engagements. Confirm that AI risk registers are updated based on shared insights.
2. **Stakeholder Interviews:** Conduct interviews with Senior Management and key internal/external stakeholders to assess their understanding of the processes, active participation levels, and perceived value of information shared. Evaluate whether shared insights are considered actionable and impactful.

3. **Outcome & Impact Assessment:** Examine evidence of demonstrable improvements in the organization's AI risk posture, decision-making quality, and adaptive capacity due to shared intelligence. Review the quantity and quality of organizational contributions to the broader AI risk management community. Assess whether documented AI pitfalls or TEVV learnings have led to proactive adjustments in AI development or deployment practices.
4. **Completeness and Timeliness:** Confirm that all identified critical stakeholders are engaged as per the defined strategy and that communications occur at the established frequency. Evaluate the timeliness of addressing action items derived from collaboration.

Key Control Indicators (KCIs):

- **Quantitative:**
    - Number of formal internal AI risk forums/meetings held per period vs. planned.
    - Percentage of target Senior Management attendance at internal and external AI risk forums.
    - Number of AI risk insights, lessons learned, or TEVV findings formally documented and shared internally per period.
    - Number of AI risk incidents or near-misses proactively identified and/or mitigated through shared intelligence or collaborative action.
    - Number of formal contributions (e.g., white papers, policy comments, industry standards participation) made by the organization to the broader AI risk management community per year.
    - Number of documented AI risk management process improvements directly attributable to insights gained from collaboration.
- **Qualitative:**
    - Average satisfaction scores from participant surveys on the effectiveness and value of AI risk information sharing processes.
    - Qualitative assessment of the depth, relevance, and actionability of shared AI risk insights by an independent reviewer (e.g., Internal Audit).
    - Evidence of improved strategic alignment on AI risk posture across various organizational functions.
    - External recognition or citations for the organization's contributions to responsible AI risk management practices.

Requirement Statements:

1. better information sharing within and across organizations about risks, decision- making processes, responsibilities, common pitfalls, TEVV practices, and approaches for continuous improvement; (Page: 24)

2. strengthened engagement with interested parties and relevant AI actors; (Page: 24)

3. and contribute their suggestions for sharing with the broader community. (Page: 26)

4. Processes are in place for robust engagement with relevant AI actors. (Page: 29)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ___________________________________

Evidence/Comments:

____

____

____

37. AI RMF Core Functions Diverse Perspective Integration

Control Actor: Risk Management Team

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for the Risk Management Team to ensure that the implementation of AI Risk Management Framework (RMF) Core functions (GOVERN, MAP, MEASURE, MANAGE) incorporates diverse and multidisciplinary perspectives. This includes actively seeking and integrating views from AI actors both within and outside the organization to enhance the comprehensiveness and effectiveness of AI risk assessment and management.

Enhanced Implementation Guide:

The organization shall establish and maintain documented processes to ensure that the implementation of the AI Risk Management Framework (RMF) Core Functions (GOVERN, MAP, MEASURE, MANAGE) systematically incorporates diverse and multidisciplinary perspectives from both internal and external AI actors to enhance the comprehensiveness, validity, and effectiveness of AI risk assessment and management.

High-Level Control Implementation Guide: The Risk Management Team (RMT) is responsible for developing a formal policy and procedure that mandates the systematic integration of diverse perspectives across all AI RMF core functions. This policy should clearly define what constitutes "diverse perspectives" (e.g., technical, legal, ethical, business, privacy, cybersecurity, user experience, domain-specific, social impact) and "AI actors" (e.g., internal AI developers, data scientists, legal counsel, ethicists, compliance officers, business unit owners, external consultants, academic researchers, affected community representatives). The RMT will identify relevant internal and external stakeholders for each AI RMF activity and establish formal mechanisms for soliciting their input, such as dedicated cross-functional working groups, structured workshops, independent expert consultations, surveys, and documented challenge sessions. All perspectives sought, methods of integration, and the demonstrable influence of these perspectives on AI risk assessments, mitigation strategies, and RMF decisions must be meticulously documented for auditability. The RMT will also ensure relevant personnel receive training on the importance and methods of effective diverse perspective integration, and the overall process will be periodically reviewed and updated to maintain its relevance and effectiveness.

Control Frequency: The initial process establishment is a one-time activity upon AI RMF implementation. Ongoing integration of diverse perspectives will occur continuously as AI RMF core functions are executed for new or evolving AI systems, or as per scheduled RMF review cycles (e.g., quarterly, annually, or upon significant AI system changes). The process itself will be reviewed and updated at least annually, or upon significant regulatory or organizational changes.

Monitoring Frequency: Monitoring of this control will be conducted quarterly as part of the overall AI RMF effectiveness review by an independent internal audit or compliance function.

Control Type: This control is **Semi-Automated**. While the core process elements of stakeholder identification, consultation, and qualitative input synthesis remain manual, technological tools (e.g., GRC platforms, collaboration software, document management systems) can be leveraged to semi-automate aspects such as stakeholder mapping, scheduling of consultations, distribution of information, collection of structured feedback, tracking of input integration status, and maintaining an auditable trail of decisions influenced by diverse perspectives.

Guide for Evaluating Control Effectiveness: To evaluate effectiveness, an auditor should: (1) Review a sample of AI risk assessment reports, RMF implementation plans, and decision records to verify that diverse perspectives were actively sought, documented, and demonstrably considered. Evidence should include meeting minutes, stakeholder consultation logs, and specific changes or additions to risk profiles/mitigation actions stemming from diverse input. (2) Conduct interviews with members of the Risk Management Team, relevant internal AI actors (e.g., development leads, legal, ethics), and where feasible, external consultants/partners, to confirm their understanding of the process and their participation. (3) Trace selected AI risk assessments or RMF decisions to confirm how input from various perspectives influenced the final outcome. (4) Assess adherence to the established policy/procedure for diverse perspective integration, including completeness and relevance of identified "diverse perspectives" for the specific AI systems and risks being assessed.

Key Control Indicators (KCIs):
* **Percentage of AI RMF activities (e.g., risk assessments, control definitions) that document diverse perspective integration:** Target: 100%.
* **Number of distinct stakeholder groups involved in AI RMF core function activities per major AI system or RMF cycle:** Target: >X distinct groups (as defined by policy, e.g., legal,

ethics, technical, business, privacy, external expert).
* **Frequency of formal stakeholder consultation sessions for AI RMF activities:** Target: Minimum X sessions per quarter or per major AI system development phase.
* **Number of identified risks, new mitigation strategies, or policy amendments demonstrably attributable to diverse perspective input:** (Qualitative and quantitative tracking, e.g., review comments implemented).
* **Average feedback score from participating stakeholders on the clarity and effectiveness of the perspective integration process:** Target: >4 out of 5 (via periodic surveys).
* **Number of non-compliance findings or material deficiencies related to diverse perspective integration during internal or external audits:** Target: 0.

Requirement Statements:

1. AI RMF Core functions should be carried out in a way that reflects diverse and multidisciplinary perspectives, potentially including the views of AI actors out- side the organization. (Page: 25)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

38. AI RMF Profile Development and Tailored Guidance Creation

Control Actor: Risk Management Team

Control Types: administrative, preventive, detective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for the Risk Management Team to create tailored guidance, including AI RMF Current Profiles and Target Profiles, and use-case profiles, from the NIST AI RMF Playbook's suggested material for internal organizational use. Current Profiles indicate the current state of AI risk management and related outcomes, while Target Profiles indicate the desired outcomes and goals for AI risk management within the organization's context. These profiles and tailored guidance should be specific to particular settings or applications, considering requirements, risk tolerance, resources, legal/regulatory requirements, best practices, and risk management priorities. This process includes comparing AI RMF Current and Target Profiles to identify gaps that need to be addressed to meet AI risk management objectives, thereby ensuring continuous improvement and effective risk management. This ensures that the AI RMF principles and practices are effectively adapted and integrated into the organization's specific operational context, facilitating practical implementation and adherence to trustworthy AI principles and effective AI risk management.

Enhanced Implementation Guide:

The Risk Management Team shall establish, implement, and maintain a formalized process for developing and updating tailored AI Risk Management Framework (AI RMF) guidance, including current-state (Current Profiles), desired-state (Target Profiles), and use-case specific profiles, derived from NIST AI RMF Playbook material. This process must ensure that guidance is contextualized to the organization's specific operational environment, risk tolerance, resources, legal/regulatory requirements, and strategic priorities, with a continuous mechanism for comparing Current and Target Profiles to identify and prioritize AI risk management gaps for remediation and continuous improvement. The high-level implementation guide involves the Risk Management Team defining a formal methodology for AI RMF profile development, including templates for Current, Target, and Use-Case Profiles, aligning with NIST AI RMF Functions (Govern, Map, Measure, Manage). This methodology should specify data collection procedures for assessing current AI risk posture, criteria for defining desired target states based on organizational context and regulatory obligations, and a structured process for conducting gap analyses. Implementation includes assigning clear roles and responsibilities within the Risk Management Team, establishing cross-functional working groups (e.g., with legal, technical, business units) for input and validation, and formally documenting all profiles and tailored guidance. The Risk Management Team will regularly review and update these profiles and guidance based on new AI initiatives, evolving risks, and regulatory changes, ensuring communication to relevant stakeholders. This control should be performed Annually, or upon introduction of a new significant AI system/use-case, major regulatory changes impacting AI, or significant shifts in organizational risk appetite. The control's effectiveness should be monitored Quarterly by Internal Audit or a dedicated Compliance Oversight function. This control is classified as Semi-Automated, primarily leveraging manual processes supported by structured documentation tools and potentially GRC platforms for tracking. Effectiveness is evaluated by verifying that formal AI RMF Current, Target, and Use-Case Profiles, along with tailored guidance, exist for all in-scope AI systems/use-cases and are readily accessible. This includes reviewing documentation for completeness, accuracy, and alignment with the organization's operational context, risk tolerance, and regulatory requirements. Assessors should confirm that a defined process for comparing Current and Target Profiles is consistently applied, leading to documented gap analyses and actionable remediation plans. Evidence of cross-functional stakeholder engagement (e.g., meeting minutes, sign-offs) during profile development and review should be present. Additionally, verify that profiles and guidance are formally reviewed and updated according to the prescribed frequency and that identified gaps are systematically tracked for resolution. Key control indicators include the percentage of in-scope AI systems/use-cases with formally documented and approved Current, Target, and Use-Case Profiles; the number of identified AI RMF gaps from profile comparisons, tracked against remediation plans and completion rates; the timeliness of profile and guidance updates (e.g., adherence to annual review schedule); the number of material AI RMF non-conformances or audit findings directly attributable to inadequate profiling or guidance; and feedback scores from relevant stakeholders (e.g., business units, technical teams) on the clarity, utility, and practicality of tailored AI RMF guidance.

Requirement Statements:

1. Playbook users can create tailored guidance selected from suggested material for their own use (Page: 26)

2. AI RMF use-case profiles are implementations of the AI RMF functions, categories, and subcategories for a specific setting or application based on the requirements, risk tolerance, and resources of the Framework user: for example, an AI RMF hiring profile or an AI RMF fair housing profile. Profiles may illustrate and offer insights into how risk can be managed at

various stages of the AI lifecycle or in specific sector, technology, or end-use applications. AI RMF profiles assist organizations in deciding how they might best manage AI risk that is well-aligned with their goals, considers legal/regulatory requirements and best practices, and reflects risk management priorities. (Page: 38)

3. AI RMF temporal profiles are descriptions of either the current state or the desired, target state of specific AI risk management activities within a given sector, industry, organization, or application context. An AI RMF Current Profile indicates how AI is currently being managed and the related risks in terms of current outcomes. A Target Profile indicates the outcomes needed to achieve the desired or target AI risk management goals. (Page: 38)

4. Comparing Current and Target Profiles likely reveals gaps to be addressed to meet AI risk management objectives. (Page: 38)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

39. AI Governance Structure and Foundational Outcomes Establishment

Control Actor: AI Governance Body

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain a robust governance structure and institute the foundational outcomes defined within the GOVERN function of the AI Risk Management Framework. This structure ensures that AI risk management functions align with organizational principles, policies, and strategic priorities, and explicitly connects the technical aspects of AI system design and development with these organizational values and principles. This serves as a prerequisite for subsequent AI risk management activities (MAP, MEASURE, MANAGE), providing a clear and systematic starting point for comprehensive AI risk management within the organization, deeply integrating it with overall corporate objectives.

Enhanced Implementation Guide:

Enhanced Control Description: AI Governance Structure and Foundational Outcomes Establishment

Control Statement: The organization shall establish, formally document, and continuously maintain a robust AI governance structure and institute foundational AI risk management outcomes that explicitly align with organizational principles, strategic priorities, ethical values, and existing enterprise risk management frameworks. This control ensures a systematic and integrated starting point for comprehensive AI risk management by deeply connecting the technical aspects of AI system design and development with overarching corporate objectives, thereby serving as a prerequisite for subsequent AI risk management activities (MAP, MEASURE, MANAGE).

High-Level Control Implementation Guide:
The AI Governance Body (Control Actor) will implement this control through a structured, multi-phase approach. **Phase 1: Structure Definition** involves formally chartering the AI Governance Body, defining its composition (e.g., representation from legal, compliance, risk, IT, data science, and business units), establishing clear roles, responsibilities, and decision-making authority, and outlining reporting lines and escalation paths for AI-related risks and

strategic decisions. **Phase 2: Foundational Outcomes & Policy Integration** requires the definition and formal documentation of the organization's foundational AI outcomes, including but not limited to its AI ethical principles, risk appetite for AI, data governance principles for AI, and alignment with corporate strategic priorities (e.g., innovation, customer trust). This phase also entails developing or updating organizational policies (e.g., AI Policy, Responsible AI Policy, Data Ethics Policy) that reflect these foundational outcomes and integrate AI risk management into existing enterprise risk management (ERM) and compliance frameworks, ensuring explicit connections between technical AI system design and organizational values. **Phase 3: Communication & Operationalization** focuses on communicating the established governance structure, foundational outcomes, and policies across all relevant internal stakeholders, and establishing mechanisms for ongoing review and adaptation of the governance framework in response to evolving AI technologies, emerging risks, and changes in the regulatory landscape.

Control Frequency: The initial establishment of the AI Governance Structure and Foundational Outcomes is a one-time event. However, the control's *maintenance* and *review* aspects are continuous and should be performed **Annually** for formal reaffirmation and policy updates, or more frequently on an ad-hoc basis whenever there are significant changes in organizational strategy, the AI technology landscape, or emerging regulatory requirements that impact AI governance.

Monitoring Frequency: The effectiveness of this control should be monitored **Bi-annually** or **Annually** by an independent function (e.g., Internal Audit, Compliance, or Enterprise Risk Management), with ad-hoc reviews triggered by significant AI-related incidents, policy breaches, or major organizational shifts impacting AI.

Control Type: This control is primarily **Manual**, involving significant human judgment, decision-making, formal documentation, and communication. There may be aspects that are semi-automated (e.g., document management systems, communication platforms for policy dissemination), but the core governance activities remain manual.

Guide for Evaluating Control Effectiveness: Evaluating the effectiveness of this control involves assessing both the existence and the operationalization of the governance structure and foundational outcomes. This includes: **1. Documentation Review:** Verifying the existence, completeness, formal approval, and currency of key documents such as the AI Governance Body Charter/Terms of Reference, AI Policy, Responsible AI Principles, and documented foundational outcomes. **2. Meeting Minutes & Decisions Review:** Examining minutes of the AI Governance Body meetings to confirm regularity, attendance, substantive discussions of AI risks and opportunities, and evidence of formal decisions related to AI strategy, policy, and risk appetite being made and recorded. **3. Stakeholder Interviews:** Conducting interviews with key stakeholders (e.g., senior management, legal, compliance, data scientists, product owners) to assess their awareness and understanding of the AI governance structure, the foundational outcomes, and how these principles guide their daily AI-related activities. **4. Integration Check:** Confirming documented evidence of integration points between the AI governance framework and the broader Enterprise Risk Management (ERM) framework, demonstrating that AI risks are consistently identified, assessed, and managed within the context of overall enterprise risk. **5. Policy Adherence Assessment:** Periodically assessing if subsequent AI-related projects and initiatives explicitly reference and align with the established foundational outcomes and policies during their design, development, and deployment phases.

Key Control Indicators (KCIs):
* **KCI 1: AI Governance Body Meeting Frequency & Attendance:** Measures the average number of scheduled meetings held per period (e.g., quarterly) and the average attendance rate of core members. (Target: 100% of scheduled meetings held, >80% average attendance).
* **KCI 2: Formal Documentation Completeness & Approval Rate:** Percentage of required AI governance documents (e.g., Charter, AI Policy, Foundational Outcomes document) that are formally approved, current, and readily accessible. (Target: 100% completeness and formal approval).
* **KCI 3: Stakeholder Awareness Score:** Average score from periodic surveys or formal assessments measuring key stakeholders' understanding and awareness of the AI governance structure, their roles, and the foundational outcomes. (Target: >85% awareness).
* **KCI 4: Integration Rate with ERM:** Number of formal references or explicit linkages from AI risk management processes and documentation to the enterprise-wide risk register or ERM framework, demonstrating consistent risk capture. (Target: All identified material AI risks integrated into ERM).
* **KCI 5: Policy Review Cycle Adherence:** Percentage of core AI governance policies (e.g., AI Policy, Responsible AI Principles) that have been reviewed and re-approved within their defined review cycle. (Target: 100% adherence to review cycle).

Requirement Statements:

1. After instituting the outcomes in GOVERN, most users of the AI RMF would start with the MAP function and con- tinue to MEASURE or MANAGE. (Page: 26)

2. provides a structure by which AI risk management functions can align with organi- zational principles, policies, and strategic priorities; (Page: 26)

3. connects technical aspects of AI system design and development to organizational values and principles (Page: 26)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

40. AI System Risk Management Program Establishment and Implementation

Control Actor: Risk Management Team

Control Types: administrative, preventive, detective

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement a comprehensive program for the systematic anticipation, identification, assessment, and management of risks posed by AI systems throughout their lifecycle, encompassing potential impacts on users and society. This program involves designing and implementing the necessary processes, documentation, and organizational structures to achieve effective risk management outcomes, ensuring that potential sources of negative risk are proactively addressed to mitigate uncertainty and enhance the integrity of decision-making processes across the full product lifecycle. This program should specifically utilize the MAP function to enhance the organization's ability to identify risks and broader contributing factors.

Enhanced Implementation Guide:

Control Statement: Establish, implement, and continuously mature a comprehensive AI System Risk Management Program to systematically anticipate, identify, assess, and manage risks throughout the entire AI system lifecycle, from conception to retirement. This program must proactively address potential negative impacts on users, society, and the organization, thereby enhancing decision integrity and mitigating uncertainty, specifically leveraging the 'MAP function' (Methodology for Anticipation and Prioritization) to enhance risk identification and broader contributing factor analysis.

High-Level Control Implementation Guide:

1. **Program Foundation:** Define and document a robust AI Risk Management Policy and Framework aligned with the organization's enterprise risk management strategy, detailing AI risk appetite, roles, responsibilities, and governance structures (e.g., a dedicated AI Risk Management Committee).
2. **Risk Taxonomy & Methodology:** Develop a comprehensive AI-specific risk taxonomy encompassing technical, ethical, societal, operational, security, privacy, and regulatory risks. Establish standardized methodologies and tools for conducting AI risk assessments at key

stages of the AI system lifecycle (e.g., design, development, deployment, monitoring, significant updates, retirement).
3. **'MAP Function' Integration:** Design and embed specific processes, templates, and workshops that leverage the 'MAP function' to facilitate proactive anticipation, comprehensive identification, and granular analysis of AI risks and their root causes/contributing factors. This could involve structured brainstorming, scenario planning, stakeholder impact analysis, and dependency mapping.
4. **Lifecycle Integration:** Mandate and integrate AI risk assessments as mandatory gates within the AI system development lifecycle (AI-SDLC) for all new and significantly modified AI systems.
5. **Mitigation & Controls:** Develop and implement risk treatment plans for identified risks, including the design and implementation of appropriate technical and administrative controls (e.g., data quality validation, bias detection, explainability measures, human-in-the-loop processes, robust monitoring, incident response).
6. **Documentation & Reporting:** Maintain comprehensive documentation of all AI risk assessments, risk registers, mitigation plans, control implementations, and monitoring activities. Establish regular reporting mechanisms to relevant stakeholders and oversight bodies.
7. **Training & Awareness:** Conduct mandatory, regular training programs for all personnel involved in AI system development, deployment, and management, covering AI risk awareness, ethical guidelines, and program procedures.
8. **Continuous Improvement:** Establish mechanisms for continuous monitoring of AI systems, feedback loops for lessons learned from incidents or near-misses, and periodic reviews to adapt the program to evolving AI technologies and regulatory landscapes.

Control Frequency:
* **Program Review & Update:** Annually (or bi-annually, based on organizational risk profile and AI landscape volatility).
* **AI System Risk Assessments:** At minimum, per AI system's critical lifecycle phase (e.g., initial design, pre-deployment, post-significant update/retraining, and annually for high-risk, deployed systems).
* **AI Risk Management Committee Meetings:** Quarterly.
* **Training:** Annually, or upon significant program updates or new AI technology adoption.

Monitoring Frequency:
* **Internal Audit/Compliance Review:** Annually (or bi-annually).
* **Management Review of Program Effectiveness:** Quarterly.
* **Continuous Monitoring of AI Systems for Risk Indicators:** Real-time/daily/weekly,

depending on system criticality and risk velocity.

Control Type: The control is **Semi-Automated**. While the overall program design, policy setting, committee oversight, and critical decision-making are manual/administrative, the execution of certain elements (e.g., data collection for risk assessments, automated scanning for model drifts or anomalies, data analytics for 'MAP function' insights, tracking of mitigation actions) can leverage automated tools and platforms.

Guide for Evaluating Control Effectiveness:

1. **Documentation Review:** Verify the existence, completeness, and approval of the AI Risk Management Policy, Framework, methodology documents (including 'MAP function' specifics), comprehensive AI risk registers, assessment reports, and corresponding mitigation plans.
2. **Process Walkthroughs & Interviews:** Conduct interviews with the Risk Management Team, AI development teams, product owners, and other relevant stakeholders to ascertain their understanding of the program, adherence to defined processes, and the practical application of the 'MAP function' in identifying and managing risks.
3. **Sample Testing:** Select a representative sample of AI systems and verify that:
    * Mandatory risk assessments were conducted at appropriate lifecycle stages.
    * All identified high/critical risks have documented mitigation plans and assigned ownership.
    * Implemented controls are designed and operating effectively (e.g., data quality checks, bias detection, explainability features are active and providing expected outputs).
    * Continuous monitoring data is being collected, reviewed, and acted upon.
4. **Training & Awareness Verification:** Confirm that all relevant personnel have completed mandatory AI risk management training and demonstrate awareness of their responsibilities.
5. **Incident & Audit Log Review:** Analyze AI-related incident logs, near-miss reports, and prior audit findings to assess the program's effectiveness in proactively identifying and addressing risks, and its responsiveness to unexpected events.
6. **Stakeholder Feedback:** Solicit feedback from internal and external stakeholders on the perceived transparency, fairness, and overall trustworthiness of AI systems, correlating feedback to the program's efficacy.

Key Control Indicators (KCIs):

* **Percentage of in-scope AI systems with completed risk assessments:** (Target: 100%).
* **Percentage of identified high/critical AI risks with documented mitigation plans and assigned ownership:** (Target: >95%).
* **Average time to implement critical AI risk mitigation actions:** (Target: e.g., <30 days from identification).

* **Number of AI-related incidents or material adverse impacts attributable to unmanaged/unidentified risks:** (Target: Decreasing trend, ideally 0 critical incidents).
* **Percentage of mandatory AI risk management training completed by relevant personnel:** (Target: >90%).
* **Number of open audit findings or critical non-conformities related to the AI Risk Management Program:** (Target: 0 critical findings).
* **Coverage/utilization rate of the 'MAP function' in new AI system assessments:** (e.g., % of high-risk AI systems where documented 'MAP function' outputs contributed to risk identification).
* **Risk Reduction Index:** A calculated metric showing the aggregate reduction in risk exposure (based on severity/likelihood) across the AI portfolio over time.

Requirement Statements:

1. outlines processes, documents, and organizational schemes that anticipate, identify, and manage the risks a system can pose, including to users and others across society – and procedures to achieve those outcomes; (Page: 26)

2. addresses full product lifecycle and associated processes (Page: 26)

3. Anticipating, assessing, and otherwise addressing potential sources of negative risk can mitigate this uncertainty and enhance the integrity of the decision process. (Page: 30)

4. The MAP function is intended to enhance an organization’s ability to identify risks and broader contributing factors. (Page: 30)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

41. Organizational AI Competency and Practice Development

Control Actor: Human Resources

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain organizational practices and develop competencies for all individuals involved in the AI system lifecycle, including acquisition, training, deployment, and monitoring of AI systems. This ensures the organization possesses the necessary skills and operational capabilities to manage AI systems responsibly and effectively.

Enhanced Implementation Guide:

Control Statement: The organization shall establish, maintain, and continuously enhance its internal capabilities and individual competencies for the responsible and effective management across the entire AI system lifecycle, encompassing acquisition, training, deployment, monitoring, and decommissioning, ensuring robust governance and risk mitigation.

Implementation Guide: To achieve this, Human Resources, in collaboration with AI stakeholders, must:
*   **Conduct a comprehensive AI skill gap analysis:** Define critical AI-related roles (e.g., AI engineers, data scientists, MLOps specialists, AI ethics reviewers, legal counsel, project managers) and assess current competencies against required proficiencies for responsible AI development and deployment.
*   **Develop tailored training programs:** Design or procure structured training modules covering technical AI skills, responsible AI principles (e.g., fairness, transparency, accountability, privacy), AI ethics, data governance, AI risk management, regulatory compliance (e.g., GDPR, NIST AI RMF), and organization-specific AI policies and procedures. These programs should cater to varying levels of AI involvement, from executive awareness to deep technical expertise.
*   **Establish clear AI competency frameworks and career paths:** Integrate AI competencies into existing HR frameworks, defining progression paths and required skills for each AI-related role, and linking these to performance management and professional development plans.
*   **Foster a continuous learning environment:** Promote knowledge sharing through

communities of practice, internal workshops, and access to external AI research and best practices.
* **Develop and disseminate organizational AI practices:** Document clear internal guidelines, standard operating procedures, and policies for each phase of the AI system lifecycle (e.g., responsible AI development checklist, AI model validation procedures, AI risk assessment methodology) and ensure adherence.

Frequency: The control itself is **ongoing**, requiring continuous adaptation of competency frameworks and training curricula. A **formal review and update of the overall AI competency strategy and practice documentation should occur annually**.

Monitoring Frequency: Control performance should be monitored **quarterly**, assessing training completion rates, competency assessment results, and adherence to established AI practices.

Control Type: This is a **semi-automated** control. While the development of curriculum and practices is manual, tracking of training completion, skills inventories, and potentially performance against competencies can leverage HRIS systems and automated assessment tools.

Guide for Evaluating Control Effectiveness: Control effectiveness is evaluated by:
* **Auditing AI competency frameworks:** Verifying their comprehensiveness, alignment with evolving AI risks and regulatory landscapes, and integration into HR processes.
* **Reviewing training participation and assessment records:** Confirming that target populations are completing relevant training and demonstrating acquired knowledge.
* **Interviewing key stakeholders:** Assessing the level of understanding and confidence in applying responsible AI principles and practices across different roles.
* **Analyzing post-implementation project reviews:** Identifying instances where lack of competency or adherence to practices contributed to AI system failures, ethical concerns, or compliance breaches.
* **Assessing the completeness and adoption of documented AI lifecycle practices:** Verifying that established guidelines are comprehensive, accessible, and consistently applied within AI development and operational teams.

Key Control Indicators (KCIs):
* **Percentage of employees in AI-related roles completing mandatory responsible AI training.**
* **Average scores on AI competency assessments (where applicable).**

*** Reduction in AI system incidents or issues directly attributable to human error or lack of adherence to AI practices.**
*** Employee self-assessment scores on confidence in applying responsible AI principles.**
*** Number of documented AI lifecycle practices and their adoption rate within relevant teams.**
*** Time-to-proficiency for new hires in critical AI roles.**

Requirement Statements:

1. and enables organizational practices and competencies for the individuals involved in acquiring, training, deploying, and monitoring such systems; (Page: 26)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

42. AI System Incident Response, Recovery, and Risk Information Sharing

Control Actor: Risk Management Team, Incident Response Team, Information Security Officer

Control Types: preventive, detective, corrective, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain robust organizational practices to enable comprehensive AI system testing throughout the lifecycle, facilitate the timely identification and response to AI-related incidents (including security, performance, ethical issues, and previously unknown risks as they are identified), and promote effective information sharing regarding AI risks, vulnerabilities, and performance insights within the organization and with relevant external stakeholders, including affected communities. These practices ensure proactive risk management and continuous improvement of AI system trustworthiness. This includes developing explicit risk treatment plans to effectively respond to, recover from, and communicate about incidents or adverse events affecting AI systems, ensuring procedures are followed for both known and emergent risks. These plans, encompassing response, recovery, and communication strategies for identified and measured AI risks, must be thoroughly documented and regularly monitored to ensure their effectiveness and readiness.

Enhanced Implementation Guide:

Control Statement: The organization shall establish, implement, and maintain a comprehensive AI System Incident Response, Recovery, and Risk Information Sharing program that proactively manages AI-related risks, ensures timely incident identification and resolution, facilitates effective system recovery, and promotes transparent communication regarding AI system trustworthiness across its entire lifecycle, including response to known and emergent risks.

High-Level Control Implementation Guide: This control is implemented by developing and maintaining explicit, AI-specific incident response plans (IRPs) that cover security breaches, performance degradation, ethical drift, bias, and previously unknown or emergent risks across all phases of the AI system lifecycle (design, development, deployment, operation, decommissioning). These plans must define clear roles, responsibilities, escalation paths, and communication protocols for internal stakeholders (e.g., development teams, legal, compliance, executive leadership) and external parties (e.g., affected users, regulatory bodies, affected communities). Implementation includes integrating AI-specific testing methodologies (e.g., adversarial testing, bias detection, performance stress testing, explainability evaluations) into the Continuous Integration/Continuous Deployment (CI/CD) pipeline to proactively identify vulnerabilities and risks. Robust monitoring tools must be deployed to detect anomalies and trigger alerts for potential incidents. Furthermore, an effective information sharing framework must be established, including internal dashboards for risk insights and external mechanisms for transparent communication of AI risks, vulnerabilities, and performance insights to relevant stakeholders. Regular training and tabletop exercises must be conducted to ensure personnel readiness and validate plan effectiveness. Post-incident reviews (PIRs) are mandatory to capture lessons learned and drive continuous improvement of AI systems and the response program.

Control Frequency:
*   **Testing:** Continuous, integrated into the AI/MLOps lifecycle, prior to significant deployments, and periodically (e.g., quarterly) for deployed systems.
*   **Incident Response & Recovery Plan Execution:** Event-driven (as incidents occur).
*   **Plan Review & Update:** Annually, or after any significant incident, regulatory change, or major AI system modification.
*   **Exercises/Drills:** Bi-annually.
*   **Risk Information Sharing:** Continuous (monitoring for insights), event-driven (for incidents), and regularly scheduled (e.g., quarterly) for aggregated risk reporting.

Monitoring Frequency:
*   **AI System Monitoring for Incidents:** Continuous.
*   **Incident Response Effectiveness:** Post-incident review (within 5 business days of incident resolution).
*   **Programmatic Effectiveness (Plans, Readiness):** Quarterly reviews by the Risk Management Team and Internal Audit.

Control Type: Semi-automated. Incident detection often leverages automated monitoring tools (e.g., anomaly detection, drift monitoring), while the response, recovery, communication, and plan management aspects are largely manual and procedural, guided by documented protocols.

Guide for Evaluating Control Effectiveness: Effectiveness is evaluated by assessing the organization's ability to timely identify, respond to, recover from, and communicate about AI system incidents and adverse events. This includes reviewing AI incident logs for completeness, accuracy, and adherence to defined response and recovery procedures. Success in meeting defined Service Level Agreements (SLAs) for incident resolution (MTTD, MTTR) is a key indicator. Audits of AI risk treatment plans and their associated documentation (e.g., version control, approval records) ensure they are current, comprehensive, and align with organizational policies. The results of tabletop exercises and post-incident reviews provide evidence of readiness and areas for improvement, with verification of action item completion. Feedback from internal stakeholders and external communities on the clarity, timeliness, and helpfulness of information sharing is also critical.

Key Control Indicators (KCIs):

1. **Mean Time To Detect (MTTD) AI Incidents:** Average time from the occurrence of an AI incident to its detection.
2. **Mean Time To Respond (MTTR) AI Incidents:** Average time from AI incident detection to the initiation of response actions.
3. **Mean Time To Recover (MTTR) AI Systems:** Average time from AI incident detection to the full restoration of affected AI system functionality.
4. **Percentage of AI Incidents Resolved within SLA:** Proportion of incidents that met predefined resolution targets.
5. **Number of Critical/High Severity AI Incidents:** Count of incidents categorized as critical or high severity.
6. **AI Incident Recurrence Rate:** Number of times similar incidents reoccur after initial resolution.
7. **Completion Rate of Post-Incident Review (PIR) Action Items:** Percentage of corrective actions from PIRs that have been implemented.
8. **Frequency of AI System Testing Activities:** Number of planned adversarial tests, bias assessments, or performance tests conducted versus planned.
9. **Stakeholder Satisfaction Score for AI Risk Communication:** Survey-based score measuring the effectiveness of internal and external communication regarding AI risks and incidents.
10. **Percentage of AI Systems with Documented & Tested IR/Recovery Plans:** Proportion of operational AI systems that have comprehensive, reviewed, and exercised incident response and recovery plans.

Requirement Statements:

1. Organizational practices are in place to enable AI testing, identification of incidents, and information sharing. (Page: 29)

2. Risk treatment comprises plans to respond to, recover from, and communicate about incidents or events. (Page: 36)

3. Procedures are followed to respond to and recover from a previously unknown risk when it is identified. (Page: 37)

4. MANAGE 4: Risk treatments, including response and recovery, and communication plans for the identified and measured AI risks are documented and monitored regularly. (Page: 38)

5. MANAGE 4.3: Incidents and errors are communicated to relevant AI actors, including affected communities. (Page: 38)

6. Processes for track- ing, responding to, and recovering from incidents and errors are followed and documented. (Page: 38)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

43. External AI Risk Impact Feedback Integration

Control Actor: Compliance Management Team

Control Types: administrative, preventive, detective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain organizational policies and practices to systematically collect, consider, prioritize, and integrate feedback from external parties (e.g., individuals, communities, advocacy groups, research organizations) regarding potential individual and societal impacts associated with AI risks throughout the AI system lifecycle. This ensures that diverse perspectives and real-world experiences are leveraged to inform, enhance, and continuously improve the organization's AI risk management processes and mitigate adverse outcomes.

Enhanced Implementation Guide:

The organization shall establish, implement, and maintain a robust framework for systematically collecting, evaluating, prioritizing, and integrating feedback from diverse external stakeholders regarding potential individual and societal impacts associated with AI risks throughout the AI system lifecycle. This control is primarily **semi-automated**, leveraging automated channels for initial feedback collection and potential AI-assisted categorization, but relies heavily on **manual** human oversight, review, prioritization, decision-making, and integration into AI governance and development processes. **Implementation guidance** includes: (1) Formalizing a policy for external AI risk feedback integration, outlining channels, responsibilities, and processes. (2) Establishing secure, accessible, and diverse external feedback mechanisms (e.g., dedicated online portals, public forums, direct outreach to advocacy groups). (3) Defining clear criteria and a structured process for analyzing, categorizing, and prioritizing feedback based on severity, relevance, and feasibility of action. (4) Forming a cross-functional committee (including compliance, legal, AI ethics, product, and engineering) to review prioritized feedback and propose actionable responses. (5) Developing a systematic procedure to integrate feedback into AI system design, development, testing, deployment, and post-market monitoring, ensuring timely updates to risk assessments, mitigation strategies, and system parameters. (6) Maintaining comprehensive documentation of all feedback received, analyses performed, decisions made, and corrective actions taken. (7) Where appropriate and feasible, establishing communication channels to inform external parties about actions taken based on their feedback, fostering transparency. **Control**

frequency for feedback collection is continuous; review and prioritization should occur at least **monthly** for active AI projects and **quarterly** for overarching policy adjustments, with critical feedback addressed immediately. **Monitoring frequency** by the Compliance Management Team shall be **quarterly**, with an annual review by Internal Audit. **Evaluating control effectiveness** will involve: (1) Assessing adherence to documented policies and procedures for external feedback processing. (2) Reviewing a sample of documented feedback cases to ensure systematic collection, comprehensive analysis, clear prioritization, and evidence of integration into AI risk management. (3) Verifying the diversity and breadth of external stakeholder engagement. (4) Confirming that feedback has led to measurable improvements in AI risk mitigation or policy adjustments. (5) Evaluating the timeliness of feedback processing and action implementation. **Key control indicators (KCIs)** include: (a) Number of unique external feedback submissions related to AI risks per quarter. (b) Percentage of high-priority feedback items leading to documented changes in AI systems or risk management processes. (c) Average time from receipt of critical feedback to initiation of corrective action. (d) Diversity index of external feedback sources (e.g., number of distinct stakeholder categories engaged). (e) Number of documented AI risk mitigations directly attributable to external feedback. (f) Trend of reported adverse outcomes linked to AI systems post-implementation of feedback-driven changes.

Requirement Statements:

1. Organizational policies and practices are in place to collect, consider, prioritize, and integrate feedback from those external to the team that developed or deployed the AI system regarding the potential individual and societal impacts related to AI risks. (Page: 29)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ___________________________________

Evidence/Comments:

____

____

44. AI System Design Feedback Integration

Control Actor: AI Development Team

Control Types: administrative, preventive, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain mechanisms to enable the AI Development Team to regularly incorporate adjudicated feedback from relevant AI actors into AI system design and implementation. This ensures continuous improvement and refinement of AI systems based on stakeholder input and operational insights, fostering adaptability and trustworthiness throughout the AI system lifecycle.

Enhanced Implementation Guide:

The AI Development Team shall establish, maintain, and execute a formal process for the systematic collection, adjudication, and integration of feedback from relevant AI actors (e.g., end-users, operators, subject matter experts, legal, ethics, security, and business stakeholders) into AI system design and implementation, ensuring continuous improvement, adaptability, and trustworthiness across the AI system lifecycle. This control is primarily **semi-automated**, leveraging specialized tools for feedback submission and tracking, with manual human intervention for adjudication and critical decision-making processes. **High-level implementation guidance** includes defining clear feedback channels (e.g., dedicated portal, structured meetings, incident reporting, user forums), establishing a standardized process for feedback triage, prioritization, impact assessment, and formal adjudication by designated personnel, ensuring decisions are documented. Integrated feedback must demonstrably inform updates to requirements, design specifications, user stories, and code changes. A communication loop must be established to inform feedback providers of the status and resolution of their input. **Control frequency** for feedback collection is continuous, adjudication should occur at least bi-weekly or prior to each sprint planning/major release, and integration into design/implementation should align with sprint cycles or release cadences. The overall process effectiveness should be reviewed at least annually. **Monitoring frequency** by compliance or internal audit will be quarterly, complemented by continuous line management oversight. **To evaluate control effectiveness**, evidence must demonstrate the existence of a documented feedback management process, comprehensive feedback logs/databases (showing submission, categorization, adjudication decisions, and disposition), traceability of adjudicated feedback items to tangible design changes, code modifications, or updated system documentation, and evidence of communication back to feedback providers. Interviews with the AI Development Team and relevant AI actors can validate process adherence and perceived effectiveness. **Key control indicators** include: percentage of feedback items adjudicated within defined service level agreements (SLAs), percentage of critical/high-priority adjudicated feedback items integrated into design or implementation within a specified timeframe, average time taken from feedback submission to integration, user satisfaction scores with the feedback process, and the trend of unaddressed critical feedback items over time.

Requirement Statements:

1. Mechanisms are established to enable the team that developed or deployed AI systems to regularly incorporate adjudicated feedback from relevant AI actors into system design and implementation. (Page: 29)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ___________________________________

Evidence/Comments:

____

____

____

45. Third-Party AI System Contingency and Incident Response

Control Actor: Operations Team

Control Types: preventive, detective, corrective, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain robust contingency processes to effectively handle failures or incidents occurring in high-risk third-party data or AI systems. These processes ensure timely response, mitigation of negative impacts, and continuity of operations, minimizing disruption and ensuring the trustworthiness and resilience of AI systems reliant on external components.

Enhanced Implementation Guide:

Control Statement: The Organization shall establish, document, and regularly test comprehensive contingency and incident response plans specifically for high-risk third-party AI systems and their underlying data components. These plans must ensure the timely detection, effective mitigation, and rapid recovery from failures or incidents, thereby minimizing operational disruption, safeguarding data integrity, and upholding the trustworthiness and resilience of AI systems reliant on external components.

High-Level Control Implementation Guide: To implement this control, the Operations Team, in collaboration with relevant stakeholders (e.g., IT, legal, procurement, risk management), must: 1. Conduct a risk assessment to identify and classify all high-risk third-party AI systems and associated data dependencies, defining criticality levels, Recovery Time Objectives (RTOs), and Recovery Point Objectives (RPOs). 2. Develop and document specific Incident Response Plans (IRPs) and Business Continuity Plans (BCPs) for each identified high-risk system, incorporating unique AI considerations such as model degradation, data poisoning, bias drift, explainability loss, and adversarial attacks. 3. These plans must clearly define roles, responsibilities, escalation paths, communication protocols (internal, external, and with the Provider Organization), data backup and restoration procedures, and alternative system/data source activation strategies. 4. Establish clear triggers for incident declaration and response, integrating with existing organizational incident management frameworks. 5. Integrate contractual obligations and Service Level Agreements (SLAs) with third-party providers into the incident response framework, detailing their expected support, communication channels, and resolution timelines. 6. Conduct regular training and awareness programs for all personnel involved in managing or responding to third-party AI system incidents. 7. Maintain a centralized, accessible repository for all plans, documentation, test results, and post-incident review reports. 8. Establish mechanisms for continuous monitoring of third-party AI system performance and health to enable proactive detection of anomalies or potential incidents.

Control Frequency: The initial establishment of plans and documentation is a one-time activity. Plans must be reviewed and updated at least annually, or upon any significant change to the third-party AI system, its data dependencies, contractual terms, or following a major incident. Comprehensive tabletop exercises and simulated incident response tests should be conducted annually for all high-risk systems. Incident response execution occurs on an as-needed basis when an incident is detected.

Monitoring Frequency: The effectiveness of this control should be monitored on a quarterly basis through review of incident logs, resolution metrics, and adherence to established RTO/RPO targets. A formal audit or control effectiveness review should be conducted annually to assess the completeness, accuracy, and currency of plans, and the performance of annual tests.

Control Type: This control is primarily **Semi-automated**. While the establishment, documentation, testing, and post-incident review aspects are inherently manual processes requiring human judgment and action, the detection of incidents often relies on automated monitoring tools and alerts from third-party systems. Execution of response may involve automated scripts or tools for recovery actions, but critical decision-making and coordination remain human-driven.

Guide for Evaluating Control Effectiveness: Control effectiveness is evaluated by assessing the following: 1. **Documentation Completeness:** Existence, accuracy, and currency of documented IRPs and BCPs for all identified high-risk third-party AI systems. 2. **Testing Efficacy:** Evidence of regular, documented testing (e.g., tabletop exercises, simulation drills) demonstrating the ability to meet defined RTOs and RPOs, with identified gaps promptly addressed and remediated. 3. **Incident Response Performance:** The Organization's ability to detect, respond to, and recover from actual incidents within defined RTOs/RPOs, evidenced by incident reports, root cause analyses, and timely implementation of corrective actions. 4. **Roles and Responsibilities:** Clarity and understanding of assigned roles, responsibilities, and communication protocols among all involved parties (internal and external). 5. **Training Adherence:** Evidence that all relevant personnel have received adequate and recurrent training on incident response plans. 6. **Contractual Alignment:** Demonstrated adherence to contractual SLAs with third-party providers regarding incident notification, support, and resolution.

Key Control Indicators (KCIs):
* **KCI 1: Plan Coverage:** Percentage of high-risk third-party AI systems with approved and current IRPs/BCPs (Target: 100%).
* **KCI 2: Test Completion Rate:** Percentage of scheduled annual contingency tests and drills completed (Target: 100%).
* **KCI 3: RTO/RPO Adherence:** Average actual Recovery Time (RTA) and Recovery Point (RPA) vs. defined Recovery Time Objective (RTO) and Recovery Point Objective (RPO) for critical third-party AI system incidents (Target: RTA ≤ RTO, RPA ≤ RPO).
* **KCI 4: Incident Resolution Time:** Average time to fully resolve critical incidents involving third-party AI systems from detection to closure.
* **KCI 5: Post-Incident Review Closure Rate:** Percentage of corrective actions identified from post-incident reviews that are implemented and closed within defined timelines (Target: >90%).
* **KCI 6: Training Compliance:** Percentage of relevant personnel completing annual incident response training (Target: 100%).
* **KCI 7: Repeat Incidents:** Number of repeat incidents stemming from previously identified and supposedly resolved root causes related to third-party AI systems.

Requirement Statements:

1. Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk. (Page: 29)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

46. AI RMF MAP Function Enhancement via Diverse Stakeholder Engagement and Feedback Integration

Control Actor: AI Governance Team

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes, including documented practices and assigned personnel, to enhance the implementation of the AI Risk Management Framework's (AI RMF) MAP function. This involves proactively incorporating perspectives from diverse internal teams and engaging regularly with relevant AI actors, external collaborators, end users, and potentially impacted communities. This engagement facilitates the integration of feedback regarding positive, negative, and unanticipated impacts of AI systems, which is crucial for preventing negative risks and developing trustworthy AI systems. The level of engagement should vary based on the AI system's risk level, internal team composition, and organizational policies, ensuring comprehensive and context-sensitive AI risk management. This specifically includes considering the degree to which each measurement type provides unique and meaningful information to the assessment of AI risks, ensuring measurement strategies are optimized for effectiveness.

Enhanced Implementation Guide:

AI RMF MAP Function Enhancement via Diverse Stakeholder Engagement and Feedback Integration - Enhanced Control Description

Control Statement: The organization shall establish, document, and maintain a robust process for the continuous enhancement of the AI Risk Management Framework's (AI RMF) MAP (Measure, Analyze, Protect) function by systematically incorporating perspectives from diverse internal teams and engaging regularly with relevant AI actors, external collaborators, end-users, and potentially impacted communities. This process must ensure feedback regarding positive, negative, and unanticipated impacts of AI systems is proactively collected, analyzed, and integrated to prevent negative risks, develop trustworthy AI systems, and optimize measurement strategies to yield unique and meaningful information for AI risk assessment.

High-Level Control Implementation Guide: The AI Governance Team, in collaboration with AI development, legal, ethics, and business units, shall implement this control by:

1. **Developing a Tiered Stakeholder Engagement Strategy:** Define engagement levels (e.g., informational, consultative, collaborative, empowering) tailored to the AI system's risk level, internal team composition, and organizational policies. This includes identifying key internal stakeholders (AI product/engineering teams, legal, compliance, ethics, business owners) and external stakeholders (end-users, affected communities, domain experts, regulators, third-party providers).
2. **Establishing Formal Feedback Mechanisms:** Implement structured channels for feedback collection, such as regular forums, workshops, surveys, direct interviews, user testing, and complaint/grievance mechanisms. Ensure these mechanisms are accessible and inclusive for all identified stakeholder groups.
3. **Defining Feedback Analysis and Integration Procedures:** Document processes for analyzing collected feedback, identifying recurring themes, emerging risks, and opportunities for improvement in AI system design, development, deployment, and particularly the RMF's measurement types and strategies. Clearly delineate how feedback will translate into actionable recommendations and updates to the AI RMF and specific AI system risk profiles.
4. **Assigning Roles and Responsibilities:** Clearly define personnel responsible for stakeholder identification, engagement planning, feedback collection, analysis, integration into the RMF MAP function, and communication of outcomes.
5. **Continuous Improvement Loop:** Establish a mechanism to review the effectiveness of engagement processes and feedback integration, ensuring lessons learned are applied to future iterations. This specifically includes assessing if measurement strategies are truly optimized for effectiveness in capturing comprehensive AI risks.

Control Frequency:

* **Stakeholder Engagement:** Varies based on AI system risk level and lifecycle stage. For high-risk AI systems, engagement should be **quarterly** or at significant development/deployment milestones. For moderate-risk systems, **bi-annually**. For low-risk systems, **annually** or upon major system changes.
* **Feedback Integration Review:** The AI Governance Team should review collected feedback and its integration into the RMF MAP function **monthly** for high-risk AI systems, and **quarterly** for all other relevant systems.
* **Overall Process Documentation Update:** At least **annually**, or as regulatory guidance evolves or significant internal changes occur.

Monitoring Frequency:

* **Internal Monitoring: Quarterly** reviews by the AI Governance Team or internal audit

function to assess adherence to the established processes and the effectiveness of feedback integration.
*   **External Review/Audit: Annually** or as mandated by regulatory requirements or independent third-party assessments.

Control Type: Semi-automated (Relies on manual processes for engagement and qualitative feedback analysis, supported by automated tools for survey distribution, data collection, stakeholder management, and reporting).

Guide for Evaluating Control Effectiveness:
1. **Documentation Completeness and Adherence:** Verify the existence and regular update of a comprehensive Stakeholder Engagement Plan and documented feedback integration procedures. Confirm adherence to the tiered engagement model for various AI systems.
2. **Feedback Integration Quality:** Assess whether collected feedback is systematically analyzed and demonstrably leads to improvements in the AI RMF MAP function, particularly evidenced by enhancements to measurement strategies (e.g., new metrics, refined assessment methodologies). Review action logs derived from feedback.
3. **Stakeholder Participation and Satisfaction:** Evaluate the breadth and depth of stakeholder engagement, and where feasible, gather feedback from stakeholders on the effectiveness of engagement channels and the perceived responsiveness to their input.
4. **Risk Mitigation Impact:** Review incident reports and risk assessments to determine if the engagement process effectively identified and mitigated previously unanticipated negative impacts or emerging risks, contributing to the development of more trustworthy AI systems.

Key Control Indicators (Metrics):
*   **Percentage of High-Risk AI Systems with Documented and Executed Tiered Engagement Plans:** (Target: 100%)
*   **Number of AI RMF MAP Measurement Strategy Enhancements/Refinements:** (e.g., Average 3-5 significant enhancements per year based on stakeholder feedback).
*   **Percentage of Critical/High-Priority Feedback Items Integrated into Action Plans:** (e.g., >85% of identified critical feedback leading to a documented action).
*   **Stakeholder Feedback Responsiveness Rating:** (Average score > 4.0 on a 5-point scale from stakeholder surveys regarding their perception of engagement and responsiveness).
*   **Reduction in Post-Deployment Unanticipated Negative Impacts:** (Trend showing a decrease in new, unanticipated negative impacts identified post-deployment, suggesting proactive identification through engagement).
*   **Number of Feedback-Driven AI System Improvements:** (Count of specific AI system enhancements or mitigations directly attributable to stakeholder feedback).

Requirement Statements:

1. Implementation of this function is enhanced by incorporating perspectives from a diverse internal team and engagement with those external to the team that developed or deployed the AI system. Engagement with external collaborators, end users, potentially impacted communities, and others may vary based on the risk level of a particular AI system, the makeup of the internal team, and organizational policies. Gathering such broad perspec- tives can help organizations proactively prevent negative risks and develop more trustwor- thy AI systems... (Page: 30)

2. MAP 5.2: Practices and personnel for supporting regular en- gagement with relevant AI actors and integrating feedback about positive, negative, and unanticipated impacts are in place and documented. (Page: 33)

3. The degree to which each measurement type provides unique and meaningful information to the assessment of AI risks should be considered. (Page: 33)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

47. Post-Deployment AI Risk Management, Measurement Contextualization, and Stakeholder Feedback

Control Actor: AI Governance Team, Risk Management Team, AI Development Team, AI Operations Team, System Owners

Control Types: administrative, preventive, detective, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for the AI Governance Team and Risk Management Team to effectively manage AI risks after a decision has been made to proceed with an AI system. This involves systematically utilizing the MEASURE and MANAGE functions of the AI Risk Management Framework, in conjunction with established policies and procedures from the GOVERN function, to ensure ongoing risk identification, assessment, mitigation, and monitoring throughout the operational lifecycle of AI systems. Crucially, this includes continually applying the MANAGE function to deployed AI systems, adapting to evolving methods, contexts, risks, and stakeholder needs or expectations, and implementing mechanisms to sustain their value. This specifically includes ensuring objective, repeatable, and scalable test, evaluation, verification, and validation (TEVV) processes, along with relevant metrics, methods, and methodologies, are established, followed, and documented after completing the MEASURE function. Furthermore, establish and integrate feedback processes for end users and impacted communities to report problems and appeal system outcomes into AI system evaluation metrics. Measurement approaches for identifying AI risks and AI system trustworthiness must be connected to deployment contexts and informed through consultation with domain experts and relevant AI actors to validate consistent performance across deployment contexts and the AI lifecycle. This contextual information gleaned from expert consultation and input from relevant AI actors must be utilized to decrease the likelihood of AI system failures and negative impacts. Gather and assess feedback specifically about the efficacy of measurement approaches for AI risks, including the identification of measurable performance improvements or declines based on consultations with relevant AI actors (including affected communities) and field data about context-relevant risks and trustworthiness characteristics. This proactive approach ensures continuous adherence to trustworthiness principles and supports responsible AI deployment, ensuring the continuous application and refinement of the MEASURE function to AI systems as knowledge, methodologies, risks, and impacts evolve over time.

Enhanced Implementation Guide:

Control Statement: Establish, maintain, and continuously refine a robust, adaptive, and stakeholder-informed post-deployment AI risk management framework that integrates ongoing test, evaluation, verification, and validation (TEVV) processes, performance monitoring, contextualized risk assessment, and active feedback mechanisms from end-users and impacted communities, to ensure the continuous trustworthiness, value, and responsible operation of AI systems throughout their operational lifecycle.

High-Level Control Implementation Guide: The AI Governance and Risk Management Teams, in collaboration with AI Development and Operations Teams and System Owners, shall define and implement comprehensive policies and procedures for post-deployment AI risk management, drawing upon the MEASURE and MANAGE functions of the AI Risk Management Framework. This includes establishing objective, repeatable, and scalable TEVV processes with defined metrics and methodologies, ensuring they are consistently applied and documented for all deployed AI systems. Implement continuous monitoring mechanisms to track AI system performance, trustworthiness characteristics (e.g., bias, fairness, reliability, security), and emerging contextual risks. Integrate user-friendly feedback channels for end-users and impacted communities to report issues, appeal outcomes, and provide insights into AI system efficacy and impact; ensure this feedback actively informs AI system evaluation metrics and operational adjustments. Regularly consult with domain experts and relevant AI actors (including affected communities) to contextualize AI risk measurements, validate consistent performance across diverse deployment contexts, and inform adaptive risk mitigation strategies. All risk identification, assessment, mitigation actions, and feedback analyses must be meticulously documented and reviewed to ensure proactive adaptation to evolving methods, contexts, risks, and stakeholder needs.

Control Frequency: Continuous (for automated monitoring and ongoing adaptation); Periodic (e.g., quarterly or semi-annually for formal TEVV updates, risk reviews, expert consultations, and feedback analysis reviews); Event-driven (upon identification of new risks, significant performance degradation, or critical stakeholder feedback).

Monitoring Frequency: Continuous (for automated performance and risk monitoring); Monthly (for review of automated alerts, feedback logs, and performance dashboards by AI Operations/Risk Management Teams); Quarterly (for formal oversight and efficacy assessment by the AI Governance Team).

Control Type: Semi-automated (leveraging automated tools for continuous monitoring and data collection, complemented by significant manual oversight, expert judgment, stakeholder engagement, and adaptive decision-making).

Guide for Evaluating Control Effectiveness: Effectiveness is evaluated by assessing the presence and adherence to documented post-deployment AI risk management policies and procedures. Verify that TEVV processes are consistently applied, documented, and demonstrate objective evaluation of AI system performance and trustworthiness characteristics. Review evidence that feedback mechanisms are actively utilized by end-users and impacted communities, and that their input demonstrably influences AI system adjustments, risk mitigation strategies, and performance improvements. Audit records to confirm that expert consultations and contextual information are systematically integrated into risk measurement and validation processes, leading to demonstrable reductions in AI system failures or negative impacts. Analyze incident logs and corrective action plans to confirm timely and effective responses to identified risks and system anomalies. Validate the proactive and adaptive nature of risk management by tracing how evolving risks, contexts, and stakeholder needs are continuously addressed.

Key Control Indicators (KCIs):

- **AI System Performance Drift:** Percentage deviation from established performance or trustworthiness benchmarks over time.
- **Risk Mitigation Effectiveness:** Percentage reduction in critical/high-severity post-deployment AI risks identified and mitigated.
- **Feedback Loop Efficacy:**
  - Number of user/community feedback reports/appeals received per AI system.
  - Average time to resolution for reported issues and appeals.
  - Percentage of feedback-driven improvements implemented on AI systems.
  - Quantitative/qualitative measure of stakeholder satisfaction with feedback processes.
- **TEVV Adherence & Coverage:**
  - Percentage of deployed AI systems with up-to-date and completed TEVV documentation.
  - Compliance rate with defined TEVV methodologies and metrics.
- **Incident Reduction Rate:** Percentage decrease in AI system failures or negative impacts post-deployment.
- **Contextual Adaptability:** Number of documented instances where AI system parameters, models, or operational procedures were adapted based on new contextual information or expert consultation.

Requirement Statements:

1. If a decision is made to proceed, organizations should utilize the MEASURE and MANAGE functions along with policies and procedures put into place in the GOVERN function to assist in AI risk management efforts. (Page: 30)

2. After completing the MEASURE function, objective, repeatable, or scalable test, evaluation, verification, and validation (TEVV) processes including metrics, methods, and methodolo- gies are in place, followed, and documented. (Page: 33)

3. It is in- cumbent on Framework users to continue applying the MEASURE function to AI systems as knowledge, methodologies, risks, and impacts evolve over time. (Page: 33)

4. MEASURE 3.3: Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established and integrated into AI system evaluation metrics. (Page: 36)

5. MEASURE 4: Feedback about efficacy of measurement is gathered and assessed. (Page: 36)

6. MEASURE 4.1: Measurement approaches for identifying AI risks are connected to deployment context(s) and informed through consultation with domain experts and other end users. (Page: 36)

7. MEASURE 4.2: Measurement results regarding AI system trust- worthiness in deployment context(s) and across the AI lifecycle are informed by input from domain experts and relevant AI ac- tors to validate whether the system is performing consistently as intended. (Page: 36)

8. MEASURE 4.3: Measurable performance improvements or de- clines based on consultations with relevant AI actors, in- cluding affected communities, and field data about context- relevant risks and trustworthiness characteristics are identified and documented. (Page: 36)

9. Contextual information gleaned from expert consultation and input from relevant AI actors – established in GOVERN and carried out in MAP – is utilized in this function to decrease the likelihood of system failures and negative impacts. (Page: 36)

10. It is incumbent on Framework users to continue to apply the MANAGE function to deployed AI systems as methods, contexts, risks, and needs or expectations from relevant AI actors evolve over time. (Page: 36)

11. Mechanisms are in place and applied to sustain the value of deployed AI systems. (Page: 37)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

48. Continuous Application of AI RMF MAP Function

Control Actor: AI Governance Team

Control Types: administrative, preventive, detective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for the AI Governance Team to continuously apply the MAP function of the AI Risk Management Framework as the context, capabilities, risks, benefits, and potential impacts of AI systems evolve over time. This ensures ongoing identification, assessment, and understanding of AI risks and opportunities throughout the AI system lifecycle, adapting to changing circumstances and fostering proactive risk management.

Enhanced Implementation Guide:

The AI Governance Team shall establish and maintain robust, documented processes for the continuous application of the AI Risk Management Framework's (RMF) MAP (Measure, Analyze, Prioritize) function throughout the entire lifecycle of all AI systems, including those from Provider Organizations. This control ensures ongoing, adaptive identification, assessment, and understanding of AI risks, benefits, and opportunities in response to evolving contexts, capabilities, and potential impacts.

High-Level Implementation Guide: The AI Governance Team must define clear procedures for data collection, analysis, and prioritization of AI risks. This includes leveraging automated monitoring tools for collecting AI system performance, telemetry, and impact data (Measure); conducting periodic and event-driven risk assessments, trend analysis, and root cause analysis (Analyze); and systematically ranking identified risks based on severity, likelihood, and alignment with organizational risk appetite to inform treatment plans and resource allocation (Prioritize). Integration points with AI development, deployment, and operational lifecycles must be defined, ensuring outputs from MAP activities feed directly into control refinements, risk treatment plans, and policy updates. Regular training on the AI RMF and MAP processes must be provided to all relevant stakeholders.

Control Frequency:

- **Measure**: Continuous (automated data collection), Monthly (manual data review/reporting).
- **Analyze & Prioritize**: Quarterly (structured reviews), or ad-hoc upon significant changes

(e.g., model retraining, new use cases, regulatory updates, incident occurrence, or changes in operational context).

Monitoring Frequency:
* **Internal Oversight**: Quarterly by AI Governance Team leadership.
* **Executive Review**: Bi-annually by relevant committees (e.g., Risk Committee, AI Steering Committee).
* **Independent Assurance**: Annually by Internal Audit or an independent assurance provider.

Control Type: Semi-automated (leveraging automated data capture for measurement, with manual oversight, expert analysis, and prioritization by the AI Governance Team).

Guide for Evaluating Control Effectiveness: Effectiveness can be evaluated by reviewing documented MAP procedures and evidence of their execution (e.g., complete and up-to-date AI risk registers, documented risk assessments, action plans, and remediation tracking). Assess whether identified risks are appropriately classified, prioritized, and mitigated in a timely manner. Verify through stakeholder interviews that roles, responsibilities, and processes are understood and adhered to. Examine incident reports to confirm that the MAP function effectively identified and addressed root causes or contributing risk factors. Analyze trends in critical risk exposure and the rate of new risk identification versus remediation.

Key Control Indicators (KCIs):
* **Percentage of AI systems with a current risk assessment through MAP**: Target > 95%.
* **Average time from identification of an emerging AI risk to its assessment and initial prioritization**: Target < 30 days.
* **Number of critical/high-severity AI risks lacking a defined treatment plan**: Target 0.
* **Percentage of high-priority risk mitigation actions completed on schedule**: Target > 90%.
* **Reduction in the number of AI-related incidents attributable to unmanaged risks**: Continuous downward trend.
* **Frequency of AI Governance Team meetings dedicated to MAP reviews**: Consistent with defined schedule (e.g., 4+ times per year).

Requirement Statements:

1. It is incum- bent on Framework users to continue applying the MAP function to AI systems as context, capabilities, risks, benefits, and potential impacts evolve over time. (Page: 30)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

49. AI System Functionality and Trustworthiness Measurement and Documentation for Risk Management

Control Actor: Risk Management Team, Information Security Officer

Control Types: administrative, preventive, detective, corrective, compensating

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain comprehensive processes for the Risk Management Team to track relevant metrics for AI systems' trustworthy characteristics (e.g., validity, reliability, safety, fairness, privacy, security, transparency), social impact, and human-AI configurations. These processes are integral to effectively measuring AI risks. Concurrently, ensure that all critical aspects of AI systems' functionality and trustworthiness, including the specific measurement approaches used for identifying AI risks, are thoroughly documented as part of these AI risk measurements, along with the results of such measurements, including identified measurable performance improvements or declines. Utilize these measurement outcomes to inform management decisions regarding recalibration, impact mitigation, system removal, or implementation of compensating, detective, deterrent, directive, and recovery controls, particularly when tradeoffs among trustworthy characteristics arise. This supports informed decision-making and continuous improvement in AI risk management.

Enhanced Implementation Guide:

The enhanced compliance control dictates that the Organization, through its Risk Management Team and in collaboration with the Information Security Officer, shall establish, maintain, and continuously execute comprehensive processes for measuring, tracking, and meticulously documenting the functionality, trustworthiness characteristics (including validity, reliability, safety, fairness, privacy, security, and transparency), social impact, and human-AI configurations of all AI systems, encompassing those from Provider Organizations. This requires the thorough documentation of specific measurement methodologies employed for identifying AI risks, the detailed results of such measurements (including identified performance improvements or declines), and the subsequent management decisions (e.g., recalibration, impact mitigation, system removal, or implementation of appropriate compensating, detective, deterrent, directive, and recovery controls), particularly emphasizing decisions made when tradeoffs among trustworthy characteristics are identified. **Control Implementation** involves defining quantifiable metrics for each characteristic, establishing integrated data collection mechanisms (both automated and manual) across the AI lifecycle, setting clear performance thresholds, and formalizing documentation standards including templates for measurement plans, results, and decision rationales. A structured review and decision-making process, involving relevant stakeholders, must be established to ensure timely and informed responses to identified risks. This control operates with **continuous frequency** for active AI systems, supplemented by periodic (e.g., quarterly or semi-annual) comprehensive reviews and ad-hoc executions triggered by significant AI system changes or incidents. The control is **semi-automated**, with automated aspects for data collection and monitoring, while analysis, interpretation, documentation, and management decision-making remain manual. **Control Effectiveness Evaluation** will ascertain the existence of documented policies and processes, verify the consistent and timely operation through evidence of measurements, complete documentation, and clear audit trails linking results to actions, and confirm its effectiveness by assessing the accuracy of risk identification, promptness of addressing issues, and whether decisions lead to desired outcomes, particularly concerning the intentional management of characteristic tradeoffs. **Key Control Indicators (KCIs)** for performance measurement include the percentage of AI systems with defined and measured trustworthiness metrics, the completeness and timeliness of AI risk measurement documentation, the number of identified AI risks directly attributed to trustworthiness issues, the average time from risk identification to management decision/action, trends in AI system trustworthiness scores, the proportion of management decisions directly informed by measurement outcomes, the documented resolution of characteristic tradeoffs, and outcomes from internal/external audits of AI risk measurement processes.

Requirement Statements:

1. AI risk measurements include documenting aspects of systems’ functionality and trustworthiness. (Page: 33)

2. Measuring AI risks includes tracking metrics for trustworthy characteristics, social impact, and human-AI configurations. (Page: 33)

3. Where tradeoffs among the trustworthy characteristics arise, measurement provides a traceable basis to inform management decisions. Options may include recalibration, impact mitigation, or removal of the system from design, development, production, or use, as well as a range of compensating, detective, deterrent, directive, and recovery controls. (Page: 33)

4. Measurement outcomes will be utilized in the MANAGE function to assist risk monitoring and response efforts. (Page: 33)

5. MEASURE 4.1: ...Ap- proaches are documented. (Page: 36)

6. MEASURE 4.2: ...Results are documented. (Page: 36)

7. MEASURE 4.3: ...are identified and documented. (Page: 36)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

50. Independent Review of AI System Testing and Bias Mitigation

Control Actor: Internal Audit

Control Types: administrative, detective, corrective, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement processes for independent review by Internal Audit to improve the effectiveness of AI system testing and to proactively identify and mitigate internal biases and potential conflicts of interest within AI system development, deployment, and evaluation processes. This ensures objectivity and enhances the trustworthiness of AI systems by providing an impartial assessment of existing control effectiveness and risk management practices.

Enhanced Implementation Guide:

The enhancement of the compliance control record for "Independent Review of AI System Testing and Bias Mitigation" is detailed as follows:

Control Statement: Establish and maintain an independent Internal Audit function to provide objective assurance over the effectiveness of AI system testing, comprehensive bias identification, and robust mitigation controls, including proactive identification and management of potential conflicts of interest throughout the AI system development, deployment, and evaluation lifecycle, thereby enhancing AI trustworthiness and objectivity.

High-Level Control Implementation Guide: Internal Audit will develop and execute risk-based audit programs specifically tailored for AI systems, commencing with a thorough review of AI governance frameworks, model risk management policies, and adherence to relevant legal, ethical, and internal guidelines. This involves assessing the adequacy of AI model validation reports, performance metrics, and robustness testing results against established internal standards and industry best practices. Furthermore, Internal Audit will evaluate the sufficiency of fairness metrics, bias detection techniques, and the completeness of data provenance records and ethical impact assessments. Crucially, procedures will be established to scrutinize processes for identifying and managing conflicts of interest within AI development teams, data selection, model training, and deployment decisions. The review will involve examining documented policies and procedures, performing independent data analysis or reperformance of selected controls and tests where appropriate, and conducting interviews with key stakeholders such as AI developers, data scientists, product owners, legal counsel, and ethics committee members. Audit findings will be formally reported, actionable recommendations will be provided to management, and the timely remediation of identified deficiencies will be diligently tracked. Internal Audit will ensure its staff possess adequate AI-specific knowledge through ongoing training or engage qualified subject matter experts as needed to maintain technical proficiency.

Control Frequency: This control should be performed at least annually for critical AI systems as determined by a comprehensive risk assessment, and/or upon the launch or significant update of high-impact AI models, or as triggered by emerging risks or significant regulatory changes.

Monitoring Frequency: The performance of this control should be monitored quarterly by the Head of Internal Audit to ensure adequate coverage, resourcing, and effectiveness of the audit function’s AI assurance activities, with an annual comprehensive review conducted by the Audit Committee.

Control Type: Predominantly Manual, leveraging Semi-Automated tools for data analysis, documentation review, and audit program execution to enhance efficiency and coverage.

Guide for Evaluating Control Effectiveness: Control effectiveness will be evaluated by assessing the completeness, accuracy, and actionability of Internal Audit reports, particularly regarding the clarity and relevance of findings related to testing gaps, identified biases, or conflicts of interest. Key indicators of effectiveness include the timely and effective remediation rate of audit findings, especially high-risk deficiencies pertaining to AI bias and testing, and the adequacy of audit coverage across the organization's critical AI systems and the various stages of the AI lifecycle. Positive stakeholder feedback on the value, objectivity, and constructiveness of the independent reviews from AI development teams and senior management will also be considered. Ultimately, the absence of significant AI-related incidents (e.g., regulatory fines, public relations crises, operational failures, or demonstrable unfair outcomes) directly attributable to unmitigated biases or testing deficiencies that should have been identified by the independent review will serve as a strong indicator of effectiveness.

Key Control Indicators (KCIs):
*   **Coverage Rate:** Percentage of critical AI systems reviewed by Internal Audit versus the total critical AI systems identified (target: 100%).
*   **Finding Severity & Volume:** Average number of high-severity AI bias mitigation

deficiencies identified per audit engagement.
* **Testing Strength:** Average number of high-severity AI system testing control weaknesses identified per audit engagement.
* **Remediation Rate:** Percentage of high-risk AI-related audit findings remediated within agreed-upon timelines (target: >90%).
* **Issue Closure Time:** Average time to closure for high-risk AI-related audit findings.
* **Conflict Identification:** Number of instances of identified and addressed conflicts of interest within AI development and deployment processes.
* **Competency Development:** Internal Audit team's cumulative hours of AI-specific training or professional certifications acquired per quarter/year.

Requirement Statements:

1. Processes for independent review can improve the effectiveness of testing and can mitigate internal biases and potential conflicts of inter- est. (Page: 33)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ___________________________________

Evidence/Comments:

____

____

____

51. AI Measurement Methodologies Adherence and Transparency

Control Actor: Legal & Compliance

Control Types: administrative, preventive, detective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes to ensure that all AI metrics and measurement methodologies adhere strictly to established scientific, legal, and ethical norms. This includes implementing practices to conduct these measurements in an open and transparent manner, providing clarity on their design, application, and results to foster trust and accountability across the AI lifecycle.

Enhanced Implementation Guide:

Control Statement: Establish and enforce robust processes to ensure that all AI metrics and measurement methodologies employed across the organization and by its provider organizations strictly adhere to established scientific, legal, and ethical norms. This encompasses implementing comprehensive practices to conduct these measurements with demonstrable openness and transparency, providing clear, auditable documentation on their design, application, and results to foster trust, accountability, and explainability throughout the entire AI lifecycle.

High-Level Control Implementation Guide: Implementation requires a multi-faceted approach. Firstly, define and formalize a comprehensive AI Measurement Policy that articulates the organization's scientific rigor (e.g., statistical validity, data provenance), legal compliance (e.g., data protection, non-discrimination laws), and ethical principles (e.g., fairness, accountability, transparency) for all AI metrics and methodologies. Secondly, establish an AI Governance Committee, composed of representatives from Legal, Compliance, Data Science, Ethics, and relevant business units, responsible for reviewing and approving all AI measurement methodologies prior to model deployment or significant modification. This committee will ensure methodologies align with policy, adequately assess risks (e.g., bias, privacy impact), and are appropriately documented. Thirdly, mandate the creation and maintenance of detailed documentation for each AI model's measurement methodology, covering metric selection rationale, data sources, preprocessing, validation techniques, assumptions, limitations, and impact assessments. Fourthly, develop and implement transparency protocols, including standardized reporting templates and accessible dashboards, to clearly communicate measurement design, application, and results to relevant internal stakeholders (e.g., leadership, model owners) and, where legally or ethically required, external parties (e.g., regulators, affected individuals). Fifthly, integrate these requirements into the AI development lifecycle, from design and training to deployment, monitoring, and retirement, ensuring continuous adherence. Finally, for third-party AI solutions, contractual agreements must stipulate adherence to equivalent standards, coupled with audit rights to verify compliance with established measurement methodologies and transparency requirements.

Control Frequency:
* **Preventive:** Performed prior to the deployment of any new AI model or a significant modification to an existing model's measurement methodology.
* **Detective:** Conducted during quarterly or bi-annual reviews of all operational AI systems.
* **Reactive:** Initiated immediately upon detection of an anomaly, incident, or stakeholder complaint related to AI measurement integrity, accuracy, or transparency.

Monitoring Frequency:
* **Continuous:** Automated monitoring of AI model performance against defined metrics and thresholds (e.g., drift, bias indicators, explainability scores) for production systems.
* **Quarterly/Bi-annually:** The AI Governance Committee or an independent oversight function should review samples of AI model measurement methodologies, documentation, and transparency reports.
* **Annually:** A comprehensive audit of AI measurement methodology adherence and transparency practices across the entire organization, potentially including third-party engagements.

Control Type: Semi-Automated.
* **Automated Components:** Include continuous monitoring of AI model performance metrics, automated checks for data integrity, bias detection, and anomaly flagging against predefined thresholds.
* **Manual Components:** Encompass policy drafting and approval, methodology review and approval by human experts/committees, manual documentation and attestation, design of transparency frameworks, stakeholder communication, and independent internal audit activities.

Guide for Evaluating Control Effectiveness: Evaluating effectiveness involves assessing both process adherence and outcome quality. Key evaluation steps include:

1. **Policy & Procedure Verification:** Confirm the existence, formal approval, and regular review of the AI Measurement Policy and supporting procedures.
2. **Methodology Approval Traceability:** Inspect records (e.g., AI Governance Committee minutes, formal approvals, risk assessments) to verify that all new or significantly modified AI model measurement methodologies have undergone the prescribed review and approval process.
3. **Documentation Completeness & Accuracy:** Sample AI models in production and critically assess the completeness, accuracy, clarity, and accessibility of their associated measurement methodology documentation, ensuring it adequately addresses scientific, legal, and ethical considerations.
4. **Transparency Protocol Adherence:** Review communication protocols, reporting mechanisms, and actual reports/dashboards to confirm that AI measurement results, design, and application are clearly and transparently communicated to relevant internal and, where applicable, external stakeholders in accordance with policy.
5. **Training & Awareness Assessment:** Verify that relevant personnel (e.g., AI developers, data scientists, model owners) have completed mandatory training on AI measurement standards, ethical considerations, and transparency requirements.
6. **Audit & Incident Review:** Analyze findings from internal and external audits, as well as incidents, complaints, or escalations related to AI measurement accuracy, fairness, or transparency, to identify root causes and control deficiencies.
7. **Third-Party Oversight Review:** Examine contractual agreements, audit reports (e.g., SOC 2 Type 2 for AI-as-a-Service providers), and assurance statements from third-party AI providers to ensure alignment with organizational measurement and transparency standards.

Key Control Indicators (KCIs):

- **Percentage of AI models with formally approved measurement methodologies:** Target 100%. This indicates preventive control effectiveness.
- **Number of documented deviations or exceptions from established AI measurement standards:** Target zero or near-zero, with clear remediation plans for any identified issues.
- **Completion rate of mandatory AI ethics and measurement training for relevant staff:** Target >95% to ensure competency and awareness.
- **Average internal audit rating/score for AI measurement and transparency controls:** Aim for consistently high scores (e.g., no critical findings; average severity of findings < 2 on a 5-point scale).
- **Number of regulatory inquiries, external complaints, or critical incidents related to AI measurement accuracy, fairness, or transparency:** Aim for zero, demonstrating effective risk mitigation.
- **Mean time to resolve critical deficiencies identified in AI measurement**

methodologies: Lower is better (e.g., < 30 days), indicating prompt corrective action.
* **Percentage of production AI models with continuous, automated performance and bias monitoring enabled:** Target >90%, showing integrated detective controls.
* **Documentation completeness score:** Average score across sampled AI models for adherence to documentation requirements (e.g., 90% completeness on a checklist).

Requirement Statements:

1. Metrics and measurement methodologies should adhere to scientific, legal, and ethical norms and be carried out in an open and trans- parent process. (Page: 33)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

52. Systematic Documentation for AI Risk Management and Transparency

Control Actor: Compliance Management Team, Information Security Officer

Control Types: administrative, preventive, detective

Related Entities: Organization, Provider Organization

Control Description:

Implement and maintain systematic documentation practices established in the GOVERN function and utilized across the MAP and MEASURE functions of the AI Risk Management Framework. These practices are crucial for bolstering overall AI risk management efforts, increasing transparency regarding AI systems, and enhancing accountability throughout their lifecycle.

Enhanced Implementation Guide:

This control mandates the establishment, implementation, and rigorous maintenance of systematic documentation practices, originating from the GOVERN function of the AI Risk Management Framework, to ensure comprehensive record-keeping, transparency, and accountability for AI systems throughout their lifecycle, explicitly including risk mapping (MAP) and performance measurement (MEASURE) activities. The high-level implementation guide includes: (1) Defining and enforcing mandatory documentation standards and templates for all AI systems, covering design, data, training, risk assessments, ethical considerations, performance metrics, and incident logs. (2) Integrating documentation requirements as mandatory checkpoints within the AI System Development Lifecycle (SDLC). (3) Clearly assigning roles and responsibilities for documentation creation, review, approval, and maintenance to relevant teams (e.g., AI development, data science, legal, compliance, risk). (4) Implementing a secure, centralized, and version-controlled repository (e.g., GRC platform, specialized AI governance tool) for all AI-related documentation, including that from Provider Organizations. (5) Providing regular training and awareness programs to all relevant personnel on documentation requirements and best practices. (6) Establishing a formal process for periodic review and update of documentation to reflect system changes, evolving risks, and regulatory amendments. This control operates with a **Continuous** frequency, meaning documentation is created and updated dynamically as AI systems progress through their lifecycle and in response to significant changes or events, supplemented by **Annual** comprehensive framework reviews. **Monitoring Frequency** should be **Quarterly** by the Compliance Management Team or Internal Audit, involving sampling AI system documentation for completeness, accuracy, and adherence to standards, with an **Annual** holistic assessment of the documentation framework's effectiveness. This control is primarily **Semi-automated**, leveraging structured templates, centralized digital repositories with version control, and potentially automated reminders for review and integration with development pipelines, while creation and review remain largely manual. Effectiveness is evaluated by assessing documentation completeness, accuracy, consistency, and adherence to defined standards across a sample of AI systems; verifying its accessibility, comprehensibility, and auditability; confirming timeliness of creation and updates; ensuring appropriate stakeholder involvement in documentation review and approval; and demonstrating that documentation is actively utilized in risk assessments, incident response, and decision-making processes. Key Control Indicators (KCIs) for measuring performance include: the percentage of AI systems with complete and approved documentation; the documentation review adherence rate (percentage of scheduled reviews completed on time); the number of non-conformities or weaknesses identified during internal or external audits related to AI documentation; the average time to retrieve specific AI system documentation during inquiries; the percentage of documentation updates following identified significant changes to AI systems; the training completion rate for personnel involved in documentation; and the number of documented deviations from prescribed templates.

Requirement Statements:

1. Systematic documentation practices established in GOVERN and utilized in MAP and MEASURE bolster AI risk management efforts and increase transparency and accountability. (Page: 36)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

53. AI System Inconsistent Performance Remediation and Deactivation Protocol

Control Actor: AI Operations Team, Senior Management

Control Types: preventive, detective, corrective, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish, implement, and apply comprehensive mechanisms and protocols, and ensure clear understanding and assignment of responsibilities, for the timely supersession, disengagement, or deactivation of AI systems. This control specifically applies when AI systems demonstrate performance or outcomes that are inconsistent with their intended use, ensuring immediate mitigation of potential negative impacts and maintaining the trustworthiness and safety of AI operations. This includes defining triggers for deactivation, escalation procedures, and communication plans for such critical actions.

Enhanced Implementation Guide:

The AI System Inconsistent Performance Remediation and Deactivation Protocol is a critical control designed to ensure the trustworthiness and safety of AI operations by promptly addressing deviations from intended performance or outcomes.

Control Statement: Establish, implement, and consistently apply a comprehensive protocol for the timely supersession, disengagement, or deactivation of AI systems when their performance or outcomes are inconsistent with their intended use, thereby ensuring immediate mitigation of potential negative impacts and maintaining AI system trustworthiness and safety. This protocol must clearly define triggers for deactivation, escalation procedures, assigned responsibilities, and communication plans for critical actions.

High-Level Control Implementation Guide:
1. **Protocol Definition:** Develop a formal "AI System Inconsistent Performance & Deactivation Protocol" document. This protocol must define:
    * **Triggers:** Specific thresholds or conditions for performance inconsistency (e.g., drift in accuracy/precision/recall, bias detection beyond acceptable limits, ethical violations, security vulnerabilities, regulatory non-compliance, system instability, data quality degradation).
    * **Roles & Responsibilities:** Clearly assign roles for detection (e.g., AI Operations, MLOps

engineers), assessment (e.g., Data Scientists, AI Governance Committee), decision-making for deactivation/remediation (e.g., Senior Management, Product Owners), execution (e.g., IT Operations, AI Operations), and communication (e.g., Legal, PR, Senior Management).

* **Escalation Matrix:** A multi-tiered escalation path detailing whom to notify, when, and with what information based on the severity and urgency of the inconsistency.
* **Deactivation/Remediation Procedures:** Step-by-step guides for:
    * Safe and graceful deactivation (e.g., stopping inferences, redirecting traffic, data archival).
    * Root cause analysis and impact assessment.
    * Remediation options (e.g., model retraining, data cleansing, re-engineering, system recalibration).
    * Testing and re-validation criteria before re-deployment or activation of a successor system.
* **Communication Plan:** Internal (e.g., affected teams, leadership) and external (e.g., customers, regulators, third-party providers if applicable) communication templates and channels, including required disclosures.
* **Record Keeping:** Requirements for documenting all detected inconsistencies, assessment findings, decisions, actions taken, timelines, and outcomes.
* **Post-Mortem/Review:** A process for conducting post-deactivation reviews to identify lessons learned and improve the protocol.

2. **Implementation & Training:**
    * Formally approve the protocol by relevant senior management.
    * Disseminate the protocol to all relevant stakeholders.
    * Conduct mandatory training for all personnel involved in AI operations, monitoring, incident response, and senior management on the protocol's content, roles, responsibilities, and procedures.
    * Integrate monitoring tools and dashboards to automate the detection of performance inconsistencies based on defined triggers.
3. **Application:**
    * Continuously monitor AI system performance against defined triggers.
    * Upon trigger activation, initiate the assessment and escalation procedures defined in the protocol.
    * Execute remediation or deactivation actions in a timely and controlled manner, adhering to the defined steps and communication plan.
    * Maintain detailed records of each incident and the actions taken.

Control Frequency:

* **Protocol Review:** Annually, or upon significant changes in AI systems, regulatory

landscape, or organizational risk appetite.
* **Control Performance (Application):** Ad-hoc, triggered immediately upon the detection of AI system performance or outcome inconsistencies.

Monitoring Frequency: Quarterly, as part of routine compliance monitoring activities, with annual deep-dive audits performed by internal audit or an independent compliance function.

Control Type: Semi-automated. Detection of inconsistencies can be largely automated through continuous monitoring tools and predefined thresholds. However, the assessment, decision-making (e.g., severity classification, choice between remediation vs. deactivation), escalation, communication, and execution of deactivation/remediation actions typically involve significant manual oversight, human judgment, and procedural steps.

Guide for Evaluating Control Effectiveness:
* **Documentation Adequacy:** Review the AI System Inconsistent Performance & Deactivation Protocol for completeness, clarity, and approval status. Verify that it addresses all required elements (triggers, roles, escalation, comms, etc.).
* **Training Verification:** Examine training records to confirm that all relevant personnel have received mandatory training on the protocol and understand their responsibilities.
* **Incident Sample Testing:** Select a sample of AI system inconsistent performance incidents (detected or reported) from the past monitoring period. For each incident:
    * Verify that the inconsistency was detected/reported promptly.
    * Confirm that the assessment and escalation procedures were followed according to the protocol.
    * Review the decision-making process for remediation or deactivation.
    * Verify the timeliness and effectiveness of the actions taken (e.g., remediation applied, system deactivations performed as per procedure).
    * Assess adherence to communication plans (internal/external).
    * Confirm complete and accurate record-keeping.
    * Evaluate if negative impacts were effectively mitigated as intended.
* **Stakeholder Interviews:** Conduct interviews with AI Operations, MLOps, Data Scientists, Risk, Legal, and Senior Management to assess their understanding of the protocol, roles, and challenges in execution.
* **Simulation/Tabletop Exercise:** Conduct a tabletop exercise or controlled simulation of an inconsistency event to test the protocol's effectiveness in real-time response scenarios (optional, but highly recommended for critical systems).

Key Control Indicators (KCIs):

* **Number of Inconsistent Performance Incidents Detected:** Total instances where AI system performance deviated from intended use.
* **Mean Time to Detect (MTTD) Inconsistency:** Average time from the occurrence of an inconsistency to its detection.
* **Mean Time to Remediate/Deactivate (MTTR/MTTD):** Average time from detection of inconsistency to the full resolution (remediation or deactivation).
* **Protocol Adherence Rate:** Percentage of inconsistent performance incidents where the deactivation/remediation protocol was fully followed (e.g., proper escalation, documentation, communication).
* **Effectiveness of Mitigation:** Number/percentage of incidents where potential negative impacts (e.g., financial loss, reputational damage, safety risks, fairness violations) were successfully mitigated or averted due to timely intervention.
* **Training Completion Rate:** Percentage of relevant personnel who have completed mandatory training on the protocol.
* **Number of Protocol Reviews Conducted:** Instances where the protocol was reviewed and updated as per schedule or triggered by significant changes.
* **Audit Findings:** Number of non-conformities or observations related to AI system inconsistent performance remediation and deactivation processes from internal or external audits.

Requirement Statements:

1. Mechanisms are in place and applied, and respon- sibilities are assigned and understood, to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use. (Page: 37)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

54. Pre-trained Model Monitoring and Maintenance for AI System Development

Control Actor: AI Operations Team, Information Security Officer

Control Types: detective, preventive, corrective, administrative

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for regularly monitoring and maintaining pre-trained models utilized in AI system development. This ensures their ongoing performance, integrity, and trustworthiness throughout the development lifecycle, as part of continuous AI system monitoring and maintenance. This process covers the identification and remediation of issues to uphold the reliability and security of AI systems.

Enhanced Implementation Guide:

This control mandates the establishment and continuous operation of a robust framework for monitoring and maintaining all pre-trained models integrated into AI systems throughout their entire lifecycle, from development to production operation.

Control Statement: To proactively ensure the sustained performance, integrity, trustworthiness, reliability, security, and ethical compliance of pre-trained models, thereby upholding the overall quality and dependability of AI systems by detecting and remediating issues like performance degradation, data/concept drift, bias, and security vulnerabilities.

High-Level Control Implementation Guide: The AI Operations Team, in collaboration with the Information Security Officer, shall define and implement a comprehensive model monitoring strategy. This includes identifying critical performance metrics (e.g., accuracy, precision, recall, F1-score), data drift indicators (e.g., input feature distributions), concept drift measures, fairness metrics (e.g., demographic parity, equalized odds), and explainability stability. Automated monitoring tools must be deployed to continuously collect these metrics and compare them against predefined thresholds, triggering alerts upon deviation. A clear incident response and remediation process must be established, outlining steps for root cause analysis, model retraining, recalibration, patching, or replacement. Version control for models and associated datasets is mandatory, alongside thorough documentation of all monitoring activities, findings, and remediation actions. Integration with MLOps pipelines is crucial to automate monitoring deployment and remediation workflows.

Control Frequency: Monitoring of critical performance and drift metrics shall be continuous (real-time or near real-time) through automated systems. Comprehensive model health reviews and scheduled maintenance activities (e.g., validation of model retraining, security audits) shall occur at a minimum of quarterly, or upon significant changes to input data, model architecture, or business requirements.

Monitoring Frequency: The effectiveness of this control itself shall be reviewed and assessed quarterly by internal audit or the AI governance committee, as part of regular internal controls assurance activities, to ensure documented procedures are followed and performance targets are met.

Control Type: Semi-automated. This control heavily leverages automated tools for continuous data collection, drift detection, and alert generation. However, human intervention is required for defining thresholds, conducting root cause analysis of detected issues, making decisions on remediation strategies (e.g., whether to retrain, recalibrate, or redeploy), and overseeing the execution of complex maintenance tasks.

Guide for Evaluating Control Effectiveness: Evaluation of control effectiveness involves verifying the existence and approval of a formal Model Monitoring and Maintenance Policy and associated Standard Operating Procedures (SOPs). This includes examining logs from automated monitoring systems to confirm continuous operation, appropriate threshold settings, and evidence of alerts being generated when deviations occur. Reviewing incident management records will confirm timely triage, root cause analysis, and resolution of identified model issues, ensuring adherence to defined Mean Time To Detect (MTTD) and Mean Time To Remediate (MTTR) service level agreements (SLAs). Further, sampling models from the active inventory will confirm their inclusion in the monitoring regime, and validating version control logs will ensure traceability of model updates and remediations. Documentation of regular review meetings involving the AI Operations Team and Information Security Officer, including discussions of monitoring reports and remediation strategies, provides additional evidence of control operation.

Key Control Indicators (KCIs):
* **Coverage Rate:** Percentage of production pre-trained models under active, automated monitoring.
* **Drift Detection Rate:** Number of instances of significant data or concept drift detected per month/quarter.

*   **Mean Time To Detect (MTTD):** Average time from the occurrence of a model performance degradation or drift event to its detection.
*   **Mean Time To Remediate (MTTR):** Average time from detection of a model issue to its full resolution and model redeployment.
*   **Remediation Success Rate:** Percentage of remediated models that show improved or restored performance post-intervention.
*   **Alert-to-Resolution Ratio:** Ratio of critical alerts generated to critical alerts successfully resolved within specified SLAs.
*   **Bias Fluctuation Index:** A metric tracking the stability or change in fairness metrics over time, indicating potential bias amplification.
*   **Security Vulnerability Scan Frequency:** Regularity of security scanning for models and their underlying components.

Requirement Statements:

1. Pre-trained models which are used for develop- ment are monitored as part of AI system regular monitoring and maintenance. (Page: 37)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

55. AI System Post-Deployment Operational Management and Lifecycle Protocols

Control Actor: Information Security Officer

Control Types: administrative, preventive, detective, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish and implement comprehensive post-deployment AI system operational management plans. These plans must include specific mechanisms for continuous monitoring of AI system performance and impacts, capturing and evaluating input from users and other relevant AI actors, defining clear processes for appeal and override of AI system decisions, establishing protocols for decommissioning AI systems when no longer needed or performing as intended, and implementing robust change management procedures for system updates and modifications. These measures, overseen by the Information Security Officer, ensure ongoing trustworthiness, accountability, and adaptability throughout the AI system's operational lifecycle.

Enhanced Implementation Guide:

Ensure ongoing trustworthiness, accountability, and adaptability of AI systems throughout their operational lifecycle post-deployment through the establishment and rigorous implementation of comprehensive operational management plans. This involves the continuous monitoring of AI system performance, accuracy, and societal impacts; the diligent capture and evaluation of input from users and relevant AI actors; the definition and clear communication of processes for appeal and human override of AI system decisions; the establishment of robust protocols for the secure and compliant decommissioning of AI systems when no longer needed or performing as intended; and the implementation of robust change management procedures for all system updates and modifications. The Information Security Officer is responsible for overseeing the implementation and adherence to these protocols.

- **Control Statement:** Establish, implement, and maintain comprehensive post-deployment operational management plans for all AI systems to ensure their continuous performance, accountability, and adaptability across their lifecycle, encompassing monitoring, feedback, decision override, change management, and decommissioning.
- **High-Level Control Implementation Guide:** Develop a standardized "AI System

Operational Management Framework" or specific plans for each AI system encompassing: (a) Automated and manual mechanisms for continuous monitoring of AI system performance (e.g., accuracy, bias, drift, resource utilization) and impacts (e.g., fairness, unintended consequences), with defined thresholds for alerts. (b) Formal channels for capturing and evaluating input from users and AI actors, including dedicated feedback loops, incident reporting, and structured review processes. (c) Documented, accessible, and communicated procedures for users to appeal or request override of AI system decisions, detailing escalation paths, human review criteria, and decision-making authorities. (d) Detailed "AI System Decommissioning Protocols" outlining data retention/deletion, model archival, system shutdown, and stakeholder notification steps. (e) Integration of AI system updates and modifications into the organization's existing change management framework, requiring comprehensive risk assessment, impact analysis, pre-implementation testing, version control, and formal approvals. The Information Security Officer, in collaboration with AI System Owners, MLOps, and Data Governance teams, is responsible for the design, implementation, and oversight of these protocols.

- **Control Frequency:** Continuous operational execution based on established protocols (e.g., ongoing monitoring, ad-hoc feedback/appeal processing, event-driven change management and decommissioning), with underlying plans and procedures reviewed and updated at least annually or when significant changes occur to AI systems or regulatory requirements.
- **Monitoring Frequency:** Quarterly, with ad-hoc reviews triggered by significant AI system incidents, performance deviations, major updates, or new AI system deployments, conducted by the Information Security Officer or a delegated governance function.
- **Control Type:** Semi-automated (leveraging automated tools for continuous monitoring of performance and impacts, complemented by manual review, decision-making, procedural adherence, and human intervention for feedback evaluation, appeal/override, and change management approvals).
- **Guide for Evaluating Control Effectiveness:** Verify the existence, completeness, and approval of AI operational management plans, monitoring reports/dashboards, feedback logs, appeal records, change management documentation (e.g., test results, risk assessments), and decommissioning protocols. Conduct process walkthroughs for a sample of AI systems, tracing execution of monitoring, feedback resolution, appeal processing, and change implementation. Interview relevant personnel (e.g., AI System Owners, MLOps Engineers, Data Scientists, Legal, Compliance) to confirm awareness and consistent application of protocols. Review incident reports to assess the effectiveness of these controls in managing AI-related incidents or deviations.
- **Key Control Indicators:** Percentage of deployed AI systems with a documented and approved operational management plan; average time to identify and resolve AI system

performance deviations or bias concerns; number of documented user feedback submissions and their resolution rates; number of AI system decision appeal requests and average resolution time; adherence rate to AI system change management procedures (e.g., percentage of changes following full protocol); number of incidents related to unauthorized AI system modifications or uncontrolled decommissioning; percentage of AI system decommissioning protocols completed within defined timelines.

Requirement Statements:

1. MANAGE 4.1: Post-deployment AI system monitoring plans are implemented, including mechanisms for capturing and eval- uating input from users and other relevant AI actors, appeal and override, decommissioning, incident response, recovery, and change management. (Page: 38)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

56. AI System Updates Continual Improvement and Stakeholder Engagement

Control Actor: Product Management, AI Operations Team

Control Types: administrative, preventive, detective, corrective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes to integrate measurable activities for continual improvements into AI system updates and include regular engagement with interested parties, including relevant AI actors. This ensures ongoing enhancement and responsiveness to evolving needs, risks, and stakeholder feedback throughout the AI system lifecycle, aligning with the MANAGE function of the AI RMF.

Enhanced Implementation Guide:

The organization shall establish and maintain auditable processes for continual improvement and stakeholder engagement regarding AI system updates, ensuring AI systems evolve responsively throughout their lifecycle. **High-level Control Implementation Guide:** Product Management and the AI Operations Team will establish structured processes for feedback collection (e.g., dedicated channels, surveys, direct interaction points) from all relevant internal and external stakeholders, including users, developers, legal, risk, and affected parties. This feedback, along with AI system performance data, identified risks, and evolving regulatory requirements, will be systematically reviewed and analyzed (e.g., via regular steering committee meetings or review boards) to identify, prioritize, and formally integrate measurable improvement activities into AI system development and update roadmaps. Ownership for implementing these improvements will be clearly assigned, and their progress tracked. A robust stakeholder engagement strategy will be developed and executed, ensuring regular, transparent communication of AI system update plans, progress, and outcomes to all interested parties, actively soliciting their input. Comprehensive documentation will be maintained for all feedback received, analysis conducted, decisions made, updates implemented, measurable improvements achieved, and stakeholder engagement activities (e.g., meeting minutes, communication logs, feedback registers). **Control Frequency:** Feedback collection is continuous; formal review and prioritization occur quarterly or more frequently based on criticality; AI system updates incorporating improvements are integrated based on established sprint/release cycles (e.g., bi-weekly, monthly, quarterly); formal stakeholder engagement sessions occur at least bi-annually, with ongoing ad-hoc communication as needed. **Monitoring Frequency:** The control's effectiveness will be monitored quarterly through internal compliance checks and annually via formal audit. **Control Type:** This is a semi-automated control, leveraging manual processes for decision-making, stakeholder engagement, and strategic planning, complemented by automated tools for aspects like performance monitoring, feedback aggregation platforms, and update deployment pipelines. **Guide for Evaluating Control Effectiveness:** Control effectiveness will be evaluated by verifying the existence, adherence to, and maturity of documented processes for feedback management, improvement integration, and stakeholder engagement. This includes reviewing evidence such as auditable logs of feedback received, clear records of analysis performed and decisions made on how feedback was addressed (or why it wasn't), traceability of implemented AI system updates back to identified improvement opportunities or stakeholder input, detailed stakeholder engagement records (e.g., meeting minutes, attendance logs, communication summaries), and reports demonstrating measurable improvements in AI system performance, fairness, or risk mitigation following updates. Furthermore, confirming clear role assignments and accountability for these processes within Product Management and AI Operations Team is crucial. **Key Control Indicators (KCIs):** Key performance metrics include the percentage of stakeholder feedback items formally reviewed and addressed within defined service level agreements (SLAs), the frequency and documented attendance rates of structured stakeholder engagement forums, the number of critical AI system risks identified and successfully mitigated through updates per quarter/year, the trend of key AI system performance metrics (e.g., accuracy, bias metrics, resource utilization) demonstrating measurable improvement post-updates, and the rate of AI system updates successfully deployed that directly incorporate identified improvements or risk mitigations.

Requirement Statements:

1. MANAGE 4.2: Measurable activities for continual improvements are integrated into AI system updates and include regular engage- ment with interested parties, including relevant AI actors. (Page: 38)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

57. AI Risk Management Resource Allocation and Prioritization

Control Actor: Senior Management

Control Types: administrative, preventive

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for Senior Management to effectively gauge and prioritize the allocation of necessary resources, including staffing and funding, to achieve AI risk management goals in a cost-effective and efficient manner. This ensures that organizational investments in AI risk management are optimized and aligned with strategic priorities and identified risk levels.

Enhanced Implementation Guide:

Control Statement: Senior Management must establish, document, and consistently execute a formal, risk-based process to assess, prioritize, and allocate adequate and appropriate resources (financial, human capital, technological infrastructure, and specialized tools) for AI risk management activities. This process must ensure resource allocation is cost-effective, auditable, and demonstrably aligned with the organization's strategic AI objectives, identified AI risk profile (including ethical, security, privacy, and performance risks), and relevant regulatory obligations, thereby optimizing investments in AI risk mitigation and ensuring the sustainability of the AI risk management framework.

High-Level Control Implementation Guide:

1. **Establish an AI Risk Management Resource Governance Framework:** Define clear roles, responsibilities, and decision-making authorities for AI risk management resource allocation within Senior Management, potentially via an AI Governance Committee or a designated cross-functional steering group. This framework should outline the process for identifying resource needs, submitting requests, review, approval, and subsequent monitoring.
2. **Develop a Structured Risk-Based Prioritization Methodology:** Implement a quantifiable methodology to assess and categorize AI risks (considering likelihood, impact, and criticality of the AI system), allowing for prioritization of mitigation efforts and resource allocation based on enterprise-wide risk appetite, strategic importance of AI initiatives, and potential impact on business objectives, individuals, and stakeholders.
3. **Conduct Regular, Comprehensive Resource Needs Assessments:** Periodically (e.g.,

annually for strategic planning, quarterly for operational adjustments) conduct thorough assessments of current and projected AI risk management resource needs. This involves gathering input from AI development teams, legal, compliance, IT security, data privacy, and ethics functions, considering existing AI deployments, new AI initiatives in the pipeline, emerging AI risks, and evolving regulatory landscapes.

4. **Integrate Resource Allocation into Organizational Planning:** Embed AI risk management resource requirements into the overall organizational budgeting cycle, strategic workforce planning, and capital expenditure processes. This includes allocating dedicated budget lines for AI risk management training, specialized personnel (e.g., AI ethicists, MLOps security engineers), technological solutions (e.g., AI model monitoring tools, risk assessment platforms), and external expertise where necessary.
5. **Perform Cost-Benefit Analysis and ROI Justification:** For significant resource allocations, conduct comprehensive cost-benefit analyses to justify investments, demonstrating the value proposition of robust AI risk management in terms of reduced potential losses, enhanced trust, and compliance, balancing risk reduction with operational efficiency.
6. **Maintain Comprehensive Documentation and Communication:** Document all resource allocation policies, procedures, decisions, and the rationale behind prioritization. Communicate these decisions effectively across relevant departments and to key stakeholders to ensure transparency, understanding, and accountability for AI risk management initiatives.
7. **Implement a Continuous Review and Adjustment Mechanism:** Establish formal mechanisms for ongoing review and iterative adjustment of resource allocations based on evolving AI risks, the effectiveness of previously implemented controls, changes in organizational strategic priorities, and shifts in the regulatory or threat landscape.

Control Frequency:

* **Quarterly:** Formal review and adjustment of AI risk exposure and resource allocations by Senior Management or delegated committee.
* **Annually:** Comprehensive resource assessment and allocation as part of the overall organizational strategic planning and budgeting cycle.
* **Ad-hoc:** Immediately upon the launch of new high-risk AI initiatives, significant changes in regulatory requirements, or identification of major AI-related incidents/near misses.

Monitoring Frequency:

* **Continuous (Embedded):** Operational tracking of resource utilization against project plans and budgets by relevant department heads.
* **Quarterly (Oversight):** Review of resource allocation decisions, budget vs. actuals for AI risk management, and alignment with the organization's AI risk profile by the designated oversight function (e.g., Risk Management, Compliance, or an AI Governance Office).

*   **Annually (Assurance):** Independent assurance review by Internal Audit or an external assurance provider of the effectiveness, efficiency, and alignment of the resource allocation process and its outcomes against the organization's AI risk management objectives and regulatory requirements.

Control Type: Semi-automated. While the ultimate strategic decision-making, prioritization, and approval of resources remain manual functions of Senior Management, the process is significantly supported and enhanced by automated tools for:
*   Aggregating AI risk assessment data and generating reports (e.g., GRC platforms).
*   Tracking budgets, expenditures, and financial performance against AI risk management initiatives.
*   Managing AI project portfolios and resource assignments.
*   Supporting workforce planning and talent management for specialized AI risk management roles.

Guide for Evaluating Control Effectiveness:
1. **Documentation Adequacy:** Verify the existence, completeness, and currency of documented policies, procedures, and frameworks governing AI risk management resource allocation. Assess if they clearly define roles, responsibilities, and decision criteria.
2. **Decision-Making Records:** Review Senior Management meeting minutes, AI governance committee records, and budget approval documentation to confirm that discussions, decisions, and formal approvals for AI risk management resource allocations, including rationale and prioritization, are consistently occurring and recorded.
3. **Financial Alignment:** Compare allocated budgets for AI risk management activities against actual expenditures. Investigate and document explanations for any significant variances (e.g., >10-15% deviation) to ensure effective resource utilization.
4. **Staffing and Competency Adequacy:** Assess whether the organization has adequate skilled personnel (e.g., AI ethicists, AI security engineers, compliance analysts specializing in AI, data scientists with risk focus) dedicated to AI risk management. This includes reviewing relevant organizational charts, job descriptions, training records, and HR data for relevant certifications or experience.
5. **Risk Profile Alignment and Mitigation Success:** Evaluate if resource allocation demonstrably addresses identified high-priority AI risks and aligns with strategic AI objectives. This can be achieved by mapping allocated resources to specific risk mitigation initiatives and assessing the completion rates and effectiveness of these initiatives.
6. **Incident Response and Remediation Capability:** Assess the organization's demonstrated ability to effectively detect, respond to, and remediate AI-related incidents. Frequent or unresolved incidents, especially those linked to resource constraints, may indicate

ineffectiveness.
7. **Stakeholder Feedback and Satisfaction:** Solicit feedback from key stakeholders (e.g., AI development teams, legal, compliance, internal audit, business unit heads) regarding the perceived adequacy, timeliness, and effectiveness of AI risk management resources.
8. **Internal/External Audit Findings:** Review historical internal and external audit findings pertaining to AI risk management and resource allocation, verifying the implementation and effectiveness of agreed-upon remediation actions.

Key Control Indicators (KCIs):
* **AI Risk Management Budget Utilization Rate:** (Actual Spend on AI Risk Management / Allocated Budget for AI Risk Management) * 100%. (Target: >90% or within an acceptable variance range, indicating effective budget management).
* **Percentage of High-Priority AI Risks Remaining Unaddressed Due to Resource Constraints:** (Number of High-Priority AI Risks with Insufficient Resource Allocation / Total High-Priority AI Risks) * 100%. (Target: 0% or minimal, with documented rationale and planned remediation).
* **Ratio of Dedicated AI Risk Management FTEs to Total AI Development FTEs:** (Number of Full-Time Equivalents in AI Risk Management / Total Full-Time Equivalents in AI Development). (Target: Benchmark against industry best practices and internal needs, indicating adequate staffing).
* **Timeliness of AI Risk Assessment and Resource Review Completion:** Percentage of scheduled AI risk assessment and resource allocation reviews completed within the defined timeframe. (Target: >95%).
* **Number of AI-Related Incidents Directly Attributable to Inadequate Resources/Controls:** (Target: 0, indicating sufficient investment in preventive measures).
* **Completion Rate of Planned AI Risk Mitigation Projects:** (Number of Completed AI Risk Mitigation Projects / Total Planned AI Risk Mitigation Projects) * 100%. (Target: >90%, reflecting effective execution of resource-backed initiatives).
* **Percentage of AI Systems/Models Undergoing Formal Risk Assessment and Resource Allocation Review:** (Number of In-Scope AI Systems/Models with Documented Risk Assessments and Resource Review / Total In-Scope AI Systems/Models) * 100%. (Target: 100% for critical/high-risk systems).

Requirement Statements:

1. This risk-based approach also enables Framework users to compare their approaches with other approaches and to gauge the resources needed (e.g., staffing, funding) to achieve AI risk management goals in a cost- effective, prioritized manner. (Page: 38)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ___________________________________

Evidence/Comments:

____

____

____

58. AI RMF Cross-Sectoral Profile Application for Common Technologies and Business Processes

Control Actor: Risk Management Team

Control Types: administrative, preventive, detective

Related Entities: Organization, Provider Organization

Control Description:

Establish and maintain processes for the Risk Management Team to govern, map, measure, and manage risks for activities or business processes common across sectors (e.g., use of large language models, cloud-based services, or acquisition), by leveraging AI RMF cross-sectoral profiles. This ensures that standardized and tailored risk management approaches are applied effectively to prevalent AI technologies and services, promoting consistent trustworthiness and compliance across diverse organizational contexts.

Enhanced Implementation Guide:

Ensure the consistent and standardized application of AI RMF cross-sectoral profiles by the Risk Management Team (RMT) to identify, assess, and mitigate AI-related risks within common technologies (e.g., Large Language Models, cloud-based services) and business processes (e.g., third-party acquisition), thereby promoting uniform trustworthiness, compliance, and effective risk governance across the organization and its provider ecosystem.

High-level Control Implementation Guide: The Risk Management Team (RMT) must establish and maintain a formal policy and detailed procedures outlining the methodology for leveraging AI RMF cross-sectoral profiles. This includes: 1) Proactively identifying and maintaining an inventory of common AI technologies and business processes used across the organization and with provider entities. 2) Systematically mapping relevant AI RMF cross-sectoral profiles to these identified common areas, ensuring appropriate tailoring for specific organizational contexts. 3) Integrating this profile-driven risk assessment into the standard risk management lifecycle for new and evolving AI deployments and existing systems. 4) Defining clear metrics for AI risk and establishing a robust risk register to track, measure, and manage identified risks through their lifecycle. 5) Ensuring continuous training and awareness for the RMT and relevant stakeholders (e.g., procurement, legal, business units) on AI RMF application. 6) Fostering collaborative mechanisms with business process owners and third-party management teams to ensure comprehensive risk identification and mitigation.

Control Frequency: Continuously, as new common AI technologies or business processes are introduced or significantly modified; at least annually for existing common AI technologies and processes, or upon significant changes to the AI RMF or organizational risk appetite.

Monitoring Frequency: Quarterly by the Internal Audit or Compliance function, and as part of the semi-annual or annual enterprise risk review cycle.

Control Type: Semi-Automated (requires significant manual judgment and process execution, supported by GRC platforms, risk registers, and automated inventory tools for tracking and reporting).

Guide for Evaluating Control Effectiveness: Control effectiveness can be evaluated by: 1) Reviewing documented policies, procedures, and evidence of AI RMF cross-sectoral profile mappings for a sample of common AI technologies/processes. 2) Verifying that risk assessments for these samples were conducted using the defined methodology, with identified risks logged and managed in the organizational risk register. 3) Interviewing RMT members, business owners, and procurement personnel to confirm understanding, adherence, and effective collaboration in applying these profiles. 4) Assessing the timeliness and completeness of risk mitigation plans derived from these assessments. 5) Confirming that all relevant RMT personnel have undergone appropriate training on AI RMF and its cross-sectoral application.

Key Control Indicators (KCIs):

- **Percentage of common AI technologies/processes with documented AI RMF cross-sectoral profile application:** Target >95%.
- **Average time from identification of new common AI technology/process to completion of AI RMF-based risk assessment:** Target <30 days.
- **Number of open AI-related risks identified through cross-sectoral profile application that exceed defined risk appetite:** Target <5 at any given time.
- **Frequency of updates to the inventory of common AI technologies/processes and their associated profiles:** At least quarterly, or on a change basis.
- **Percentage of RMT members and relevant stakeholders trained on AI RMF and cross-sectoral profile application:** Target 100% within 3 months of joining or significant standard update.
- **Number of audit findings or compliance exceptions related to the application of AI RMF cross-sectoral profiles:** Target 0 per year.

Requirement Statements:

1. Cross-sectoral profiles can also cover how to govern, map, measure, and manage risks for activities or business processes common across sectors such as the use of large language models, cloud-based services or acquisition. (Page: 39)

Implementation Status:

☐ Implemented

☐ Partially Implemented

☐ Not Implemented

☐ Not Applicable. Reason: ________________________________

Evidence/Comments:

____

____

____

End of Document