Kiteworks

GUIDE

Kiteworks' Guide to the NIST AI Risk Management Framework

A Voluntary Framework for Trustworthy AI Development and Deployment

Table of Contents

Introduction

The National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF 1.0), released on January 26, 2023 (https://doi.org/10.6028/NIST.AI.100-1), establishes a voluntary framework to help organizations design, develop, deploy, and use artificial intelligence systems in ways that manage AI risks and promote trustworthy AI development. The framework outlines characteristics of trustworthy AI systems including:

- Valid and reliable
- Safe
- Secure and resilient
- Accountable and transparent
- Explainable and interpretable
- Privacy-enhanced
- Fair –– with harmful bias managed

The framework applies broadly to all organizations involved in AI system design, development, deployment, or use, regardless of sector or size. The framework is voluntary, rights-preserving, non-sector-specific, and use-case agnostic, providing flexibility to organizations of all sizes and in all sectors. The framework covers the full AI system life cycle across the stages shown in Figure 2 of the framework document, from Plan and Design through Operate and Monitor, with explicit subcategories addressing decommissioning (GOVERN 1.7, MANAGE 4.1).

The framework describes outcomes across four core functions. The GOVERN function calls on organizations to establish AI governance structures, policies, and accountability mechanisms. The MAP function describes activities for documenting AI system purposes, categorizing AI capabilities, and characterizing potential impacts. The MEASURE function describes activities for implementing metrics and evaluation processes to assess AI trustworthiness characteristics. The MANAGE function describes activities for prioritizing risks, maximizing benefits, managing third-party dependencies, and implementing post-deployment monitoring.

Organizations using this voluntary framework should implement specific technical and organizational measures to comply with its guidance. These include establishing risk management processes aligned to organizational risk tolerance, maintaining inventories of AI systems, implementing testing and validation procedures, and creating feedback mechanisms for stakeholders. The framework requires comprehensive documentation of AI system characteristics, limitations, and risk controls throughout the system life cycle. Organizations should also establish procedures for incident response, system decommissioning, and continuous improvement of AI systems.

While the framework itself is voluntary, organizations adopting it should follow its structured approach to risk management. The framework emphasizes the importance of diverse and multidisciplinary teams in AI development and deployment, regular engagement with affected communities, and transparent communication about AI system capabilities and limitations. Organizations should establish clear roles and responsibilities for AI risk management and ensure appropriate training for personnel involved in AI system development and operation.

This guide showcases how Kiteworks can support organizations in complying with specific sections of NIST AI RMF to achieve comprehensive data governance of their critical unstructured data.

The Kiteworks Secure File Sharing and Governance Platform

Kiteworks' file sharing and governance platform enables organizations to share sensitive information quickly and securely while maintaining full visibility and control over their file-sharing activities.

Protection of Unstructured Data

Kiteworks provides comprehensive protection for unstructured data through its advanced firewall and zero-trust file sharing capabilities that ensure sensitive unstructured data remains secure throughout its life cycle. The platform protects data whether at rest or in transit across various communication channels.

Governance and Compliance

Kiteworks reduces compliance risk by consolidating advanced data governance capabilities into a single platform. Whether employees send and receive data via email, file share, automated file transfer, APIs, or web forms, all activities are tracked and monitored.

Simplicity and Ease of Use

Kiteworks offers a user-friendly interface that simplifies secure file sharing and collaboration, enabling users to easily send, receive, and manage sensitive data without compromising security. Intuitive design and seamless integration with existing workflows ensure high user adoption rates, while specialized features like Secure Data Forms enable structured data collection.

AI Risk Management and Monitoring

Kiteworks provides specialized capabilities for AI risk management through comprehensive audit trails, real-time monitoring, and automated risk detection. The platform's Compliant AI controls AI access to enterprise data, enforces consent requirements, and maintains detailed activity logs. Advanced analytics and reporting features help organizations track AI system performance, detect anomalies, and respond to incidents in accordance with NIST AI RMF guidance.

Function-by-Function Mapping

GOVERN Function Outcomes

The GOVERN function requirements included in this guide address risk management process monitoring and review, AI system inventory aligned to organizational risk priorities, safe decommissioning and phase-out of AI systems, organizational practices for AI testing and incident identification, feedback mechanisms for incorporating input from AI actors into system design and implementation, and contingency processes for handling failures in high-risk third-party AI systems.

Function	Outcome	Kiteworks Solution
GOVERN 1.5	Ongoing monitoring and periodic review of the risk management process and its outcomes are planned and organizational roles and responsibilities clearly defined, including determining the frequency of periodic review.	Kiteworks supports ongoing monitoring and periodic review of data governance processes through its persistent audit event log, which captures all file, folder, user, mail, attachment, and comment events with entity resolution and detailed descriptions. The platform provides admin activity audit trails with list, detail, and CSV export capabilities for forensic review and periodic assessments. Compliance Reports from the Data Policy Engine document adherence to regulatory controls based on individual policy configurations, supporting systematic review of data governance outcomes. Admin role management allows organizations to define custom roles with fine-grained per-component permissions, ensuring clear assignment of responsibilities for data governance oversight.
GOVERN 6.2	Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk.	Kiteworks implements comprehensive contingency processes for handling third-party system failures through its high-availability cluster administration features that report cluster state, perform rolling reboots and OS upgrades with per-node prechecks/postchecks and migration capabilities. The platform includes database cluster failure recovery mechanisms that allow admin-triggered recovery of EPG database clusters with status tracking and progress checks for existing recoveries in flight. Maintenance mode toggles provide local and cluster-wide maintenance-mode scripts that block user traffic during upgrades or incident response. Cross-region replication with consistency checks ensures data availability through replication rules with thresholds and alerting, locked locations, replication statistics, and ACFS consistency checks/inconsistencies plus scheduled runner, maintaining system resilience even when third-party components fail.

MAP Function Outcomes

The MAP function requirements included in this guide address the definition of AI system tasks and methods, specification and documentation of targeted application scope based on system capability and context, and identification and documentation of internal risk controls for AI system components including third-party AI technologies.

Function	Outcome	Kiteworks Solution
MAP 3.3	Targeted application scope is specified and documented based on the system's capability, established context, and AI system categorization.	Kiteworks documents and enforces the data access boundaries within which AI systems operate. Attribute-based folder and file permission evaluation combines user, role, parent container, attachment context, and source attributes to produce allow/deny decisions that define which data an AI system is permitted to reach. Policy-based feature entitlements evaluate per-profile feature lists to gate access and establish clear operational context at the data layer. Content tagging and metadata through REST APIs ensure classification labels are carried forward across file versions and used for ABAC lookups, providing an auditable record of the data scope made available to AI systems.
MAP 4.2	Internal risk controls for components of the AI system, including third-party AI technologies, are identified and documented.	Kiteworks comprehensively identifies and documents internal risk controls for AI system components through its risk policy engine within the Data Policy Engine, allowing administrators to define risk policies with rules, directives, approvers, tag/form triggers, and automated actions on content. The platform implements multi-engine antivirus/ATP/ICAP scanning with quarantine capabilities, integrating multiple scan backends including F-Secure, FireEye DOD, and ICAP CSI services with quarantine/unquarantine actions and virus assertion. ICAP-based content scanning integration sends files to third-party ICAP scanners with configurable service URLs, request/response headers, chunked encoding, and result parsing. Enterprise content source connectors through Enterprise Connect integrate backend ECM/storage systems both in the cloud and on-premises with licensed server counts, ensuring comprehensive risk control coverage across all AI system components.

MEASURE Function Outcomes

The MEASURE function requirements included in this guide address risk measurement approaches and documentation of unmeasurable trustworthiness characteristics, regular assessment of metric effectiveness, data-layer evidence supporting validity and reliability, safety controls and governance infrastructure availability, security and resilience evaluation, transparency and accountability risk documentation, data context governance for responsible use, environmental impact documentation of AI data flows, ongoing risk identification and tracking, feedback processes for end-users and affected communities, and measurement approaches informed by domain experts.

Function	Outcome	Kiteworks Solution
MEASURE 1.1	Approaches and metrics for measurement of AI risks enumerated during the MAP function are selected for implementation starting with the most significant AI risks. The risks or trustworthiness characteristics that will not -- or cannot -- be measured are properly documented.	Kiteworks implements comprehensive AI risk measurement approaches through its Compliance Report – Audit Log feature that audits log events corresponding to selected risk policies and reports on those actions, with tag-based data breakdown capabilities requiring an Advanced Governance license. The risk policy reporting dashboard produces background job reports on risk-policy violations, enabling systematic measurement of identified risks. The CISO risk dashboard with geolocation map provides a specialized console that surfaces file scan details, geo-located events, communication flows, IP-to-server mappings, and top user activities. These tools surface data-layer risk signals and policy violations, supporting organizations in their broader effort to measure and document AI risks.
MEASURE 1.2	Appropriateness of AI metrics and effectiveness of existing controls are regularly assessed and updated, including reports of errors and potential impacts on affected communities.	Kiteworks enables regular assessment and updating of AI metrics and control effectiveness through its comprehensive admin activity audit trail that supports list, detail, and CSV export endpoints for forensic review of control implementations. The platform implements event validation and masking for API admin calls, validating event schemas, masking sensitive values, validating client/host/user-agent and IPs, and logging API admin calls to ensure metric accuracy. Threat reports and insider threat dashboards generate comprehensive insider/outsider threat reports covering communications, top actors, and country activity with CSV export capabilities. These tools allow organizations to regularly assess the appropriateness of metrics, identify errors, and evaluate potential impacts on affected communities through detailed reporting and analysis capabilities.

Function	Outcome	Kiteworks Solution
MEASURE 2.5	The AI system to be deployed is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented.	Kiteworks governs what data AI systems can access and documents access controls at the data layer, providing evidence that supports validity and reliability assessments. Compliance Reports from the Data Policy Engine show adherence to regulatory controls governing data handling. Asset classification and risk policy rules enforce content and user conditions, maintaining per-asset file metadata that documents the data boundaries within which AI systems operate. Content tagging with Microsoft Information Protection, Titus, and custom sensitivity labels creates a data-classification record attached to every file, documenting the data context made available to AI systems and the access limitations applied. These controls produce the audit evidence organizations need to demonstrate that the data layer governing AI system inputs is operating within defined and documented boundaries.
MEASURE 2.8	Risks associated with transparency and accountability -- as identified in the MAP function -- are examined and documented.	Kiteworks examines and documents transparency and accountability risks through its persistent audit event log service that captures all events with entity resolution for files, folders, users, mail, attachments, and comments with default single-line descriptions per event type, ensuring complete transparency. The admin activity audit trail supports list, detail, and CSV export endpoints for forensic review, providing documented accountability for all administrative actions. Event transaction tracking uses EventTransaction entities to group related events, with a cleanup task and query helpers for time-range splits supporting eDiscovery-style log navigation. These capabilities comprehensively examine and document risks associated with transparency and accountability as identified in the MAP function, maintaining clear audit trails for all system activities.
MEASURE 2.9	The AI model is explained, validated, and documented, and AI system output is interpreted within its context -- as identified in the MAP function -- to inform responsible use and governance.	Kiteworks governs the data context within which AI systems operate, providing the documented data trail that supports responsible use and governance at the data layer. Every file accessed by an AI agent through Kiteworks Compliant AI is identity-verified, ABAC-evaluated, FIPS 140-3 encrypted, and captured in a tamper-evident audit log -- creating a complete record of what data entered the AI workflow, who authorized it, and under what policy. Content tagging with Microsoft Information Protection and Titus sensitivity labels attaches classification records to every file, documenting the regulatory context of data made available to AI systems. Comments and tasks on content enable teams to document decisions about data access and flag items for review, supporting governance of AI system inputs and outputs at the data layer. These controls produce the authoritative, policy-enforced record of what regulated data entered AI workflows and under what classification -- the foundation for responsible use and governance.

Function	Outcome	Kiteworks Solution
MEASURE 2.12	Environmental impact and sustainability of AI model training and management activities -- as identified in the MAP function -- are assessed and documented.	Kiteworks Compliant AI governs and logs every AI agent interaction with enterprise data at the data layer -- providing a structured, tamper-evident record of what data AI workflows consumed, when, and under whose authorization. This audit trail, streamed in real time to SIEM, gives organizations the data-access documentation needed to support broader environmental impact assessments of AI operations when combined with infrastructure-level monitoring tools. Kiteworks does not directly measure energy consumption, compute utilization, or carbon footprint of AI model training; organizations should use dedicated infrastructure monitoring for those metrics. What Kiteworks provides is the authoritative, policy-enforced record of the regulated data flowing into AI workflows -- a necessary input to any complete AI sustainability assessment.
MEASURE 3.1	Approaches, personnel, and documentation are in place to regularly identify and track existing, unanticipated, and emergent AI risks based on factors such as intended and actual performance in deployed contexts.	Kiteworks implements comprehensive risk tracking through threat reports and insider threat dashboards that generate reports covering communications, top actors, and country activity with CSV export for identifying emergent AI risks. Risk and content-firewall policy management enables administrators to add/update/delete risk policies, link policies, and run content firewall in report-only mode with quarantine triggers based on detected risks. Security Analytics utilizing the CISO Dashboard and Splunk App visualizations reveal suspicious traffic patterns including anomalies in download locations, traffic levels, user activities, and individual file activities across all data exchange channels such as email, file shares, SFTP, MFT, and Microsoft Teams. These approaches provide personnel with documentation to regularly identify and track anomalies and policy violations across data exchange channels at the data layer.
MEASURE 3.2	Risk tracking approaches are considered for settings where AI risks are difficult to assess using currently available measurement techniques or where metrics are not yet available.	Kiteworks addresses difficult-to-assess AI risks through its attribute-based policy engine within the Data Policy Engine that evaluates policies, rules, and actions against operations, user/entity/execution context, and LDAP attributes to produce allow/deny decisions even when traditional metrics are unavailable. Risk policy reporting generates background job reports on risk-policy violations, providing tracking mechanisms for emerging risk patterns. Kiteworks Compliant AI governs AI agent access to enterprise data at the data layer regardless of which model, agent framework, or prompt is in use -- enforcing authenticated identity via OAuth 2.0, ABAC policy evaluation, FIPS 140-3 encryption, and tamper-evident audit logging on every interaction. This data-layer enforcement applies even in novel AI deployment scenarios where model-level metrics do not yet exist, providing a consistent, policy-driven control point independent of AI system maturity. These approaches enable organizations to track AI risks in settings where conventional measurement techniques are insufficient or where standardized metrics have not yet been developed.

Function	Outcome	Kiteworks Solution
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.	Kiteworks establishes comprehensive feedback processes through Secure Data Forms (Advanced Forms) that enable end-users to collect structured feedback data in modern enterprise forms leveraging the platform's centralized security and compliance policies and unified audit log. Advisory contacts management allows administrators to manage security advisory contacts for notifications and alerts, creating channels for problem reporting. Web forms, both standalone and embedded, provide configurable forms with custom fields supporting standalone or embedded hosting, profile assignment, and per-form access control for external intake of feedback. These feedback mechanisms are integrated into the system's evaluation metrics through the audit log and reporting capabilities, ensuring that end-user reports and appeals are systematically captured and incorporated into AI system assessments.
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. Approaches are documented.	Kiteworks connects measurement approaches to deployment contexts through Secure Data Forms where end-users licensed for building Advanced Forms can create, manage, and distribute sophisticated forms within Kiteworks using a simple, self-service user interface for domain expert consultation. Comments and tasks on content allow collaborators to attach comments and be assigned tasks, with task-owner/assignee listing and task share links facilitating expert input on risk measurement approaches. Secure folder collaboration and membership management adds, updates, and removes folder and file members with role-based permissions, invitations, and delegated sharing to engage domain experts. These capabilities ensure measurement approaches for identifying AI risks are informed by consultation with domain experts and end-users, with all approaches documented through the platform's comprehensive audit and reporting mechanisms.

MANAGE Function Outcomes

The MANAGE function requirements included in this guide address procedures for responding to and recovering from previously unknown risks, mechanisms to supersede or deactivate AI systems performing inconsistently with intended use, monitoring of risks and benefits from third-party AI resources, monitoring of pre-trained model artifacts, post-deployment monitoring including incident response and change management, and communication of incidents and errors to relevant AI actors with documented tracking and recovery processes.

Function	Outcome	Kiteworks Solution
MANAGE 2.3	Procedures are followed to respond to and recover from a previously unknown risk when it is identified.	Kiteworks supports data-layer response and recovery when previously unknown risks are identified. The Data Policy Engine can be updated to apply new risk policies immediately, quarantining or blocking access to affected content while an incident is assessed. Database cluster failure recovery enables admin-triggered recovery with status tracking and progress checks, keeping the governance infrastructure operational during incidents. Maintenance mode controls block user and agent traffic during remediation, ensuring no data moves while risk controls are being updated. These mechanisms provide documented, repeatable procedures for responding to emerging risks at the data layer.
MANAGE 3.1	AI risks and benefits from third-party resources are regularly monitored, and risk controls are applied and documented.	Kiteworks monitors third-party AI risks through Enterprise Connect, which enables governed access to data in external ECM repositories, making that data available within the Kiteworks security and policy framework alongside natively stored content. Cloud/ECM storage connectors for KP sources allow administrators to add, test, list, and edit external content sources including ECM and document repositories with per-source authentication tokens, ensuring controlled third-party access. Compliance reporting for CMMC 2.0 demonstrates compliance with CMMC 2.0 compliance controls, requiring an Advanced Governance license to document third-party risk management. These features enable organizations to regularly monitor AI risks and benefits from third-party resources while applying and documenting comprehensive risk controls across all integrated systems.
MANAGE 3.2	Pre-trained models which are used for development are monitored as part of AI system regular monitoring and maintenance.	Kiteworks monitors pre-trained models through its antivirus and ATP scanning with multi-engine orchestration that triggers per-file, bulk, and folder AV/ATP scans, aggregates results across configured scan services, and tracks verbose status to detect potential model corruption or tampering. File versioning with version history maintains numbered versions with version increment, version collections, and per-version location/scan metadata, enabling tracking of pre-trained model iterations and changes. Kiteworks Compliant AI governs all AI agent interactions with enterprise data through authenticated identity verification, ABAC-enforced access controls, and a tamper-evident audit trail streamed to SIEM -- providing a complete, policy-enforced record of every operation performed on or with pre-trained model artifacts stored in the Kiteworks environment. These capabilities ensure pre-trained models are continuously monitored as part of regular AI system maintenance, with all interactions and modifications tracked through the unified audit system.

Function	Outcome	Kiteworks Solution
MANAGE 4.1	Post-deployment AI system monitoring plans are implemented, including mechanisms for capturing and evaluating input from users and other relevant AI actors, appeal and override, decommissioning, incident response, recovery, and change management.	Kiteworks supports post-deployment monitoring at the data layer -- governing, logging, and auditing every interaction between AI agents and enterprise data after systems go live. The MFT Server Operations Console monitors data workflows at scale, tracking job status and surfacing anomalies in automated data flows. The Data Policy Engine enforces access controls continuously and generates incident records when policy violations occur, feeding the tamper-evident audit log streamed to SIEM. Advisory contacts and email notification templates ensure relevant stakeholders receive timely alerts when data-layer incidents are detected. Kiteworks does not manage AI model performance, override AI system decisions, or control decommissioning of AI systems -- its role is to ensure the data those systems access remains governed, encrypted, and fully audited throughout their deployment.
MANAGE 4.3	Incidents and errors are communicated to relevant AI actors, including affected communities. Processes for tracking, responding to, and recovering from incidents and errors are followed and documented.	Kiteworks supports incident communication through email notification templates and advisory contacts management, ensuring relevant stakeholders receive timely alerts when data-layer incidents are detected, with all events captured in the unified audit log. Email notification templates provide reusable templates stored and instantiated with dynamic variables for notifications and direct sends, enabling consistent incident communication. Advisory contacts management allows administrators to manage security advisory contacts for notifications and alerts, ensuring relevant AI actors receive timely incident information. These capabilities support the documentation and communication of data-layer incidents, providing an auditable record organizations can use as part of their broader incident response processes.

The information provided in this guide does not, and is not intended to, constitute legal advice; instead, all information, content, and materials available in this guide are for general informational purposes only. Information in this guide may not constitute the most up-to-date legal or other information. Add-on options are included in this guide and are required to support compliance.