Please refer to the original text for accuracy.

Practical Manual for Establishing a Chief AI Officer and Implementing AI Governance

(Version 1.00)

March 1, 2026

Japan AI Safety Institute (J-AISI)

AISI Japan
AI Safety Institute

Table of Contents

Summary

This manual provides practical guidance primarily for private-sector operators on how to establish a Chief AI Officer (CAIO) and integrally implement the promotion of AI strategy, AI governance, risk management, security and privacy, data governance, talent development, procurement and vendor management, and audit and monitoring. As generative AI and Agentic AI become part of business infrastructure, companies are expected to establish arrangements that balance value creation with the protection of rights, safety, fairness, privacy, and other interests. In response to these needs, this manual presents an integrated architecture for treating AI as a management issue.

The CAIO is positioned not merely as a technology executive, but as the accountable executive who provides company-wide, cross-functional oversight of AI strategy, governance, risk management, talent development, and procurement management. This manual recommends an organizational model in which the CAIO is appointed as an independent C-suite executive reporting directly to the CEO, and an AI Governance Office (also referred to as the AI Enablement Team) is established under the CAIO. It further recommends establishing a "Company-wide AI Steering Committee" comprising the CAIO, Chief Data Officer (CDO), Chief Information Officer (CIO), Chief Technology Officer (CTO), Chief Information Security Officer (CISO), Legal and Compliance, Data Protection Officer (DPO), Human Resources, and key business divisions. Through this structure, the aim is to enable management-integrated decision-making on matters such as whether to adopt high-risk use cases, the allocation of investment, risk tolerance, and responses to major incidents, while minimizing fragmented initiatives and ambiguity of accountability.

While referencing domestic and international frameworks including the AI Guidelines for Business, the NIST AI Risk Management Framework (NIST AI RMF), ISO/IEC 42001 (AI management system: AIMS), ISO/IEC 23894 (AI risk management methodologies), the EU AI Act, and U.S. OMB Memorandum M-24-10, this manual presents a practical operating model combining elements such as an AI inventory, AI Impact Assessment (AIIA), Model card and Datasheet (for datasets), a risk register, Key Performance Indicator (KPI) and Dashboard, and internal audits and external certifications. In particular, for Rights and safety impacting AI—AI that could have a significant impact on human life, physical integrity, liberty, property, or other fundamental rights and safety—the manual sets out additional control expectations, including approval by the CAIO and the Company-wide AI Steering Committee, Human oversight and final decision (Human-in-the-loop: HITL), and enhanced governance such as impact assessment, fairness assessment, and remediation

processes. Note that “Rights and safety impacting AI” in this manual refers to a subset of high-risk AI for which significant impacts on individuals’ rights and interests or on life, physical integrity, or safety are anticipated. This may include domains that are classified as high-risk under the AI Guidelines for Business or the EU AI Act but is not limited to them. In addition, the manual organizes, in a coherent structure, the skills and qualities expected of the CAIO; roles and responsibilities; organizational design and size-based operating models (large enterprises, mid-sized companies, and startups); key points for internal collaboration and engagement with external stakeholders; compliance and regulatory responses; security and privacy; data governance and quality assurance; procurement and vendor management; education and talent; KPI and dashboard design; major risks and mitigation measures; audit, monitoring, and reporting; use-case-specific application examples; key points for templates; and example first-year implementation plans. In doing so, it provides a pathway for organizations to design and introduce a CAIO-centered AI governance structure in phases, considering their size, industry, risk profile, and alignment with existing internal controls, and to continuously improve by referencing this manual as a “Living Document.”

1. Overview

1.1 Background

Generative AI and Agentic AI are becoming increasingly embedded in business operations beyond the conventional scope of AI-driven automation and analytics. They are expanding opportunities for value creation, for example, through conversational agents and personalized recommendations in customer-facing operations; improved efficiency in document drafting and analytical work in back-office functions; and predictive maintenance and optimization in manufacturing and logistics. At the same time, they are amplifying complex and interrelated risks, including legal compliance, security, privacy, fairness, and accountability. Challenges such as transparency in algorithmic decision-making, the prevention of discrimination, the appropriate use of personal data, and the provenance and integrity of generated outputs cannot be adequately addressed through IT governance alone. In practice, trial-and-error efforts within individual departments alone are increasingly insufficient to keep pace with effective management, creating a need for governance and operational frameworks specific to AI.

Both domestically and internationally, frameworks such as the AI Guidelines for Business[1], the NIST AI Risk Management Framework (NIST AI RMF)[2], the EU AI Act[3], and ISO/IEC 42001[4] are being developed. For example, in the United States, governance requirements for federal agencies have been set out premised on the designation of a Chief AI Officer (CAIO). In Japan, national administrative organs are also required to appoint a CAIO under the Digital Agency's "The Guideline for Japanese Governments' Procurements and Utilizations of Generative AI for the sake of Evolution and Innovation of Public Administration"[5]. These developments indicate that not only public institutions, but also private-sector operators are increasingly expected to manage AI in an integrated manner as a management issue, rather than treating it solely as a technical theme.

[1] Ministry of Economy, Trade and Industry: AI Guidelines for Business

[2] NIST: AI Risk Management Framework

[3] The EU Artificial Intelligence Act

[4] ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system

[5] Digital Agency: The Guideline for Japanese Governments' Procurements and Utilizations of Generative AI for the sake of Evolution and Innovation of Public Administration

Meanwhile, in many companies, the adoption of generative AI tools and AI services has progressed under the discretion of individual departments, with investment decisions, quality standards, risk assessments, and vendor selection often conducted in a fragmented manner. As a result, not only may it become difficult to maximize return on investment (ROI) due to duplicated or insufficient investments, but there is also an increased likelihood that risks—such as legal violations, security incidents, the manifestation of discrimination or unfairness, and reputational damage—will be discovered only after the fact. As AI utilization becomes core business infrastructure, governance premised on such decentralized initiatives is no longer fit for purpose.

Against this backdrop, there is a growing need for a dedicated executive lead who can integrate AI strategy, AI governance, risk management, talent development, and procurement and vendor management from a company-wide perspective, and who can continuously fulfill accountability to senior management. Internationally, there is also a growing movement to establish positions such as the CAIO, and this manual puts the CAIO at the center of its approach. The specific rationale for the CAIO role and its mission is described in detail in “3.The Necessity for and Mission of the CAIO.”

Positioning such a dedicated leader at the core and establishing arrangements that balance AI-driven value creation with the protection of rights, safety, fairness, privacy, and other interests is increasingly becoming an essential prerequisite for private-sector operators to utilize AI sustainably and reliably, maintain and strengthen competitiveness and social trust, and respond to domestic and international regulatory trends.

1.2 Purpose of This Manual

The purpose of this manual is to provide standard practical guidance for private-sector operators when establishing and operating a CAIO, and to support organizations in balancing AI-driven value creation (maximizing ROI) and responsible AI use (risk reduction and regulatory compliance). By clarifying the role of the CAIO and presenting, in an integrated manner, elements such as organizational design, processes, evaluation, oversight, education, and procurement, this manual enables each organization to advance AI utilization steadily in a manner aligned with its own context.

In addition, this manual aims to translate domestic and international frameworks—including the AI Guidelines for Business, the NIST AI Risk Management Framework (NIST AI RMF), the EU AI Act, and ISO/IEC 42001—into implementable practices that the CAIO and related functions in private-sector organizations can embed in day-to-day operations.

It further seeks to organize functions such as AI governance, risk management, talent development, procurement, and audit as a coherent architecture. The primary intended users are the CAIO and the CAIO’s operational staff, senior executives considering the appointment of a CAIO, and stakeholders in legal and compliance, the IT function, data functions, human resources, and business divisions.

1.3 Scope

This manual focuses on company-wide operations that cut across management, business, technical, and compliance functions. Specifically, it covers elements such as organizational design, roles and responsibilities, operational processes, templates, Key Performance Indicators (KPI), maturity models, and audit and reporting mechanisms, and covers the full lifecycle, from planning and implementation through operations and vendor selection and procurement. Rather than providing guidance confined to a single function, it aims to establish a foundation for enabling AI to operate safely and effectively across the enterprise as a whole.

While this manual is primarily intended for private-sector operators in Japan, it also contains content that may be informative for organizations operating internationally and for public institutions facing similar challenges. At the same time, it does not provide detailed technical specifications—such as individual algorithm design or model architecture but instead places emphasis on governance, organizational, process, and operational management frameworks surrounding AI.

1.4 Assumptions and Constraints

This manual provides generalized guidance with due regard to differences in company size, industry, and jurisdiction, and does not present a single “one-size-fits-all” answer that can be applied to all organizations without modification. In applying this manual, organizations should tailor the guidance to be effective for their own context, taking into account their risk appetite, regulatory environment, business priorities, and alignment with existing frameworks such as an information security management system (ISMS) and internal control frameworks.

In addition, this manual is not intended as legal advice in any specific jurisdiction. Decisions regarding compliance with applicable laws and regulations should be made under the oversight of an organization’s legal and compliance functions, and, where appropriate, supported by external expert advice. Because the domestic and international guidelines, standards, and laws referenced may change due to future amendments or new enactments, this manual should itself be maintained as a “Living Document”, premised on regular review by the CAIO and relevant functions.

2. Definitions

Term	Definition
CAIO (Chief AI Officer)	A senior executive with overall accountability for integrally leading AI-related strategy, AI governance, risk management, talent development, and procurement. Positioned at a level such as reporting directly to the CEO, the role is responsible for cross-business decision-making and coordination.
AI governance	A framework that includes the establishment and implementation of AI principles (fairness, safety, transparency, accountability, etc.) and policies, operational oversight, and audit—representing the overall set of rules and mechanisms that enable an organization to leverage AI in a trustworthy manner. The definition in the AI Guidelines for Business is as follows, and this manual follows that definition: "The design and operation of technological, organizational, and social systems by stakeholders for the purpose of managing risks posed by the use of AI at levels acceptable to stakeholders and maximizing their positive impact (benefit)."
AIIA (AI Impact Assessment)	A structured procedure for systematically assessing the impacts of AI on rights and safety, fairness, privacy, and other interests, and an assessment process that, prior to the introduction or modification of AI, clarifies potential impacts and mitigation measures.
PIA (Privacy Impact Assessment)	An assessment process that, in advance, identifies privacy risks and impacts associated with the collection, use, and sharing of personal data, and clarifies mitigation measures and the allocation of responsibilities.
Model card / Datasheet	Transparency documentation that standardizes and records, for AI models and datasets, information such as

	design intent, training data, evaluation results, limitations, and intended scope of use.
AI inventory	A company-wide inventory that provides a consolidated view of AI use cases, models, data, vendors and partners used or developed across the organization, and that should serve as the foundational information for management and oversight.
Company-wide AI Steering Committee	A cross-company decision-making body, chaired by the CAIO, that sets high-level direction for AI strategy and AI governance.
Rights and safety impacting AI	AI used in use cases that could have a significant impact on human life, physical integrity, liberty, property, or other fundamental rights and safety.

3. The Necessity for and Mission of the CAIO

3.1 Necessity

Generative AI and Agentic AI are becoming core infrastructure for creating business value, while also involving complex risks from the perspective of legal compliance, security, privacy, fairness, and accountability. In addition, domestic and international guidelines and international standards increasingly call for unified policies and AI governance structures. However, there are limits to relying solely on existing roles—such as the Chief Information Officer (CIO), Chief Technology Officer (CTO), Chief Data Officer (CDO), and Chief Information Security Officer (CISO)—to integrate these requirements at the operational level and embed them into company-wide decision-making and day-to-day operations.

In many companies, AI-related functions are siloed and fragmented—for example, AI strategy is handled by corporate planning, technology strategy by the CTO, data governance by the CDO, security by the CISO, and legal and ethical matters by legal and compliance. As a result, structural challenges are likely to arise, including: (1) investment decisions and use-case selection remain optimized within individual departments, making portfolio management and maximizing ROI across the organization difficult; (2) standards for governance and risk management vary by department, and accountability becomes unclear; and (3) there is no single point of contact that can provide unified accountability to senior management, the board of directors, and external stakeholders.

To address these challenges, an accountable executive is essential—one that can take a company-wide view from both value creation and risk management perspectives and integrate functions across strategy, governance, risk, talent, and procurement. This role is fulfilled by the CAIO. Operating at the highest level of management, such as reporting directly to the CEO, the CAIO is expected to lead cross-functionally on the standardization of AI governance, the design of risk-based operational processes, continuous monitoring through KPI and dashboard, and alignment with external regulations and international standards.

In summary, the main reasons why establishing a CAIO is considered necessary are as follows:

- To connect management strategy with the use of AI, and to optimize AI investments and use cases as a company-wide portfolio.
- To standardize AI principles, policies, approval flows, and audit across the organization, and to clarify “who is accountable for what”.

- To continuously operate risk management related to rights and safety, fairness, privacy, and other interests across the full lifecycle, and to respond swiftly to changes in the internal environment, including regulatory changes.
- To design talent development, cultural adoption, vendor management, external certification, and related elements as an integrated whole, thereby balancing AI-driven value creation with social trust.

3.2 Mission

The mission of the CAIO is to maximize business value creation through AI while protecting rights and safety, fairness, privacy, security, and other interests, so that the organization can leverage AI in a manner that earns societal trust. To fulfill this mission, the CAIO designs an integrated architecture that connects the functions of strategy, governance, risk, talent, and procurement, and embeds it into company-wide decision-making and day-to-day operations.

This mission can be organized into the following four objectives.

1) Alignment With Management Strategy and Value Creation

Align the roadmap and investment portfolio for leveraging AI with management strategy and allocate limited resources to the highest-value use cases. With a view to both short-term efficiency gains and medium- to long-term competitive advantage, continuously visualize and improve the ROI of AI investments.

2) Establishing Trustworthy AI Governance

Based on AI principles and applicable laws and guidelines, design and update policies, approval flows, and oversight and audit mechanisms, and keep risks related to rights and safety, fairness, privacy, and security within acceptable levels. Ensure consistent controls and transparency across the full lifecycle of AI systems.

3) Building Organizational Capability, Talent, and Culture

Define the necessary roles and skill standards and internalize organizational capabilities to appropriately leverage AI through training, reskilling, and talent allocation. Foster a culture that balances experimentation in frontline operations with governance, and scale insights and best practices across the organization.

4) Fulfilling Accountability to Stakeholders

Serve as a single point of contact capable of providing centralized accountability for AI strategy, risks, and response status to stakeholders such as senior management and the board of directors, employees, customers, and regulatory authorities. Through transparency reports, KPI, dashboards, and other means, continuously share the status of AI utilization and improvement processes, and build trust.

4. Skills and Competencies Required for a CAIO

A broad range of skills and competencies is required of the CAIO in order to connect AI-driven value creation to business outcomes, manage and control risks, and strengthen execution capability across the organization. The key skills and competencies expected of the CAIO are organized below. This section does not assume that a single CAIO will possess all of the elements listed here at a high level. Rather, it assumes that the CAIO will build on their own strengths as a core and complement them through role allocation with a team reporting directly to the CAIO described later.

4.1 Technical Understanding and Application Capability

- Has a foundational understanding of machine learning, natural language processing, generative AI (large language models, retrieval-augmented generation, etc.), and MLOps, and can organize key considerations (data, evaluation, operations) when applying them to the organization’s use cases.
- Understands the characteristics and risks of generative AI (hallucinations, bias, lack of transparency, etc.) and can clearly communicate internal points of attention for business use.
- Incorporates perspectives such as cybersecurity, data governance, and privacy protection, and can translate them into requirements for AI systems (data handling, access controls, audit, logs, vendor management, etc.).

4.2 Strategic Thinking and Insight Into Business and Regulatory Environments

- Connects management objectives and business challenges with the characteristics and advantages of AI and can develop an AI adoption roadmap from a company-wide optimization perspective.
- Taking into account marketability, feasibility, and risks, can design and drive performance measurement (KPI/ROI, etc.) and continuous improvement cycles.
- Incorporates external knowledge from government–industry collaboration, academia and research institutions, startups, and other sources, and improves the quality of decision-making for both business and governance.

4.3 Capability to Drive Organizational Transformation and Leadership

- Can lead cross-functional projects and drive them forward while coordinating interests among stakeholders.
- Can implement AI not merely as the introduction of tools, but as a transformation of organizational culture that includes business process redesign, talent allocation, and evaluation systems and related policies.
- Can systematically promote AI literacy improvement, training, and reskilling for executives and employees, and build mechanisms for talent development.
- Maintains a stance of continuously learning about the latest AI technologies and societal trends and agilely reflecting them in measures and rules.

4.4 Ethics, Legal, and Risk Management Capability

- Has an advanced understanding of AI ethics and legal compliance and can operationalize them as internal rules and review and approval processes.
- Can identify and assess potential risks such as false or incorrect outputs from hallucinations and bias and design an end-to-end risk management process that covers establishing and executing mitigation measures and continuous monitoring.
- When making decisions on leveraging high-risk AI, can coordinate with organizations and committees responsible for risk management, such as an AI ethics committee or a risk committee, and escalate appropriately.
- Maintains an awareness of accountability and can sustain governance from the perspectives of fairness, transparency, and societal trust.

4.5 Adaptability and a Forward-looking Mindset

- Continuously tracks trends in international standards and state-of-the-art technologies and maintains a global perspective to adapt them to the organization’s environment.
- In the fast-changing AI domain, can flexibly shift direction with agile thinking and make rapid decisions.

5. Roles and Responsibilities

While the CAIO’s roles and responsibilities are wide-ranging, it is important to clarify accountability and cooperation arrangements among relevant stakeholders to ensure that they function in practice. The CAIO has final accountability for the design and oversight of company-wide AI governance, while final business-specific decisions remain with business leaders; where significant risks exist, the CAIO escalates matters to the board of directors. In addition, for use cases that qualify as Rights and safety impacting AI, the default principle is that, where they are to be deployed to production, separate approval should be obtained from the Company-wide AI Steering Committee and it is desirable for the CAIO to have veto authority, including the ability to halt or prohibit deployment to production.
As chair of the Company-wide AI Steering Committee, the CAIO leads alignment across stakeholders and company-wide optimization in decision-making.

5.1 Management and Strategy

The CAIO develops the company-wide AI strategy, sets the roadmap and investment allocation, and leads the management of ROI. This includes not only setting goals, but also establishing criteria for selecting specific use cases, policies for technology selection, platform strategy, and policies for relationships with partners. Working with the CEO, business leaders, and the finance function, the CAIO integrates expectations for business value, risk tolerance, and resource constraints to make investment decisions. Risk tolerance should be aligned with the enterprise risk management (ERM) framework and reviewed regularly in coordination with the CFO and the risk management function. The CAIO is expected to balance short-term outcomes with long-term capability-building, and to minimize duplicated investments and opportunity losses, including through the use of the AI inventory. By defining KPI for business value and trustworthiness and visualizing them in dashboards, the CAIO establishes a foundation for dialogue with senior management and the board of directors.

5.2 Governance and Implementation of AI Principles

The CAIO establishes policies for implementing AI principles, positions the development and maintenance of foundational documentation—such as AIIA, the AI inventory, model cards, and datasheets—as core elements of governance, and builds approval flows and audit plans. Approval flows should incorporate joint decision-making by the CAIO, business leaders, the CISO, and the legal and compliance functions, balancing accountability and assurance. In addition, in coordination with the legal and compliance functions and the CDO, the CAIO translates AI principles into concrete operational rules and documents criteria for exception and mitigation decisions. These policies and processes should be operated in an integrated manner with management systems such as ISO/IEC 42001, and reviewed regularly (for example, every six months) in line with the PDCA cycle and changes in regulation and technology.

5.3 Risk Management

AI risks span a wide range, including adverse impacts from misclassification, discrimination driven by bias, misinformation, security breaches, privacy violations, and non-compliance with laws and regulations. With reference to frameworks such as ISO/IEC 23894[6], it is recommended to carry out, in a systematic manner, identification (threats, vulnerabilities, impacts, and mapping of relevant stakeholders), analysis (qualitative and quantitative assessment, scenario analysis, and measurement of whether discriminatory impacts arise from fairness and bias), evaluation (comparison against acceptable levels and prioritization), and treatment (avoidance, mitigation, acceptance, and transfer). For each use case, a responsible owner (a business owner or a model owner) should be clearly designated for risk assessment. The CAIO is responsible for standardizing assessment design and providing final approval. For use cases that could have a significant impact on rights and safety, apply the additional controls specified in section “5.8 Additional Internal Controls for Rights and Safety Impacting AI”, and conduct more rigorous assessment, approval, and monitoring.

Risk management should be operated through the following steps.

1) Establish risk treatment policies, roles and responsibilities, and oversight and audit plans, and document organizational risk tolerance.
2) Identify use cases for leveraging AI, and understand the stakeholders involved and potential impacts on rights and safety.
3) Design evaluation and test plans and metrics related to performance, robustness, fairness, explainability, security, and privacy.
4) Implement risk mitigation measures, operate controls, perform monitoring, and, when an incident occurs, take response actions and corrective measures, thereby driving continuous improvement.

[6] ISO/IEC 23894: 2023 Information technology — Artificial intelligence — Guidance on risk management

5) Work collaboratively with the CISO, the legal function, and the risk management function to maintain effective controls grounded in operational practice. Integrate the AI risk register with the company-wide risk register and align it with other financial and operational risks.

5.4 Security and Privacy

In coordination with the CISO, the Data Protection Officer (DPO), and the legal function, define an overall approach and control requirements so that AI-specific threats and privacy risks (prompt injection, model extraction, data poisoning, etc.) are managed in a manner aligned with existing information security management practices.
As a general principle, evaluate and oversee the integration of security and privacy from the design stage (secure-by-design, privacy-by-design, etc.). For specific design and operational methods, see “10. Security and Privacy”.

5.5 Data Governance and Quality Assurance

Oversight of data governance and quality assurance for data used in AI, as well as the legality of licenses, is an important responsibility of the CAIO. In coordination with the CDO and the legal function, standardize the data lifecycle, set quality metrics, and ensure rigorous confirmation of licenses for third-party materials. A perspective is required that balances improved model performance with the prevention of rights infringement. Because data quality and license legality underpin not only model performance but also fairness, explainability, and transparency, they should be managed consistently through documentation such as datasheets. For details, see “11. Data Governance and Quality Assurance”.

5.6 Procurement and Vendor Management

Procurement and vendor management should be operated across three layers: policy, evaluation, and monitoring. In coordination with the procurement function, the legal function, the CISO, the DPO, and other relevant stakeholders, the CAIO designs evaluation criteria and a vendor risk assessment process, including vendor lock-in, and is responsible for the final decision on whether to adopt high-risk AI. For detailed evaluation perspectives, contract clauses, and post-contract monitoring requirements, see “12. Procurement and Vendor Management”. For related checklists, see “18. Template Key Points”.

5.7 External Conformity Assessment and Certification

Develop and update, as an “external certification roadmap”, the certification strategy (target schemes, scope, selection of certification bodies, and timing for obtaining certification), alignment of mutual recognition and jurisdiction-specific requirements, and an audit readiness plan. Review domestic accreditation trends (e.g., accreditation of certification bodies for AI management system certification by ISMS-AC, JAB, and others) and overseas accreditation trends (e.g., ANAB, IAS, RvA, and others) every quarter and reflect them in procurement and market entry plans. Where appropriate, pursue integrated audits with existing certifications such as ISMS to minimize duplicated controls and the audit burden.

5.8 Additional Internal Controls for Rights and Safety Impacting AI

For use cases that use Rights and safety impacting AI, the default principle is that approval should be obtained from the CAIO and the Company-wide AI Steering Committee, and human oversight and final decision (Human-in-the-loop: HITL) within workflows should be mandated. Standardize and regularly review impact assessment, fairness assessment, discrimination mitigation, child sexual abuse material (CSAM) measures, user notification, and the design of human-led remediation pathways. In cooperation with the legal function, business functions, and customer support, establish specific remediation procedures and points of contact for inquiries, and set KPI for complaint handling and remediation and recurrence prevention.

5.9 Innovation and PoC Management

Clarify the approach from identifying use cases through PoC (Proof of Concept), criteria for transition to production, and exit criteria. Define operating rules for a testing environment that enables rapid trials within the boundaries of the rules and make phased adoption—such as shadow operation and limited releases—the standard approach. In collaboration with business functions, the CTO and data science teams conduct value hypothesis testing and risk assessment in parallel, and balance speed and safety.

5.10 Training and Talent

Define roles related to leveraging AI, set skill standards for those roles, institutionalize training programs, and embed AI use as a sustainable organizational capability through reflection in evaluation and compensation. In coordination with the human resources function and business functions, establish career paths and provide continuous learning opportunities. In addition, continuously provide AI literacy training for the board of directors and senior management, as well as briefings on risk and governance, to support management-level understanding and the fulfillment of accountability. For details, see “13. Training and Talent”.

6. Governance Process Workflow

To make the introduction and operation of AI safe and repeatable, it is essential to establish “gates” at each stage of the lifecycle and clearly define the required assessments, documentation, and approvals. This approach aligns with documentation and evidence requirements under the AI management system (AIMS) in ISO/IEC 42001 and with the ”documentation and evidence” expectations in the NIST AI Risk Management Framework (NIST AI RMF).

1) At the idea stage, develop a business value hypothesis and conduct an initial assessment of scope of impact, then begin preparing the AIIA.
2) Before moving to a PoC, conduct the AIIA to specify impacts on rights and safety, fairness, and privacy, and define necessary design measures (stakeholder notification, human review, remediation procedures, use restrictions, etc.).
3) During the PoC, execute evaluation metrics and test plans for performance, fairness, robustness, security, and privacy, and record the results in the model card and datasheet.
4) At the go-live approval gate, the business owner, the CAIO, the CISO, and the legal and compliance functions make a joint decision and confirm that evaluations and measures meet requirements.
5) During operations, monitor performance, fairness, incidents, and complaints through dashboards, and connect anomaly detection to remediation and re-evaluation.
6) Through prescribed reviews and periodic (e.g., quarterly) internal audits, verify policy compliance, completeness of records, and the effectiveness of remediation. A key principle here is a documentation-centered approach: by maintaining the AIIA, model card, datasheet, approval records, and audit reports, and preserving the basis for decisions, the organization ensures transparency of operations and supports explainability.

This workflow corresponds to the following chapters: “7. Organization Design and Recommended Structure”, “14. KPI, Measurement, and Dashboard”, “15. Key Risks and Mitigation Measures”, “16. Audit, Monitoring, and Reporting” and “17. Use Case–Specific Examples of Application”. This section presents the overall structure.

7. Organization Design and Recommended Structure

In designing an AI governance structure, this chapter is premised on four principles: (1) an independent oversight function, (2) second line of defense, (3) proportionality based on company size and risk, and (4) phased expansion based on maturity.

7.1 Recommended Structure

The recommended structure is to establish the CAIO as an independent member of the C-suite, reporting directly to the CEO. Ensuring the CAIO's independence enables neutral cross-business decision-making, prioritization, and oversight. In addition, establish the AI Governance Office (also referred to as the AI Enablement Team) under the CAIO, with core functions including AI strategy development, standardization, operational oversight, and self-assessment. Further, establish the Company-wide AI Steering Committee and, through a decision-making body comprising the CAIO, the CDO, the CIO, the CTO, the CISO, the legal and compliance functions, the DPO, the human resources function, and key business divisions, make timely decisions closely linked to management on whether to adopt use cases, policy revisions, business prioritization and investment allocation, and responses to significant incidents.

As a standard, the Company-wide AI Steering Committee should be held at least once a month. From an operational perspective, it is desirable to set standard agenda items: (1) review of new use cases, (2) review of key KPIs, and (3) reporting of incidents and complaints and remediation plans.

The Company-wide AI Steering Committee is a decision-making body chaired by the CAIO. In principle, it is expected to have final decision-making authority, or to serve an advisory function to the executive management committee, with respect to the items below. Which positioning is adopted should be selected based on each organization's governance structure and clearly specified in internal rules and related documents.

- Use case approval (approval based on the AIIA and evaluation results).
- Resource allocation (people, budget, and platforms).
- Response policy for significant incidents (company-wide decision-making).

In addition, the AI Governance Office should be assigned functions such as the following.

- Development and revision of AI policies and guidelines (through periodic updates and in response to changes in regulation and risks).
- Development of and operational support for templates such as the AIIA, model cards, and datasheets.

- Secretariat for use case reviews (inventory management, agenda preparation, and records management).
- Consolidation and reporting of KPIs and dashboards.
- Point of contact for coordination with self-assessment and internal audits.

Through this structure, organizations can avoid fragmented initiatives and enable decision-making that centrally considers both value and risk. The key advantages of this structure are as follows.

- The organization can make rapid decisions on priorities and investment allocation with direct linkage to management.
- Standardization of governance and exception management can be applied consistently (reducing variation across functions).
- The organization can coordinate interests among security, legal, and privacy functions and business divisions from an independent perspective.

7.2 Alternative Structural Options

Where AI use cases are limited—for example, where no high-risk AI exists—or where regulatory risk is moderate or lower, an alternative to the recommended structure is to have the CDO concurrently perform CAIO functions. The advantage of this approach is close integration with data strategy, strengthening linkages with data quality and governance for AI use. On the other hand, there are concerns that the certainty of independent oversight may be reduced and that cross-functional coordination with legal and security functions may become less agile. Priorities related to impacts on rights and safety may be overshadowed by data-related issues, potentially biasing weighting in decision-making. Accordingly, selecting this structure requires careful consideration, taking into account the organization's risk appetite and organizational culture.

In regulated industries, expectations for independent oversight are high, and it is desirable to separate oversight functions from functions that promote leveraging data and AI. Therefore, this alternative option is, as a general rule, not recommended.

7.3 Coordination With Each Function

It is important not for the CAIO to manage everything directly, but to clearly define role allocation with the CxO leading each function and to vary the intensity of coordination accordingly. However, the appropriate level of coordination with internal functions differs depending on the function under consideration.

- S**trategy development:** coordination is strongest with business functions, the CDO, the CIO, and the CTO, integrating perspectives on business value, technical feasibility, and data resources.
- **Governance and risk management:** coordination is strongest with CISO and the legal and compliance functions, embedding perspectives on security, privacy, and regulatory compliance into operations.
- **Innovation:** coordinate closely with business functions, data science, the CIO, and the CTO, and share criteria for experimentation and transition to production.
- **Talent and culture:** human resources, business functions, and data science take the lead in advancing role definitions, skill standards, and training.

These coordination arrangements are not fixed. As an operational practice, it is practical to create and update a RACI (Responsible/Accountable/Consulted/Informed) every quarter by use case, and to review, as needed, the allocation of responsibilities and the strength of approval flows.

Example RACI: for approval of a high-risk AI use case, designate the business owner as Responsible, the CAIO as Accountable, the CISO, legal, and the DPO as Consulted, and the IT function as Informed.

7.4 Structural Option for Mid-Sized Companies

For the purposes of this manual, a mid-sized company refers to an organization with up to several hundred employees, and with fewer than several dozen AI use cases per year.

Place the CAIO reporting directly to the CEO and centralize cross-business approval and oversight. Where a dedicated CAIO is difficult, the CTO or the head of business planning may concurrently perform the CAIO function, and an independent safety lead (information management officer) with stop authority should be appointed to secure the second line of defense. Where it is difficult to establish a dedicated AI governance office, operate a virtual structure of one to three people to manage the AIIA, the use case inventory, and model registration and monitoring.

Simplify governance bodies to a monthly review (AI lead, safety lead, and business owner) and quarterly CEO approvals.

Legal and privacy functions should coordinate with external advisors, and reviews should always be conducted when handling personal data. The IT lead should be responsible for selecting foundational infrastructure and tools, and when using vendors, security, data processing agreements, and checks on cross-border data should be mandatory.

Where the relevant responsible functions do not exist, consider the following alternative functions or organizations.

- **If there is no IT lead:** an external MSP (Managed Service Provider) or the security team of a cloud vendor.
- **If there is no legal function:** outside counsel or an external DPO service.
- **If there is no safety lead:** an information security officer or a personal information protection manager.

7.5 Structural Option for Small Startups

For the purposes of this manual, a small startup refers to a startup with up to several dozen employees.

Simplify decision-making while ensuring a minimal second line of defense. The CAIO function is performed concurrently by the CEO or the CTO, who assumes final accountability for go/no-go decisions for release. Designate a non-product member as the safety owner and grant release-stop authority. Coordinate with external legal and security experts where necessary.

Include a fixed AI governance agenda in regular product meetings held weekly or at another suitable cadence and confirm material issues. As fixed agenda items, it is desirable to review: (1) planned releases of new AI features and the status of the AIIA, (2) recent incidents and complaints, and (3) changes in regulations and platform policies.

Before shipment, prepare checklists covering items such as the AIIA, model cards, a data use register, and vendor confirmations, and record and approve risks and mitigation measures.

When an incident occurs, based on defined severity levels, promptly stop and notify as appropriate, and conduct a post-incident review.

After beginning governance under a startup structure, it is desirable to consider transitioning to the mid-sized-company structure described above when, for example, the organization begins production operation of high-risk AI, begins to process large volumes of personal data on an ongoing basis, or when AI-related revenue exceeds a certain threshold (e.g., more than 20% of total revenue).

8. Collaboration Structure and Stakeholder Engagement

8.1 Internal Collaboration

Because AI operations cannot be completed within a single function, key internal executives should work closely with the CAIO while fulfilling their respective roles. The table below reorganizes, from the CAIO's perspective, the coordination with each function presented in "7.3 Coordination With Each Function" and sets out concrete coordination points. The CAIO and each executive should collaborate on these coordination points through the Company-wide AI Steering Committee (monthly to quarterly) and through individual working groups, so that their respective activities can be mutually optimized. In particular, for critical use cases related to high-risk AI, organizations should follow a common process: AIIA review → steering committee approval → go-live review.

Role example	Coordination Points
Chief Technology Officer (CTO)	◦ Discuss the selection of AI technologies and implementation approaches and share the technology roadmap. ◦ Consult on how AI will be leveraged in products or production lines. ◦ Integrate AI systems into existing infrastructure from the perspectives of infrastructure, architecture, and operations. ◦ Coordinate on establishing technical standards, developing technical talent, and allocating AI-related R&D budgets.
Chief Data Officer (CDO)	◦ Jointly drive data collection, quality management, and governance required for AI projects. ◦ Cooperate on establishing and enforcing rules for ethical data use and privacy protection in data use. ◦ Continuously discuss alignment between data strategy and AI strategy, and update priorities.
Chief Information Officer (CIO)	◦ Collaborate on company-wide optimization of IT systems associated with AI adoption, cost management, IT governance, and establishing IT standards.

	◦ Share plans for designing integration between AI-related systems and core business operations, as well as the overall plan for advancing digital transformation (DX). ◦ Identify IT-related threats and challenges that increase with AI use and consider mitigation measures.
Chief Information Security Officer (CISO)	◦ Jointly develop security enhancements enabled by AI and risk response measures. ◦ Cooperate on measures for AI data protection and the prevention of model misuse (unauthorized use, data leakage, etc.). ◦ Exchange views regularly on updates to security standards and response measures.
Chief Supply Chain Officer (CSCO)	◦ Collaborate on developing strategy for AI deployment across the supply chain (forecasting, optimization, risk management, etc.). ◦ Discuss supply chain data use and AI and data collaboration with external partners. ◦ Drive training and change management to support the adoption of AI tools in frontline operations.
Chief Human Resource Officer (CHRO)	◦ Jointly plan and execute initiatives for recruiting and developing AI talent, reskilling, and improving AI literacy. ◦ Work together on work process transformation, organizational design, job role shifts and change management driven by AI deployment. ◦ Exchange views on the impacts of AI deployment on employees, including from the perspectives of ethics, transparency, and fairness.
Chief Operating Officer (COO)	◦ Collaborate on deployment planning for leveraging AI in work process transformation and on setting KPIs. ◦ Drive business process redesign (to-be design) and standardization and agree on operational design to embed use of AI into day-to-day operations (role allocation, procedures, and delineation of responsibilities).

	◦ Establish cycles for post-go-live quality and stable operations (service-level agreements: SLA /service-level indicators: SLI, monitoring and logs, performance measurement) and continuous improvement, and coordinate on escalation and remediation operations when deviations or defects occur.
Chief Executive Officer (CEO)	◦ Appoint the CAIO and clarify the CAIO's mission, authorities, and resources (staffing and budget). ◦ Consult with the CAIO on alignment between the company-wide AI strategy, management strategy, and risk appetite and make the final decision. ◦ Receive regular reports from the CAIO and the Company-wide AI Steering Committee and make decisions on matters involving high-risk AI and on significant incidents. ◦ Chair discussions and lead decision-making at the board of directors and executive management committee level on reviews of the AI governance structure and policies.
Chief Financial Officer (CFO)	◦ Collaborate on budget allocation and prioritization for AI investments (CAPEX/OPEX) from a company-wide portfolio perspective. ◦ Jointly design and operate a balanced evaluation of AI investment ROI/KPI, costs, and risks (value-based management). ◦ Coordinate on designing and reviewing investment decision rules—taking into account financial constraints and risk appetite—including approval gates and evaluation timing.
Business Owner	◦ Collaborate on identifying and selecting critical use cases, developing value hypotheses, and setting KPIs. ◦ Agree on approaches to PoC and evaluation and on scaling decisions (go-live/exit), together with risk assessment. ◦ Organize business and operational requirements (adoption in frontline operations, quality, and delineation of responsibilities) and lead coordination with development, operations, and governance functions.

Legal/Compliance	◦ Collaborate on establishing and updating internal rules (rules for use, review and approval processes, etc.) based on interpretations of AI-related laws and guidelines. ◦ Jointly confirm and remediate matters related to contracts and procurement (vendor management, rights clearance, auditability, clauses related to accountability, etc.). ◦ Coordinate on escalation, regulatory engagement, internal investigations, and corrective measures when matters involve high-risk AI or when significant incidents occur.
Data Protection Officer (DPO)	◦ Oversee data protection design, including the Privacy Impact Assessment (PIA), and coordinate on implementing privacy-by-design. ◦ Organize requirements for handling personal data (purpose, minimization of collection and use, retention periods, data subject requests, etc.) and reflect them in requirements definition and reviews for AI projects and AI use cases. ◦ Coordinate on data protection risk assessment and remediation, including vendor management and cross-border transfers, and on incident response when significant incidents occur.

8.2 Stakeholder Engagement

Engagement with external stakeholders is also important. The following sections summarize engagement with various stakeholder groups.

Stakeholder	Engagement
Customers and Users	◦ Collect expectations and complaints (customer support, user research, net promoter score: NPS, etc.). ◦ Establish notification and remediation processes (complaint channels, appeals, re-review).
Affected communities and individuals	◦ Engage through dialogue with representative groups, advisory boards, and focus groups.

	◦ For high-risk use cases, incorporate consultation as part of the AIIA process.
Regulatory Authorities	◦ Conduct consultations and pre-consultations, incorporate guidance, and maintain incident reporting channels for significant incidents.
Partners and Vendors	◦ Conduct reviews of technology, contracts, and transparency (sharing model cards and datasheets and exercising audit rights).
External Auditors and Certification Bodies	◦ Confirm the status of corrective actions through audits such as ISO/IEC 42001 audits and reflect findings in improvement plans.

Incorporate feedback received into the AI inventory, risk register, and KPI dashboards, and review it together with improvement measures in the Company-wide AI Steering Committee. For high-risk AI, engage stakeholders more frequently than usual (e.g., regular dialogue with affected communities and participation in impact assessments).

In addition, publish transparency reports (intended use, performance/limitations, fairness assessment results, update history, etc.) and provide information through FAQs and a portal site to establish a foundation for dialogue with stakeholders.

Coordinate with public relations, investor relations, legal, customer support, and other relevant functions to align stakeholder engagement policies and messaging. Individual inquiries and complaint handling should be handled by customer support and designated contact functions, with AI governance-related issues escalated to the CAIO.

9. Compliance and Regulatory Response

9.1 AI Guidelines for Business

The AI Guidelines for Business provide a framework for translating principles such as human-centeredness, safety, fairness, privacy protection, security, transparency, and accountability into concrete business processes and operational rules. This manual as a whole also uses the AI Guidelines for Business as a primary reference framework.

The CAIO and the AI Governance Office are responsible for mapping each principle in the AI Guidelines for Business to the organization’s AI policies and documentation formats, including AIIA templates and templates such as model cards and datasheets, and embedding them into day-to-day operations. Specifically, the following efforts are expected.

- Map each principle of the AI Guidelines for Business to the organization’s “AI principles” and “code of conduct”, and document those relationships.
- Incorporate items related to safety, fairness, privacy protection, security, and accountability into templates such as AIIA, model cards, and datasheets, so that the status of implementation of these principles can be visualized for each use case.
- For Rights and safety impacting AI, reflect relevant parts of the AI Guidelines for Business in the AIIA as mandatory check items and confirm them during Company-wide AI Steering Committee review.

Because the AI Guidelines for Business themselves are a “Living Document” that may be revised and supplemented over time, it is desirable for the CAIO, in coordination with the legal and compliance functions, to conduct a gap analysis between the guidelines and the organization’s policies and operations at least once a year as a general rule. The results should be reported to the Company-wide AI Steering Committee, which will decide on any necessary remediation plans and revision policies.

9.2 EU AI Act

The EU AI Act entered into force on August 1, 2024, and will begin full application on August 2, 2026. Its objectives are to promote the uptake of human-centered, trustworthy artificial intelligence, ensure a high level of protection of fundamental rights against harmful impacts of AI systems—including health, safety, democracy, the rule of law, and environmental protection—and support innovation.

The EU AI Act adopts a risk-based approach, establishes four risk levels, and sets requirements and regulatory obligations according to the relevant risk level. It also sets out rules for general-purpose AI models that can learn and perform a broad range of tasks and can be integrated into other AI systems.

- **Unacceptable risk**
  - Examples: subliminal techniques, social scoring, emotion inference systems in workplaces or educational institutions, real-time remote biometric identification systems in public spaces for law enforcement purposes, etc.
  - Prohibited in principle.
- **High risk**
  - Examples: machinery, medical devices, biometric identification, critical infrastructure, education, employment, law enforcement, migration management, etc.
  - Strict regulatory requirements for providers, importers, distributors, and deployers, including risk management, data governance, preparation of technical documentation, human oversight measures, conformity assessment procedures, log retention, and other obligations.
- **Limited risk**
  - Examples: generative AI, AI systems that interact with natural people, emotion recognition systems, etc.
  - Limited transparency obligations, such as marking AI-generated content and disclosing the use of AI.
- **Minimal risk**
  - Examples: AI systems other than those above.
  - Generally permitted (with voluntary codes of conduct recommended).

In addition to the above, with respect to general-purpose AI (GPAI), cross-cutting obligations are imposed particularly on providers of GPAI models. Where a GPAI model qualifies as having systemic risk under certain criteria, it becomes subject to additional obligations.

The EU AI Act also defines requirements and responsibilities by actor type, using the terms ”provider”, “deployer”, “importer”, “distributor”, and “authorised representative”, among others.

To ensure alignment with the EU AI Act for AI systems provided or used in the EU, the CAIO and the AI Governance Office should undertake at least the following efforts:

- Create and periodically update an AI inventory that maps the organization’s AI use cases to the EU AI Act risk categories.
- For each use case, determine whether the organization qualifies as a “provider”, “deployer”, “importer”, or “distributor”, and organize the obligations required for each role (technical documentation, logs, quality management, monitoring and reporting, etc.).
- For systems that may fall under high-risk use cases or GPAI, establish internal processes aligned with frameworks such as ISO/IEC 42001, ISO/IEC 23894, and the NIST AI Risk Management Framework (NIST AI RMF), and enable requirements to be operated within the management system.
- Systematically store and manage technical documentation, records of training data, logs, AIIA, model cards, datasheets, and related materials in a manner that satisfies the EU AI Act’s evidence requirements.

Monitor the implementation status of the EU AI Act and developments in secondary legislation (implementing acts and delegated acts) in coordination with the legal and compliance functions, and revise policies and operations as necessary.

In applying the EU AI Act, relationships with other EU laws cannot be disregarded, including the General Data Protection Regulation (GDPR) (personal data protection), product safety-related legislation, and online platform regulation. The CAIO should treat these as part of an integrated compliance design and, in collaboration with the legal and compliance functions, establish arrangements so that integrated judgments can be made on a case-by-case basis.

9.3 Other Jurisdictions and Related Laws

Leveraging AI requires alignment not only with rules specific to AI, but also with a wide range of existing laws and regulations. Typical areas of consideration include the following:

- **Privacy and data protection:** the Act on the Protection of Personal Information (APPI), GDPR, etc.
- **Consumer protection:** misleading representations, false or exaggerated advertising, duties of explanation, fairness of contract terms, etc.
- **Labor and employment law:** preventing discrimination in hiring, evaluation, and placement, and impacts on the working environment, etc.
- **Intellectual property:** rights in training data and rights relating to generated outputs, etc.

- **Fair competition and antitrust law:** algorithmic collusion and discriminatory treatment, etc.
- **Other sector-specific regulations:** industry-specific laws and guidelines in finance, healthcare, education, transportation, and other sectors, etc.

In addition, cross-border data transfers, terms of use for cloud and AI services, and platform policies—often referred to as “soft law” or contractual constraints—also function, in practice, as part of compliance. The CAIO and the AI Governance Office, in coordination with the legal and compliance functions, should conduct “regulation mapping” in the following ways:

- Create and periodically update a regulatory matrix that lists, for each major use case, potentially applicable laws, guidelines, and platform policies.
- For high-risk AI, directly incorporate relevant requirements from applicable laws and guidelines into the AIIA and operational rules.
- At the stage of considering new business expansion (expansion into new markets or jurisdictions, entry into regulated industries, etc.), the CAIO and the legal function should jointly conduct an initial risk and legal assessment and use it as input for Go/No-Go decisions and the assessment of prerequisites.
- Share information on amendments to major laws and guidelines and on significant court precedents from the legal function to the AI Governance Office, and revise policies and templates as needed.

Through these efforts, the organization aims to move beyond case-by-case responses and toward embedding jurisdiction-specific requirements into use case design and operational processes.

9.4 Basic Framework and Actors for Conformity Assessment Schemes

Many international standards and regulations related to AI are premised on a conformity assessment framework. Conformity assessment generally involves the following levels and actors:

- **First-party assessment (self-declaration of conformity):** an organization assesses and declares that its systems and management meet applicable requirements.
- **Second-party assessment:** customers, partners, and other parties assess an organization as part of a business relationship.

- **Third-party assessment and certification:** a third-party certification body accredited by an accreditation body assesses and certifies an organization based on international standards and other criteria.
- **Assessment by regulatory authorities or designated bodies:** in certain regulatory regimes, competent authorities or designated bodies conduct assessment and certification.

In addition, actors that support conformity assessment schemes include accreditation bodies that accredit certification bodies, scheme owners that operate certification schemes, and customers and regulatory authorities that rely on certification results.

Based on the maturity of the organization's AI governance and its business strategy, the CAIO and the AI Governance Office are responsible for designing which areas should undergo which level of conformity assessment. This includes, for example, the following:

- Establish a policy on which standards and schemes (e.g., ISO/IEC 42001, ISO/IEC 27001, industry-specific schemes) to pursue third-party certification for.
- For Rights and safety impacting AI and high-risk use cases, design the combination of conformity assessment approaches—such as self-declaration of conformity, internal audits, third-party assessment, and notifications to regulatory authorities—by linking them to use case classification.
- Select appropriate certification bodies and confirm that they hold valid accreditation for the relevant schemes.
- Manage certification validity periods, planned surveillance audits, scope, findings, and the status of corrective actions as an "external certification and registration inventory."
- Reflect findings and recommendations obtained through certification and external assessments in improvements to AI policies, processes, KPIs, and audit plans, and share them through the Company-wide AI Steering Committee.

These policies and mechanisms should be designed in an integrated manner with company-wide risk management, while maintaining alignment with "5.7 External Conformity Assessment and Certification" and "16.4 External Audit and Assessment".

9.5 Policy for Leveraging ISO/IEC 42001 (AIMS) Certification

ISO/IEC 42001 is an international standard for an AI management system (AIMS) and aims to manage AI-specific risks and governance within a management system framework. By obtaining ISO/IEC 42001 certification, an organization can demonstrate to third parties that its AI governance mechanisms are established at a certain level and are being operated on a continuous basis in line with the PDCA cycle. At the same time, certification does not mean that “compliance is complete once obtained.” Rather, it is one means of externally demonstrating continuous improvement of AI governance.

When leveraging ISO/IEC 42001 certification, the CAIO and the AI Governance Office should consider at least the following points:

- **Scope design:** design, based on business strategy and the risk profile, which organizational units, processes, and systems related to AI activities will fall within the scope of the AIMS. Also consider the potential for integration with existing management systems such as ISO/IEC 27001.
- **Gap analysis:** analyze the extent to which current AI governance (policies, AIIA, risk management, training and talent, etc.) aligns with ISO/IEC 42001 requirements and develop improvement plans starting with the highest-priority gaps.
- **Management system integration:** in coordination with functions responsible for ISMS, QMS, and other systems, integrate internal audits, management reviews, and corrective and preventive action processes to the extent feasible to reduce duplication and operational burden.
- **Selection of and engagement with certification bodies:** select appropriate certification bodies and align the audit plan (initial audit, surveillance audits, and recertification audits) with the medium- to long-term AI governance plan.
- **Integrated operation with other frameworks:** operate, in an integrated manner within the AIMS, the principles, processes, and controls set out in the AI Guidelines for Business, the NIST AI Risk Management Framework (NIST AI RMF), ISO/IEC 23894, and other frameworks, thereby reducing duplication and gaps.
- **External communications and stakeholder engagement:** appropriately use certification status, scope, and operational status as inputs for explanations to customers and users, regulatory authorities, shareholders, and other stakeholders, while organizing messaging so that certification is not treated as an undue “blanket exemption.”

Whether to pursue ISO/IEC 42001 certification depends on business strategy, the regulatory environment, counterparties’ requirements, and other factors. In all cases, however, it is useful to leverage the AIMS concept as a foundation for AI governance design.

10. Security and Privacy

10.1 AI-Specific Threats

AI systems—particularly those that leverage generative AI and large language models—face AI-specific threats in addition to threats common to conventional information systems. Because these threats manifest differently across stages of the AI lifecycle, it is important to identify and understand them in a stage-based manner. Examples are provided below but are not limited to these.

1) **Data Collection and Design Stage**
   - Inappropriate data collection and deficiencies in obtaining consent.
   - Collection and use of data that may involve rights infringement.
   - Design deficiencies in datasets containing bias (e.g., under- or over-representation of certain attributes).
2) **Training and Tuning Stage**
   - Data poisoning (malicious tampering with or injection into training data).
   - Unauthorized access to training environments (theft of models, data, or code).
   - Use of data or models in violation of license terms.
3) **Deployment and Inference Stage**
   - Prompt injection and adversarial inputs (prompt-based attacks).
   - Model extraction attacks, model inversion attacks, and membership inference attacks.
   - Leakage of confidential information or personal data through outputs.
   - Hallucinations (generation of incorrect or false information) lead to misguidance.
   - Inducing misclassification or erroneous decisions through adversarial examples.
4) **Operations, Maintenance, and Supply Chain Stage**
   - Vulnerabilities in external models, APIs, libraries, and tool chains.
   - Performance degradation or reduced safety due to model updates or parameter changes.
   - Impacts from incidents involving vendors or cloud service providers.
   - Leakage due to inappropriate storage or management of logs and monitoring data.

In coordination with the CISO and the information security function, the CAIO is responsible for developing an “AI threat catalog” that lists AI-specific threats such as those above and integrating it with existing information security policies and risk registers. In addition, because threats surrounding generative AI are evolving rapidly, it is desirable to review the catalog and mitigation policies at least once a year as a general rule, taking into account external threat intelligence, vendor information, and insights from relevant communities.

10.2 Security Design

Security for AI systems should be designed in alignment with the existing ISMS, based on Secure-by-design and Zero Trust. In addition, controls that address AI-specific threats should be added on top.

1) **Basic Design Principles**
   - **Identity and Access Management**

     Design authentication and authorization appropriately for each user, service, and API, and enforce the principle of least privilege. Strictly restrict access to administrator privileges and model management privileges and apply multi-factor authentication.
   - **Environment Separation**

     Logically and physically separate development, testing, and production environments, and avoid casual use of production data in development or testing environments.
   - **Management of Secrets**

     Manage API keys, tokens, credentials, and similar secrets in a secure store, and do not embed them in code, prompts, or repositories.
   - **Logs and Audit Trails**

     Log inputs, outputs, model versions, configuration changes, access histories, and similar items appropriately, and define tamper prevention measures and retention periods.
   - **Configuration Management and Change Management**

     Record changes to model versions, prompt templates, and configuration values, and define review and rollback procedures.
2) **AI-Specific Controls**
   - **Input Validation and Filtering**

     Establish mechanisms to detect and block prompt injections and harmful inputs. Where actions to external systems (such as tool execution) are performed, thoroughly enforce input validation and sanitization.
   - **Output Filtering and Sanitization**

Design mechanisms to detect outputs containing personal data, confidential information, harmful content, and similar items, and route them to masking, blocking, or review.

- **Safety Layer Design**

  A safety control layer implemented around large AI models to compensate for safety, consistency, and governance alignment that cannot be ensured by the model alone. Its purpose is not limited to input/output filtering; rather, it systematically applies decisions such as allow, block, or require review by interpreting conversation context, user intent, and the impacts of tool execution, using rule-based filters, additional models, and other techniques, thereby preventing dangerous instructions and undesirable behavior.

- **Access Controls for RAG and Similar Mechanisms**

  When combining retrieval-augmented generation (RAG) with internal document search, control which data can be referenced according to user permissions and design the system so that information outside a user's authorization cannot be retrieved.

3) **Data Classification and Use of External AI Services**

In coordination with the data classification referenced in "11. Data Governance and Quality Assurance" (confidential, internal-only, public, etc.), define, for each classification, whether data may be sent to external AI services and what masking requirements apply. For example:

- **Confidential information and specific personal information:** As a general rule, do not send to external AI APIs; alternatively, allow sending only after anonymization or masking.
- **Internal-only information:** Allow limited sending after confirming intended use, contractual conditions, and technical safeguards.
- **Public information:** Allow sending as necessary.

In coordination with the CISO, the CDO, and legal, the CAIO should document these rules as the "AI Use Policy" and "Technical Standards."

4) **Threat Modeling and Red Teaming**

For critical AI systems, conduct threat modeling at the design stage (identifying potential attack paths and misuse scenarios) and design controls based on the results. In addition, before go-live and periodically after go-live, conduct red teaming exercises that assume prompt injection, model extraction attacks, and similar threats to test for unexpected behavior and weaknesses.

As a general rule, results of red teaming and vulnerability assessments should be incorporated into risk registers, KPIs, and improvement plans, and shared through the Company-wide AI Steering Committee and information security committees and similar bodies.

10.3 Incident Response

AI-related incidents are not limited to cyberattacks or system outages. They also include cases that, due to AI-specific behavior or outputs, may have a significant impact on rights and safety, fairness, privacy, and other interests. In this manual, an “AI incident” generally includes cases such as the following:

- Unauthorized access, data leakage, service disruption, and similar events via models or AI services
- Incidents in which AI malfunctions or hallucinations caused significant misguidance or loss to customers and users
- Incidents in which discriminatory or unfair decisions or outputs by AI became apparent
- Unintended transmission of confidential information or personal data to external AI services
- Incidents in which AI-generated content included illegal or harmful information such as child sexual abuse material (CSAM)
- Incidents that may trigger reporting obligations to regulatory authorities, etc.

Incident response should be integrated with existing CSIRT and information security incident response processes, while incorporating AI-specific perspectives. The basic flow is as follows.

1) **Detection and Reporting**

   Identify signs of an AI incident through log monitoring and alerts, frontline reports, and complaints or inquiries from customers and users. When it is confirmed that the matter relates to AI, promptly share it with the CAIO.

2) **Initial Response**

   Implement interim measures to prevent further harm (e.g., disabling functions, blocking access, changing configurations). This may include temporarily suspending the affected system, as appropriate.

3) **Impact Assessment**

In coordination with the CISO, the CAIO, the DPO, and relevant functions (business functions, legal, customer support, etc.), assess the following:

- Impact scope (number of cases, affected individuals, geographic scope, etc.)
- Types of impact (life, physical integrity, property, privacy, reputation, etc.)
- Potential violations of laws/regulations or contractual obligations

4) Containment and Interim Remediation

As the root cause is progressively identified, implement necessary containment measures (e.g., configuration changes, model rollback, disabling specific functions, strengthening access controls).

5) Root Cause Analysis and Development of Permanent Measures

Analyze technical, organizational, and process-related factors, and develop permanent measures to prevent recurrence (e.g., design changes, revisions to operational rules, strengthened training).

6) Notification and Explanation

Depending on severity, consider notification to and explanation for the following parties, and take necessary actions in accordance with laws and regulations, contractual obligations, and internal policies:

- senior management and the board of directors
- customers and users, and partners
- regulatory authorities and supervisory bodies

7) Recording, Learning, and Improvement

Record the incident circumstances, response actions, and recurrence prevention measures, and incorporate them into the risk register, AI inventory, AIIA, model cards, and related materials. Through periodic reviews and audits ("16. Audit, Monitoring, and Reporting"), continuously improve the response process.

With respect to severity criteria, it is desirable to define as "high severity" those incidents that could have a significant impact on human life, physical integrity, liberty, property, or other fundamental rights, and to establish in advance criteria for immediate escalation to the CAIO, the CISO, the DPO, and senior management, as well as reporting to regulatory authorities where necessary.

In addition, for Rights and safety impacting AI, it is desirable to conduct tabletop exercises and simulations at least once a year that assume AI incidents, to verify the effectiveness of response arrangements and coordination.

10.4 Privacy

Privacy protection in leveraging AI should be designed based on the concept of Privacy-by-design/Privacy-by-default, while premised on alignment with personal data protection laws and regulations (APPI, GDPR, etc.).

1) **Design Principles and Linkage Between PIA and AIIA**
   - Position the Privacy Impact Assessment (PIA), which evaluates impacts on personal data and privacy in advance, as part of the AIIA or as a parallel process.
   - For Rights and safety impacting AI, require the PIA and confirm the results in the use case approval process and in the Company-wide AI Steering Committee.
   - Document as design-stage requirements principles such as specifying purposes for data collection and limiting purposes of use, data minimization, and limits on retention periods.
2) **Distinguishing and Managing Training Data and Inference Data**
   - Distinguish and manage training data (data used for model training and tuning) and inference data (data input by users during operations).
   - For inference data, establish a policy such as not using it for model retraining by default, or allowing use only where explicit consent has been obtained from the data subject.
   - For both training and inference, anonymize, pseudonymize, or aggregate data to the extent possible to reduce identifiability.
3) **Responding to Data Subject Rights**
   - Establish processes to respond to requests from data subjects—such as requests for disclosure, correction, deletion, suspension of use, or objection—regarding the use of personal data for training and inference.
   - For specific use cases or models, technically and operationally assess the feasible scope and methods of responding to deletion requests for personal data (retraining, filtering, masking, etc.) and organize response policies.
   - Provide users with clear notification and explanation of intended use of AI, data handling, and whether data is used for training.
4) **Cross-Border Transfers, Third-Party Provision, and Use of External Services**
   - Organize legal requirements (consent, contracts, verification of appropriate safeguards, etc.) where data transmission to cloud services or external AI APIs

constitutes cross-border transfer or third-party provision and reflect them in the PIA and procurement processes.

- In contracts with vendors and cloud service providers, clarify the scope of use of training data and inference data, whether reuse is permitted, retention periods, and the handling of sub processors (see “12. Procurement and Vendor Management”).

5) Role Allocation Between the CAIO and the DPO

- The CAIO is responsible for overseeing privacy protection policies and design principles in leveraging AI and reflecting them in the architecture and operations of AI systems overall.
- The DPO (or a personal information protection manager) oversees legal requirements, guidelines, and the adequacy of the PIA, and serves as the point of contact with regulatory authorities.
- Both parties should work closely on PIA and AIIA, the AI inventory, AI incident response, and training and awareness initiatives, to balance privacy protection and leveraging AI.

As described above, security and privacy are core elements of AI governance. The CAIO is expected to work collaboratively with the CISO, the DPO, the CDO, legal, and other relevant functions to maintain consistent controls across the design, operations, and audit stages.

11. Data Governance and Quality Assurance

11.1 Data Lifecycle Management

For data used in AI, it is necessary to manage the data lifecycle in alignment with the company-wide data governance framework led by the CDO, while incorporating AI-specific requirements. By establishing standards for each stage—data collection and acquisition, processing and integration, storage, use and analysis, sharing, and disposal and archiving—and by recording and maintaining who did what, when, and on what basis, an organization can ensure consistent quality and traceability.

In particular, the following points are important for data lifecycle management related to leveraging AI.

1) **Clarifying the Data in Scope**
   - Clarify the scope of data used for AI, including training data, evaluation data, and data entered during inference (including logs).
   - For each use case and model, organize which data sources are used for which purposes.
2) **Data Classification and Handling Rules**
   - Based on data classifications such as confidential information, specific personal information, internal-only information, and public information, define collection methods, storage methods, whether data may be sent to external AI services, and masking requirements.
   - In particular, for confidential information and specific personal information, as a general rule, prohibit sending to external AI services, or limit sending to cases after processing such as anonymization, pseudonymization, or aggregation.
3) **Records and Traceability (Data Lineage)**
   - Record and enable tracking of sources, the basis for acquisition (consent, contracts, laws and regulations, etc.), content of preprocessing, histories of integration and processing, storage locations, access permissions, and downstream use (which models and use cases use the data).
   - Link this information with the AI inventory and datasheets (see “18.3 Datasheets”) so that it can be understood on a per-use-case basis.
4) **Criteria for Disposal and Archiving**
   - Document retention periods and criteria for disposal and archiving for training data, evaluation data, logs, and other data types.

- After the retention period has elapsed, delete data in an irrecoverable manner, or, where necessary, archive it after anonymization or aggregation.

5) **Clarifying Roles and Responsibilities**
    - The CDO leads the overall data lifecycle management framework, while the CAIO defines additional requirements related to leveraging AI (use of external AI services, AI-specific risks, etc.).
    - Each business division and system owner operates in accordance with the defined standards and provides necessary information to the AI Governance Office and the data governance organization.

By systematically managing the lifecycle of data used for AI in this manner, an organization can achieve highly transparent operations that withstand audit and remediation, while reducing risks related to quality, privacy, and license legality.

11.2 Quality Assurance

The quality of data used in AI directly affects not only model performance but also fairness, explainability, and trustworthiness. In addition, the legality of licenses for third-party materials included in training and test data is critical from the perspective of compliance and reputational risk. This section presents a quality assurance framework from two perspectives: “data quality” and “license legality.”

1) **Data Quality**

From a data quality perspective, consider and manage at least the following items.

Confirming Representativeness

- Confirm, in light of intended use contexts, users, and target populations, whether data has been collected without undue bias (i.e., without under- or over-representation of specific groups).
- For Rights and safety impacting AI and high-risk use cases, conduct representativeness checks with particular rigor.

Managing Missing Values and Noise

- Define policies for handling missing values, inconsistent data, outliers, and similar issues, and record how such issues were handled.
- Understand the relationship between raw data and preprocessed data and assess impacts on quality.

Bias Analysis

- Analyze whether bias related to attributes such as gender, age, origin, and disability status affects outcomes.
- Consider and, where necessary, implement mitigation measures when bias is detected (resampling, weighting, constraints on performance by attribute, etc.).

Update Frequency and Data Drift

- Define update frequency and validity periods for data and periodically confirm whether outdated data is adversely affecting decision-making.
- Monitor changes in data distributions (data drift), and where thresholds are exceeded, treat them as triggers to consider retraining and re-evaluation.

Tracking Data Provenance

- Enable understanding of where data came from and how it reached its current form through processing and integration (data lineage).
- Organize provenance information in datasheets and the AI inventory so that it can be used for model explanation and audits.

Establishing Evaluation Data and Golden Data

- For key use cases, define golden datasets for performance evaluation and regression testing, and use them for performance comparisons and quality checks across model versions.
- Define evaluation metrics (Accuracy, Recall, False positive rate, etc.) and thresholds in advance, and record reasons and the scope of impact when changes are made.

For Rights and safety impacting AI and high-risk use cases, it is desirable to increase the frequency and rigor of these checks and shorten re-evaluation cycles.

2) License Legality

From the perspective of license legality, confirm that third-party materials (content, databases, software, etc.) included in training and test data are used within the scope of appropriate rights clearance and use conditions. Consider at least the following items.

Identifying Right Types and Rights Holders

- Organize the types of relevant rights, such as copyright and neighboring rights, database rights, portrait rights, and publicity rights, and identify rights holders to the extent possible.

Confirming Use Conditions and Licenses

- Confirm use conditions for open data, open source, commercial licenses, bespoke agreements, and other arrangements (whether modification is permitted, whether commercial use is permitted, redistribution conditions, attribution requirements, etc.).

- For data obtained through web scraping and similar methods, consider appropriate handling in light of site terms of use, robots.txt, copyright notices, and other factors.

Rights in Outputs and Allocation of Responsibilities

- Organize, in contracts and terms of use, matters such as ownership of outputs generated by generative AI, allocation of responsibilities in cases of third-party rights infringement, and the scope of disclaimers and indemnities (for details, see "12. Procurement and Vendor Management" and "15.6 Intellectual Property and Licenses").
- Clarify notices and use conditions for users (commercial use, redistribution, etc.) and, where necessary, reflect them in guidelines or terms of use.

Documentation and Clarifying Conditions for Reuse

- For each dataset used, document rights relationships, use conditions, and whether reuse is permitted, and reflect them in datasheets and related documentation.
- Organize conditions regarding reuse, secondary use, and third-party provision, so that they can be referenced when extending to other projects or externally.

While referring to "12. Procurement and Vendor Management" for contract clause design and "15.6 Intellectual Property and Licenses" for risk identification, this section should be used operationally as concrete checklist items referenced in day-to-day dataset design and use.

11.3 Transparency Documentation

Transparency documentation, such as model cards and datasheets, is foundational documentation that organizes an AI system's assumptions, design, performance, limitations, risks, and conditions of use, and supports accountability to internal and external stakeholders. It should be used not as a formalistic form, but as practical materials for decision-making and explanation.

1) **Positioning and Relationship With Other Documentation**

- A model card primarily documents a model's purpose, assumptions, performance, limitations, risks, and monitoring methods, and is positioned as a summary of AIIA results and risk assessment findings.

- A datasheet documents the provenance, composition, quality, bias, and license conditions of datasets used for training and evaluation and reflects outcomes from “11.1 Data Lifecycle Management” and “11.2 Quality Assurance.”
- These documents should be referenceable from the AI inventory and treated as core information for understanding the overall picture for each use case.

2) Responsible Owners for Drafting and Review

- A model card should be prepared by the model owner (development team or product team) and reviewed by the AI Governance Office under the CAIO from governance and risk perspectives.
- A datasheet should be prepared primarily by the data governance organization under the CDO and reviewed by legal and the DPO from license and privacy perspectives.
- As needed, business divisions, the security function, the quality assurance function, and other relevant functions should review the content.

3) Timing for Creation and Updates

- As a general rule, it is desirable to create and update these documents at least at the following times:
  - When a PoC or similar effort reaches a meaningful conclusion (as input for go/no-go decisions for broader deployment)
  - Before go-live (as part of the approval process)
  - When significant changes occur (changes in data sources, large-scale retraining, algorithm changes, expansion of intended use, etc.)
- Record update history and versions so that it can be confirmed later what assumptions were in place at a given point in time.

4) Storage, Access, and Version Control

- Manage transparency documentation centrally in a central repository (such as a document management system), and store approved versions separately from drafts.
- Link these documents from the AI inventory so that relevant stakeholders can access them easily.
- Retain them for an appropriate period to support audits, incident response, and explanations to regulatory authorities.

5) Policy on External Disclosure

- Determine the scope and level of detail for external disclosure of transparency documentation on a use-case basis, balancing competitive confidentiality with the need to build trust. For example:
  - For Rights and safety impacting AI, consider publishing a user-facing summary model card (purpose, key assumptions and limitations, whether human oversight is in place, etc.).
  - For internal tools and models that are sources of competitive advantage, limit disclosure to internal use, while maintaining a detailed internal version sufficient for regulatory and audit readiness.
- The disclosure scope and methods (website, transparency reports, terms of use, FAQs, etc.) should be decided by the CAIO in coordination with public relations, investor relations, and legal.

Transparency documentation, alongside AIIA, the risk register, and KPI dashboards, is an important element supporting the “visualization” of AI governance. The CAIO should promote adoption from both process and culture perspectives so that these documents are not merely formalities, but are actually used for use case approval, review, and explanation.

12. Procurement and Vendor Management

12.1 Procurement Policy

For AI-related procurement, it is important to document comprehensive evaluation criteria that include AI governance perspectives, not only technical fitness and price. In coordination with the procurement function, legal, information security function, and other relevant functions, the CAIO should establish a vendor evaluation policy that includes perspectives such as the following:

- **Technical fit:** functional requirements, performance requirements, ability to integrate with existing systems, scalability, and support model and SLAs.
- **Governance/transparency:** whether the vendor can provide model cards, datasheets, AIIA, and other materials; clarity of intended use, assumptions, and limitations.
- **Security:** certification status such as ISMS, vulnerability management processes, and mitigation policies for AI-specific threats.
- **Privacy and data protection:** personal data handling policies, conditions for use of training data and inference data, and the status of PIA implementation.
- **Fairness and explainability:** whether bias assessment, fairness assessment, and mitigation measures are in place; features and documentation related to explainability.
- **Compliance:** policies for alignment with the AI Guidelines for Business and legal and regulatory requirements (such as the EU AI Act), and consistency with sector-specific laws and regulations and platform policies.
- **Operational track record and trustworthiness:** operating track record, key customers and use cases, incident history, and the status of corrective actions.
- **Sustainability:** disclosure and reduction efforts regarding energy consumption and environmental impacts associated with training and inference.
- **Vendor lock-in risk:** whether data/models can be exported, adoption of standard technologies, and substitutability.

To the extent possible, these evaluation criteria should be disclosed to vendors in advance to ensure comparability and repeatability. In addition, for procurement related to Rights and safety impacting AI and high-risk use cases, it is desirable to increase the weighting of governance, security, privacy, fairness, compliance, and lock-in perspectives, and, where appropriate, make third-party certification or external assessments a condition.

After obtaining approval from the Company-wide AI Steering Committee, the vendor evaluation policy should be communicated internally. When applying exceptions or setting special conditions on a case-by-case basis, document the reasons and associated risks.

12.2 Examples of Mandatory Contract Clauses

When procuring AI-related services or models, it is necessary to design contract clauses that reflect AI-specific risks and governance requirements, not only performance and price. The following are typical examples of mandatory clauses; specific content should be determined on a case-by-case basis in consultation with the legal function, the DPO, and other relevant stakeholders.

1) **Clauses on Performance, Limitations, and Risks**
   - Document the functions and performance metrics of the AI system provided (e.g., accuracy, response time, availability) and the methods for measuring them.
   - Require clear statements of “what the system can and cannot do”, “intended and non-intended use scenarios”, and “known risks and limitations (hallucinations, bias, etc.)”.
   - Define the policy for model updates (frequency, notification method, compatibility, and validation period).
2) **Use Restrictions and Prohibited Uses**
   - Clearly specify permitted and prohibited purposes of use (e.g., certain surveillance uses, discriminatory uses, generation of illegal content).
   - Define whether use is permitted for purposes that qualify as Rights and safety impacting AI, and any additional conditions (such as human oversight requirements).
3) **Data Handling and Rights (Training, Evaluation, and Logs)**
   - Confirm rights (copyright, database rights, etc.) and license conditions for third-party materials included in training/test data and require that the lawful scope of use be specified in the contract.
   - For data provided by the customer (training data, inference input data, logs, etc.), clarify:
     - intended use (for service provision, for model improvement, etc.)
     - whether secondary use is permitted (including use for improving models for other customers)
     - retention periods and deletion methods

- Define ownership of outputs and the allocation of responsibilities and the scope of indemnity clauses in cases of third-party rights infringement.

4) Continuous Improvement and SLAs

- Define SLAs (service-level agreements) for model updates, vulnerability remediation, bug fixes, and responses to performance degradation.
- Agree on key KPIs and thresholds for performance and quality, and set out corrective actions, fee adjustments, the right to terminate, and other remedies where thresholds are not met.
- For significant specification changes or model changes, provide advance notice and a validation period.

5) Monitoring and Audit Rights

- Define the scope, format, and retention period for log sharing (inputs/outputs/metadata, etc.).
- Require provision of interfaces for customer evaluation and verification (APIs, logs, test environments, etc.).
- Where appropriate, establish audit rights, including on-site or remote audits, provision of third-party audit reports (SOC reports, etc.), and submission of remediation plans.

6) Security Requirements

- Define cooperation for attack resilience assessments (penetration testing, red teaming, etc.) and the conditions for conducting them.
- Document processes for vulnerability reporting and security incident reporting (reporting deadlines, content, points of contact, etc.).
- Define required levels of technical and organizational measures, including retention of input/output logs and access controls, encryption, backups, and similar measures.
- Confirm baseline mitigation policies for AI-specific threats (prompt injection, model extraction, etc.).

7) Privacy and Data Protection Requirements

- Contractually require principles such as data minimization, purpose limitation, and retention period limits.
- Define policies and conditions regarding cross-border transfers, third-party provision, and use of sub processors (including whether prior notice or consent is required).

- Organize responsibilities and practical procedures for responding to data subject rights (disclosure, correction, deletion, suspension of use, etc.).
- Define cooperation obligations where a PIA/DPIA (Data Protection Impact Assessment) is required.

8) **Environment and Sustainability**
   - Require disclosure of information on energy consumption associated with training and inference, data center efficiency metrics (e.g., PUE), and CO2 emissions intensity.
   - Where there are targets or plans to reduce environmental impacts, require sharing of an overview.

9) **Termination and Transition Clauses (Mitigating Vendor Lock-In)**
   - Define the scope for return or export of models, data, configuration information, logs, and related materials upon contract termination or significant changes in conditions.
   - Agree on transition support to another vendor or in-house systems (duration, cost, and scope).
   - To mitigate vendor lock-in, consider the use of standard formats and standard protocols, and similar measures.

After organizing these points, it is operationally useful to consolidate them into a checklist format, as follows, to support contract negotiations and reviews.

Summary of Mandatory Contract Items (Examples)

- Documentation of performance and limitations (“can/cannot do” / known risks)
- Restrictions on use (clear specification of prohibited uses)
- Rights and handling of training/test data and customer data (license legality, reuse conditions)
- Ownership of generated outputs and allocation of responsibilities in cases of third-party rights infringement
- Continuous improvement (update plans, vulnerability remediation, SLAs)
- Monitoring and audit (log sharing, agreement on evaluation metrics, audit rights)
- Security requirements (attack resilience assessments, vulnerability reporting processes, cooperation on red teaming)
- Privacy requirements (minimization, purpose limitation, deletion rights, data transfers and sub processors)

- Environmental considerations (energy use for training and inference, PUE, CO2 emissions intensity, disclosure of reduction plans)
- Termination and transition (mitigating vendor lock-in, return of models and data, transition support)

This section provides general examples. The drafting and interpretation of specific contract clauses should be conducted based on advice from each organization’s legal function and, where appropriate, external experts.

12.3 Vendor Risk Assessment

Risk assessment for AI-related vendors should be a continuous process conducted throughout the contract term, rather than a one-time review at the time of procurement. In coordination with the procurement function, legal, the information security function, and business functions, the CAIO should design a vendor risk assessment framework that covers at least the following perspectives:

- **Technical risks:** technology maturity, architectural robustness, scalability, and risks related to dependent technologies (specific cloud providers, foundation models, etc.)
- **Operational risks:** service continuity, support model and SLAs, SLA performance, and incident/outage response capability
- **Security and privacy risks:** security certifications (e.g., ISO/IEC 27001), privacy certifications, mitigation measures for AI-specific threats, incident history, and the status of corrective actions
- **Compliance risks:** alignment policies for the AI Guidelines for Business and legal and regulatory requirements (such as the EU AI Act), consistency with related laws and regulations, sector-specific laws and regulations, and platform policies, and whether conformity assessment and certification are in place (see “9.4 Basic Framework and Actors for Conformity Assessment Schemes” and “9.5 Policy for Leveraging ISO/IEC 42001 (AIMS) Certification”)
- **Ethics and fairness risks:** whether bias assessment and mitigation measures are in place, the status of ensuring explainability, and considerations regarding impacts on human rights
- **Reputational risks:** public reputation, past scandals, litigation, or findings raised by regulatory authorities

- **Strategic and dependency risks:** the vendor’s financial base, business continuity, degree of dependence on a specific vendor (vendor lock-in), and the availability of alternative vendors
- **Environment and sustainability:** policies and track record regarding energy consumption and emissions, and sustainability commitments

Assess these elements through both quantitative and qualitative methods, and, depending on the criticality and risk level of the use case, assign a vendor-specific risk score and a risk category (e.g., high/medium/low). For Rights and safety impacting AI, high-risk use cases, and initiatives involving large-scale data processing, it is desirable to apply more detailed due diligence and higher standards.

Vendor risk assessment should be reviewed periodically not only before procurement, but also at times such as the following:

- at contract renewal
- when significant incidents occur
- when regulations or guidelines change substantially
- when there are major changes in the vendor’s financial condition or business policy

Incorporate assessment results into the AI risk register and the AI inventory and use them for use case-specific risk assessment and prioritization (see “5.3 Risk Management” and “15. Key Risks and Mitigation Measures”). Also share them through the Company-wide AI Steering Committee, risk management committees, and similar bodies, and, for high-risk vendors, consider additional controls (strengthened contract terms, evaluation of alternative vendors, increased audit frequency, etc.).

In this way, it is important to embed vendor risk assessment into the end-to-end process of procurement, contracting, and operations, and to operate it continuously as part of overall AI governance.

13. Training and Talent

13.1 Designing Training Programs

To practice AI governance, it is necessary to provide role-appropriate training and reskilling for all personnel involved in AI-related work. Training programs should be designed and implemented to include the following elements.

1) **Objectives and Outcomes**
   - **Embedding an AI governance culture:** Provide all employees with a baseline understanding of AI governance and foster a culture that supports ethical and lawful use of AI across the organization.
   - **Enhancing specialized knowledge:** For technical professionals and managers, deepen specialized knowledge of AI technologies, AI-related laws and regulations, and risk management, and build the capability to apply it in projects and day-to-day work.
   - **Capability assessment and skill development:** After training, conduct capability assessments (understanding of AI literacy, examples of application in projects) and, together with feedback, develop reskilling plans.
2) **Customization by Audience**
   - **Training for Executives:** Provide senior management with training on the overall direction of AI governance and its business value. Content should include, in particular, advances in AI technology and associated risks, the importance of governance, and the latest regulatory developments (such as the EU AI Act), as well as topics related to risk, compliance, and social responsibility.
   - **Training for technical professionals:** Provide data scientists and engineers with technical training on AI algorithms, risk assessment methods (AIIA, bias assessment, etc.), and methods for model training, validation, and monitoring. Also provide tools and best practices useful for actual model development and operations.
   - **Training for non-technical staff:** For sales, marketing, support, and similar functions, provide training that builds understanding of basic ways to leverage AI, how to handle data, and ethical considerations and legal obligations (including data protection), and strengthens the ability to provide feedback on AI systems.

3) **Skill Sets and Training Content**

- **Core AI governance skills (all employees):** Teach basic understanding of AI, how AI affects business, AI risks and legal issues, and the importance of privacy protection.
- **Specialized knowledge (specialists):** Build practical knowledge of model evaluation and risk management (AIIA, bias assessment, retraining management, etc.) and laws and regulations (GDPR, the EU AI Act, etc.).
- **AI ethics skills (all employees):** Teach ethical issues in AI (bias, fairness, explainability, etc.) and practical methods for ensuring transparency.
- **Use of technologies and tools (engineers):** Strengthen skills in using tools for model development and operations (e.g., industry-standard machine learning frameworks), MLOps tools, and security tools.

4) **Continuous Training Plan**

Because AI technologies and laws and regulations are evolving continuously, training should not be treated as a one-time effort. Regular updates and follow-ups are essential. This enables employees to reflect on the latest technologies and regulatory developments in practice and enhances the maturity of the AI governance structure.

13.2 Reskilling and Career Paths

To keep pace with the rapid evolution of AI, reskilling should be understood not merely as acquiring skills, but as a process that enables employees to grow throughout their careers. In particular, for employees, including non-technical staff, it is important to help them identify new roles in AI-related work and provide opportunities to contribute as next-generation leaders.

1) **Need and Objectives for Reskilling**

As AI governance becomes essential to business operations, re-educate the workforce on required skills and increase both understanding and effective use of AI technologies. This enables organizations to accelerate new value creation through AI. Improving AI literacy contributes to operational efficiency, higher-quality decision-making, and strengthened competitive advantage.

2) **Designing AI-Related Career Paths**

- **Career paths for technical professionals**

  For engineers and data scientists, clearly design pathways from foundational AI knowledge to advanced algorithm design, risk management, and leadership roles in AI governance. For example, define career paths from technical roles to leadership roles in AI governance teams, providing opportunities not only to deepen expertise but also to develop management skills.

- **Career paths for non-technical staff**

  For sales, marketing, support, and similar functions, provide training to improve AI literacy and strengthen skills for AI deployment and use, and design career paths specialized in specific areas (AI product sales, AI customer support, etc.).

- **Company-wide career paths**

  For all employees, after they acquire foundational AI-related skills, provide pathways to serve as AI adoption leads within their respective functions, and offer broad career path options. For example, encourage role transitions to positions such as AI project managers and internal AI consultants.

3) Practice-Oriented Reskilling

To ensure reskilling outcomes are reflected immediately in day-to-day work, integrate real operational cases into training programs. By offering opportunities to learn while participating in actual AI projects, employees can test what they have learned in practice and reinforce skill retention. In particular, where there are high-risk AI projects, involving employees in those projects can help them learn, in practical terms, the importance of risk management and governance.

4) Measuring Effectiveness and Providing Feedback

After training, always measure effectiveness and periodically conduct knowledge tests, case studies, and similar exercises to confirm how acquired skills are being applied in practice. Based on feedback, adjust the content of subsequent training programs and continue to provide programs that remain aligned with the latest technologies and laws and regulations.

13.3 Evaluation and Feedback

To assess whether training programs on AI governance are effective, it is important to evaluate quantitative learning outcomes. Establish an approach to periodically assess how well training content is being applied in practice and whether it is contributing to operational efficiency and risk reduction, and to provide feedback based on those results.

1) Measuring Training Effectiveness

- **Knowledge assessment:** After training, conduct online tests and case studies to confirm whether participants have acquired a foundational understanding of AI governance.
- **Degree of practical application:** Confirm, through actual project evaluations, the extent to which participants are applying what they learned in AI projects and day-to-day work.
- **Ongoing performance indicators:** Where employees have achieved operational improvements by applying training content, measure outcomes through KPIs (e.g., success rates of AI-related projects, accuracy of risk management) and assess post-training growth.

2) Feedback and Improvement

After training, always provide feedback and confirm participants’ understanding and how they are applying the content in practice. Collect feedback from participants and reflect it in the next training program. Based on this feedback, revise training programs flexibly so that they continue to address the latest technologies, laws and regulations, and AI governance practices.

3) Reflection Across the Organization

Based on feedback and outcomes obtained, review the company-wide AI training plan and aim to raise the organization-wide level of training. Report regularly to the Company-wide AI Steering Committee to demonstrate how training contributes to strengthening AI governance across the organization and reflect this in the following year’s training strategy.

4) Linking Career Paths With Training Outcomes

For employees who achieve strong results, provide opportunities for career progression and roles as new project leaders, linking reskilling outcomes directly to practice. Establish career paths that enable promotion into AI governance leadership positions, linking training with career growth.

14. KPI, Measurement, and Dashboard

This chapter sets out concepts for KPI design, measurement, and dashboard operation to continuously understand and improve the effectiveness of AI governance. This manual recommends organizing AI-related metrics into the following five areas and monitoring them in an integrated manner through a unified dashboard:

- **Business value:** contribution to revenue and earnings, cost reduction effects, productivity improvements, customer satisfaction, etc.
- **Trustworthiness:** model performance and operational stability, data quality, trends in the occurrence of bias and errors, etc.
- **Governance:** AIIA implementation rate, number of incidents and the status of corrective actions, policy compliance status, etc.
- **Security and privacy:** security/privacy incidents, status of security testing and red teaming, PIA implementation rate, etc.
- **Talent and culture:** training participation rates, number of personnel participating in AI projects, number of internal improvement proposals and consultations, etc.

Organizations may specify, add, or revise these metrics based on business characteristics and the regulatory environment. At a minimum, however, it is desirable to cover these five perspectives in a balanced manner. The following sections describe approaches to setting and measuring KPIs and to visualizing and using them through dashboards for each area.

14.1 Setting and Managing KPIs

To operate AI governance effectively, it is important to set appropriate KPIs and manage progress on a regular basis. KPIs should be used not only to measure performance, but also to identify areas for improvement and support decision-making. The following provides a basic framework and management approach for KPI setting in AI governance.

1) Purpose and Use of KPIs

- **Supporting improvement and adjustment:** Use KPIs not only to measure the performance of tasks and projects related to AI governance, but also to identify issues and implement improvement measures early.
- **Data-driven decision-making:** KPIs provide inputs for senior management and the CAIO to make timely and accurate decisions based on data. By monitoring indicators such as AI risks, data quality, and AI model performance, KPIs function as tools to flexibly adjust governance strategy.

2) **Basic Principles for KPI Setting**

- **Specificity:** Each KPI should be measurable and clearly define what outcome is being tracked. For example, set quantitative targets such as “maintain AI model Accuracy at 90% or higher” or “maintain an on-time delivery rate of 95% for AI projects.”
- **Achievability:** KPIs should be set within realistic and achievable ranges. Overly ambitious targets may reduce motivation; therefore, set KPIs with reference to industry standards and past performance.
- **Relevance:** KPIs should align with AI governance objectives and strategy. For example, risk management KPIs should be based on risk assessment results, and evaluations of AI’s societal impacts should be linked to ethical perspectives.

3) **KPI Evaluation and Improvement Cycle**

Conduct KPI evaluations on a regular basis. Typically, evaluate KPIs every quarter and revise targets as necessary. Report evaluation results to the Company-wide AI Steering Committee and decide resource allocation and improvement measures based on KPI results. KPI management should be used not merely for tracking figures, but as a mechanism to support adjustment and improvement activities in response to progress. Through regular reviews, confirm which indicators are functioning effectively and make changes as needed.

14.2 Dashboards and Visualization

Dashboards are an important tool for quickly and intuitively understanding the status of AI governance. By visually presenting AI-related KPIs, responsible personnel can promptly identify issues and take necessary actions. Dashboards should also be designed so that stakeholders—including senior management, the CAIO, the AI Governance Office, and risk management functions—can access the data they need from different perspectives.

1) **Purpose and Use of Dashboards**

- **Information sharing:** Use dashboards so that senior management, the CAIO, and relevant functions can understand the status of AI governance in real time and make timely, data-driven decisions.
- **Early warning system:** Establish mechanisms to generate early alerts when specific KPIs fall below thresholds, enabling rapid identification and response.

- **Enhancing transparency:** To appropriately communicate AI governance status to internal and external stakeholders, visualize key indicators (data quality, AI risks, bias assessment, etc.) and support accountability.

2) **Dashboard Design and Data Visualization**
   - **Simple and intuitive design:** Design dashboards so that users can understand status at a glance. Use charts, color-coding, icons, and similar techniques effectively to present complex data clearly and succinctly.
   - **Interactive elements:** Incorporate interactive elements so stakeholders can drill down into detailed data, enabling deeper analysis.
   - **Real-time updates:** Build in mechanisms to update KPIs and related data in real time so users can respond promptly to changes.

3) **Dashboard Users and Role Allocation**
   - **Senior management:** Understand progress in company-wide AI governance and make necessary resource allocation and adjustments. Dashboards for senior management should focus on overall results of AI projects and key risk indicators.
   - **CAIO:** As the role responsible for managing overall AI governance, monitor KPIs related to project progress, risks, and quality management, and promptly communicate issues to relevant stakeholders.
   - **AI Governance Office / Risk management functions:** Monitor risk assessments and the status of AI project evaluations in detail and analyze data through dashboards to inform decisions on improvement measures.

4) **Dashboard Updates and Improvement**

   Dashboards should continuously reflect the latest data and be reviewed regularly. Update dashboard content in response to new AI projects, changes to KPIs, and new risk management indicators. Collect feedback from dashboard users and improve visualization methods and data provisioning approaches.

14.3 Regular Reviews and Improvement

KPIs and dashboards should not merely be set; their effectiveness must be maintained through regular evaluation and improvement. Improvements in AI governance performance are achieved through a continuous improvement cycle. Through the following process, review KPI and dashboard operations and implement necessary improvements.

1) Regular Reviews

At least every quarter, conduct reviews of KPIs and dashboards through the Company-wide AI Steering Committee and relevant functions (CAIO, CISO, business functions, etc.). In these reviews, confirm whether each KPI is functioning as intended, whether progress is being made toward targets, and whether any issues have been identified.

2) Feedback and Improvement

Receive feedback on KPIs and dashboards from relevant stakeholders and, where necessary, adjust indicators and evaluation methods. Based on feedback, identify new issues and areas for improvement, and propose and implement solutions by the next review.

3) Developing Improvement Actions

Based on review results, develop necessary improvement actions and reflect progress in the dashboard. Review KPIs, reallocate resources for improvement, and adjust priorities, and report progress at the next review.

4) Continuous Performance Improvement

Where targets are not met, identify issues and implement measures to address root causes. By repeating regular reviews and improvement activities, enhance AI governance performance and strengthen risk management and ethical use of AI across the organization.

15. Key Risks and Mitigation Measures

This chapter organizes risks that are particularly likely to arise in connection with establishing the CAIO function and operating AI governance, along with representative mitigation measures. The risks listed here are not an exhaustive catalog; rather, they serve as a starting point for organizations when designing their risk register. It is desirable for the CAIO to add to and further specify risk items based on business characteristics and jurisdiction-specific requirements, and to review them regularly.

15.1 Excessive Centralization

Strengthening the CAIO function can be effective for improving consistency in company-wide governance and centralizing risk management. However, if decision-making becomes excessively centralized, there is a risk that business divisions' autonomy and speed will be undermined. In addition, if all decisions are concentrated with the CAIO, bottlenecks and excessive concentration of accountability may occur.

The following designs may be considered as mitigation measures:

- **Balancing with decentralized operation**
  - Institutionalize business-led decision-making in the Company-wide AI Steering Committee, so that the business side leads use case proposals and prioritization while the CAIO coordinates and oversees them from a cross-cutting perspective.
  - Combine "decentralized experimentation" with "standardized go-live approval gates": respect each function's discretion through the PoC stage, while applying a unified approval and evaluation process for go-live.
- **Clarifying roles and authorities**
  - In AI policies for leveraging AI, specify "minimum common controls" (AIIA implementation, preparation of transparency documentation, incident reporting, etc.) and the "scope of business discretion" (detailed design of UIs and workflows, etc.).
  - Use RACI and similar tools to document role allocation and decision-making authority among the CAIO, business owners, the CISO, and the legal and compliance functions.
- **Ensuring agility in approval processes**
  - Set target lead times for approval of AI use cases and exception approval and monitor them through dashboards.

  - For low-risk and limited-scope matters, simplify approval processes on a risk basis, such as by allowing streamlined flows or post hoc reporting.
- **Enhancing transparency in exception management**
  - For exception approval, record the rationale, duration, and remediation plan, and manage them as an “exception register.”
  - After a certain period, reassess the appropriateness of exceptions; where continuation is necessary, consider revising rules so that the practice becomes standard operations rather than remaining an exception.

Through these measures, organizations can maintain the CAIO’s integrated governance function while balancing it against business autonomy and speed.

15.2 Misuse and Over-Reliance Driven by Overconfidence in AI

There is a risk of over-reliance and misuse driven by “automation bias” and “authority bias”, in which users place excessive expectations on AI performance and unconditionally trust system outputs. In particular, generative AI may present plausible but incorrect content (hallucinations) with high confidence, and this can have significant impacts on decision-making and external communications. Representative mitigation measures include the following:

- **Human-in-the-loop / Human-on-the-loop design**
  - For critical decisions (hiring and HR, credit and screening, healthcare and safety, contracts and legal judgments, etc.), require final human approval.
  - Rather than adopting AI outputs as-is, present them as “options” or “reference information”, and standardize workflows in which responsible personnel verify the rationale before making decisions.
- **Presenting uncertainty and rationale**
  - To the extent possible, provide users with AI confidence/uncertainty and the data sources used.
  - Where AI is used to draft important documents or external-facing information, build citation and rationale disclosure and fact-checking steps into the UI and processes.
- **Rules for use and training**
  - Establish usage guidelines that clearly state that “AI is not the final source of truth” and that users “must refer to primary sources and official information.”
  - In employee training, share examples of AI limitations, bias, and hallucinations, and provide practical checklists to avoid overconfidence and misuse.

- **Monitoring and remediation**
  - Classify and record incidents and complaints arising from misuse of AI outputs and analyze which processes and UI designs are contributing to misuse.
  - Based on the analysis, continuously review countermeasures such as input constraints, output filtering, and additional review steps.

Through these measures, it is important not to regard AI as an "all-purpose decision-maker", but to position it appropriately as a tool that supports human judgment.

15.3 Regulatory Uncertainty

While AI-related regulation is advancing both domestically and internationally, detailed requirements and operational interpretations can change readily, creating uncertainty in conformity determinations. When operating across multiple jurisdictions, different obligations may also be imposed on the same AI system. As mitigation measures, establish frameworks such as the following:

- **Regulatory watch and clarification of responsible owners**
  - Build a "regulatory watch" function led by the legal and compliance functions in coordination with the CAIO, and regularly review developments in AI-related laws, guidelines, and standards.
  - Clarify responsible owners for collecting and analyzing regulatory developments and reflecting them in internal rules.
- **Portfolio and requirements mapping**
  - Using the AI inventory as a foundation, create a mapping table between the organization's AI use cases/models and regulatory categories and obligations in each jurisdiction.
  - For high-risk use cases and use cases that affect rights and safety, apply more conservative criteria.
- **Agile processes for updating policies and standards**
  - Embed rapid revision processes for regulatory changes into policies, standards, and templates (AIIA, model cards, etc.), with quarterly reviews as the default.
  - Clearly record revision history, effective dates, and affected use cases, and manage them in an auditable manner.

- **Engagement with external experts**
  - Where appropriate, engage external law firms, certification bodies, and industry associations to obtain second opinions on the organization's interpretations and response policies.
  - Where regulatory authorities publish guidance or Q&As, reflect them promptly in internal policies.

While uncertainty cannot be eliminated entirely, these measures can strengthen organizational capability to keep pace with change.

15.4 Vendor Lock-In

Excessive dependence on a specific vendor's AI platform or APIs creates risks, including future cost increases, functional constraints, unfavorable changes to contractual terms, and difficulty in adopting new technologies. As mitigation measures, consider the following:

- **Reducing dependency through architecture design**
  - Use standardized interfaces and protocols to design AI components with loose coupling.
  - Prepare abstraction layers (adapters) that assume model or vendor switching and embed an "exit strategy" at the architectural level.
- **Multi-vendor strategy and portability**
  - For critical use cases, confirm the availability of alternative vendors to the extent possible, and consider test use and securing a second source.
  - Specify data portability requirements (export formats, metadata, logs) in contract clauses to prepare for future migration.
- **Contractual termination and transition clauses**
  - As set out in "12. Procurement and Vendor Management", include termination and transition clauses and define, in concrete terms, return of models and data, transition support, and support levels during the transition period.
  - Establish consultation processes for price revisions or functional changes and clarify the right to terminate.
- **Maintaining internal capabilities**
  - Maintain minimum in-house capabilities for evaluation, MLOps, and governance, and avoid over-reliance on vendors.
  - Ensure sufficient skills to internally validate evaluation reports and audit results provided by vendors.

Through these measures, organizations can maintain constructive partnerships with vendors while preserving future options.

15.5 Environmental Impacts

Training and inference for large-scale models may increase environmental impacts through higher electricity consumption and increased CO2 emissions. At the same time, there are cases where AI contributes to reducing environmental impacts, such as through optimization of business processes and energy-saving controls. As mitigation measures, the following perspectives are important:

- **Measurement and KPI integration**
    - Estimate energy consumption and CO2 emissions associated with training and inference and incorporate them into the KPIs described in "14. KPI, Measurement, and Dashboard."
    - For critical use cases, define indicators for both environmental impacts and operational efficiency gains, and visualize trade-offs.
- **Efficient model and infrastructure selection**
    - Reduce inference costs through model distillation, quantization, caching, and optimization of batch processing.
    - Rather than routinely using models that are larger than necessary, combine lightweight models and on-demand inference depending on the use case.
- **Using sustainable infrastructure**
    - Use data center or cloud provider energy efficiency metrics and renewable energy usage as factors in procurement decisions.
    - Where feasible, consider using green electricity and adopting infrastructure with environmental certifications.
- **Policies and transparency**
    - Establish basic policies on environmental impacts associated with leveraging AI and ensure alignment with ESG strategy.
    - Where appropriate, disclose environmental impact assessments and reduction efforts for key AI use cases through sustainability reporting and similar channels.

These measures help ensure that leveraging AI does not conflict with medium- to long-term sustainability goals.

15.6 Intellectual Property and Licenses

Intellectual property and licensing issues related to training/test data and generated outputs involve legal and ethical risks, including use of training or test data without authorization, copyright or trademark infringement in generated outputs, non-compliance with license terms, insufficient originality, restrictions on secondary use, and leakage of confidential information. Where accountability is unclear, it can be difficult to respond to complaints or litigation, and risks may also arise relating to publicity rights, privacy infringement, and breach of contract. As mitigation measures, controls such as the following are necessary:

- **Institutionalizing data and IP reviews**
  - In coordination with data lifecycle management and quality assurance under "11. Data Governance and Quality Assurance", establish a "data register/AI inventory" that records, for each training/test dataset and evaluation dataset, sources, licenses, and scope of use.
  - Standardize processes to confirm and record license conditions for third-party materials (images, text, code, etc.), including whether commercial use is permitted, whether modification is permitted, attribution requirements, geographic restrictions, and similar terms.
- **Policy for use of generated output**
  - Define, on a use-case basis, the permitted scope of use for generated outputs (text, images, audio, etc.) (internal-only, whether external use is permitted, whether commercial use is permitted, whether secondary use is permitted), and reflect it in terms of use and internal guidelines.
  - Where attribution statements or disclaimers are required, prepare templates so users can apply them consistently.
- **Contractual warranty and indemnification framework**
  - Ensure consistency with procurement and contract clauses under "12. Procurement and Vendor Management" and require vendors to warrant legality of intellectual property rights and licenses for training data and generated outputs.
  - Clearly define, in contracts, indemnification and allocation of responsibilities in the event of rights infringement.
- **Coordination between the CAIO and legal**

  - Based on the AI inventory and the data register, the CAIO should take an overview of which use cases contain high-risk IP and licensing issues and determine review prioritization on a risk basis.
  - In coordination with the legal function, periodically update interpretations and practical responses relating to copyright law, trademark law, contract law, and similar areas.

Through these measures, organizations can manage intellectual property and licensing risks through upfront design and contracting, rather than relying solely on reactive issue handling.

15.7 Brand and Decision-Making Risks From Misinformation, Disinformation, and Lack of Provenance

With the widespread adoption of generative AI, misinformation, disinformation, and content with unclear provenance are rapidly increasing both inside and outside organizations. If the information an organization publishes contains errors, this may lead to brand damage and legal risks. If external disinformation is mistakenly cited, it may result in errors and mis-citations in business documents and could lead to flawed decision-making. As mitigation measures, implement controls from both inbound (intake of external information) and outbound (external communications by the organization) perspectives.

- **Adding and Verifying Provenance (Outbound)**
  - For important external-facing materials, consider attaching provenance information (author, creation date, revision history, etc.), with reference to standards such as C2PA (Coalition for Content Provenance and Authenticity).
  - For official information, establish an “official information database” and/or signed distribution channels that enable authenticity to be verified, and clearly differentiate official information from disinformation.
- **Verification Gates When Using AI**
  - When using AI to draft important documents (press releases, IR materials, internal rules, customer-facing explanatory materials, etc.), embed verification gates in workflows—such as providing rationale, validating citations and sources, and dual approval—and ensure AI-generated text is not adopted as-is.
  - For external information summarized or translated by AI, require cross-checking against primary sources and confirming references as mandatory steps.

- **Evaluation of Inbound Information**
  - Establish rules for checking source reliability, whether provenance information is available, and the profile of the information publisher for information flowing in from outside (SNS posts, forums, generative AI content, etc.).
  - For information used in important decisions or policy development, adopt “confirmation from multiple sources” and “tracing back to primary sources” as core principles.
- **Integration With the Crisis Communications Plan**
  - In preparation for the spread of disinformation or false information related to the organization, include the following in the crisis communications plan: immediate presentation of primary information, disclosure of provenance evidence, dissemination of corrective information, and clear indication of the timeline for corrections and responsible parties.
  - Through simulations and exercises, establish arrangements that enable public relations, legal, and business functions to coordinate and respond rapidly.
- **Improving Internal Literacy**
  - In employee training, share typical examples of misinformation and disinformation, practical ways to detect them, and key considerations when handling generative AI content.
  - Repeatedly reinforce that, even when using AI for information search and summarization, “final fact-checking must be performed by humans.”

Through these measures, organizations can balance information authenticity with brand protection in the era of generative AI.

16. Audit, Monitoring, and Reporting

To ensure governance continues to function, a feedback loop that combines internal and external audits, monitoring, and reporting to senior management and the board of directors is essential. Through day-to-day monitoring, organizations can detect deviations and anomalies early; through internal audits and external assessments, they can verify the effectiveness of controls; and by reporting and reflecting results in management through the CAIO and the Company-wide AI Steering Committee, they can achieve continuous improvement.

16.1 Internal Audit

Internal audit is a function that independently assesses whether policies, processes, and controls related to AI governance are operating as designed and delivering the intended risk reduction effects. As a general rule, internal audit should be conducted by third-line functions such as an internal audit function. While the CAIO and the AI Governance Office are responsible for the design and operation of controls subject to audit, roles should be separated so that they do not audit themselves.

The scope of internal audit should include, at a minimum, the following:

- Status of implementation of AI policies and guidelines, and AI principles.
- Status and completeness of the AI inventory.
- Implementation rates for AIIA and PIA, model cards, datasheets, and related documentation, and the adequacy of their content.
- Status of implementation of additional controls for high-risk use cases / Rights and safety impacting AI (HITL, remediation procedures, etc.).
- Operational status of controls related to security, privacy, and data governance.
- Effectiveness of incident and complaint handling and corrective actions.
- Status and records of training and awareness initiatives.

Develop the audit plan annually and prioritize high-risk use cases and Rights and safety impacting AI on a risk basis. For low risk use cases and supporting controls, design the plan flexibly based on business characteristics and risk appetite, such as through biennial audits or thematic audits. Link audit findings to remediation plans, specify deadlines and responsible owners, and track them through completion.

To maintain alignment with external certifications, integrate schedules for annual surveillance audits and recertification audits for third-party certifications such as ISO/IEC

42001 into the internal audit plan. Manage surveillance findings in an integrated manner within internal audit remediation plans and close the loop using the same supporting evidence and records as those used for corrective action reporting to certification bodies. Reflect internal audit results and remediation status in reviews of the AI risk register and KPI dashboards.

16.2 Monitoring

Monitoring is a mechanism for continuously observing the condition of AI systems and related processes in day-to-day operations, detecting early signs of deviations and anomalies, and linking them to remediation. While internal audit is an independent verification conducted at set intervals, monitoring is conducted continuously, primarily by operational teams and second-line functions (security, compliance, the AI Governance Office, and similar functions).

Monitoring involves continuously collecting and analyzing operational logs, performance and fairness metrics, incidents, and user complaints. Visualize results through dashboards, set thresholds for key KPIs (the five areas of business value, trustworthiness, governance, security and privacy, and talent) to automate anomaly detection, and standardize workflows for retraining and remediation in response to detected model drift and degradation in data quality.

Monitoring targets may include, in addition to technical metrics (Accuracy, Recall, False positive rate, Fairness metrics, Model drift, etc.), the rate of go-live deployment of use cases, AIIA implementation rate, number of exception approvals, numbers of incidents and complaints and response lead times, rates of returns in HITL workflows, rate of access permission optimization, training participation rates, and similar indicators. For each metric, define thresholds and escalation criteria. When thresholds are exceeded, notify the CAIO or the AI Governance Office and, where necessary, implement interim measures such as temporary suspension of the use case, configuration changes, or additional reviews.

As a general rule, review monitoring results through the Company-wide AI Steering Committee every quarter and link them to continuous improvement, including policy revisions, adding or streamlining controls, and revising training content.

16.3 Reporting to Senior Management and the Board of Directors

Reporting to senior management and the board of directors is an important means of maintaining a balanced understanding of the value and risks brought by leveraging AI and making decisions aligned with strategy and risk appetite. Based on discussions in the Company-wide AI Steering Committee and the results of monitoring and audits, the CAIO is responsible for compiling regular reports for senior management and the board of directors.

As a general rule, quarterly reporting to senior management and the board of directors should present an overview that includes: a KPI summary across the five areas of business value, trustworthiness, governance, security and privacy, and talent; progress on key use cases; significant incidents and complaints and the status of corrective actions; status of regulatory response and external assessments; status of execution of key risks and mitigation measures; and priority themes for the next period. In particular, significant incidents, adoption or changes of high-risk use cases, and major regulatory changes should be included in reporting, together with alternatives and impact assessments needed for decision-making. Reporting should use a concise and comparable format, and it is desirable to show long-term trends.

For the board of directors, at least once a year, provide opportunities for reporting and discussion from a medium- to long-term perspective on the overall maturity assessment of AI governance (including certification status such as ISO/IEC 42001 and internal audit results), alignment between AI strategy and company-wide strategy, risks relating to impacts on rights and safety, and impacts on societal trust and brand.

Example: Submit a quarterly scorecard (compared with the prior quarter, year-to-date, and against targets) using a fixed format across the five areas of business value, trustworthiness, governance, security/privacy, and talent, with an attached summary of key use cases, incidents, regulatory response, and audit findings.

16.4 External Audit and Assessment

External audits and third-party assessments are effective for ensuring objectivity in AI governance and strengthening trust among customers, regulatory authorities, and society. In alignment with the conformity assessment and certification policy set out in “9. Compliance and Regulatory Response”, the CAIO should establish a policy on which external assessments to undergo for which areas.

External audit and assessment may include, for example, the following:

- audits by third-party certification bodies for management system certifications such as ISO/IEC 42001

- guideline conformity assessments and registration/certification schemes by industry associations or regulatory authorities
- technical assessments by independent expert organizations, such as evaluations of model fairness and robustness and red teaming
- vendor audits and security reviews by key customers

With a view to obtaining certifications such as ISO/IEC 42001, plan gap analyses and remediation. To support accountability to customers and regulatory authorities, appropriately disclose assessment results and remediation status. Integrate items identified through external audits and assessments into internal audit remediation plans, the AI risk register, and reviews of KPI dashboards, and use them to improve policies, standards, and operational processes. For subsequent external assessments, it is desirable to organize evidence and explanations in advance so that remediation status for prior findings can be confirmed.

17. Use Case–Specific Examples of Application

17.1 AI for Recruitment Screening

AI for recruitment screening can have a significant impact on candidates' employment opportunities and careers and should therefore be managed as a high-risk system. Before deployment, conduct an AIIA and a PIA to clarify job relevance, affected groups, and the intended use and scope of AI. From a diversity and inclusion perspective, ensure that the HR function, diversity and inclusion functions, and the legal function participate in the impact assessment.

For the model developer, require provision of model cards and evaluation reports (training data provenance, known biases and limitations, update plans, etc.) and define audit rights, change notifications, data use restrictions, and allocation of responsibilities in contracts. As evaluation indicators, incorporate fairness metrics into operational KPIs in addition to predictive performance, review them at least every quarter, and implement corrective actions when deviations occur. Align thresholds and evaluation criteria with job requirements and applicable laws and regulations, avoid full automation, and require a final decision by humans (HITL).

For candidates, provide advance notice and explanation regarding whether AI is used, its purpose, and its scope of impact, and establish an objection channel and re-review procedures, reasonable accommodations, and alternative options (such as a screening pathway that does not use AI). Ensure rigorous records management by retaining inputs, outputs, decision rationales, and human interventions, and clarifying retention periods, deletion procedures, and confidentiality controls. Enforce data minimization, confirmation of the legality of cross-border transfers, security measures, and prohibition of secondary use.

During operations, monitor indicators suggesting issues—such as disparities in rejection rates and increases in complaints—through dashboards. Where anomalies are identified, take steps such as temporarily suspending the system, changing configurations, requesting re-evaluation by the vendor, or strengthening human oversight. Re-conduct impact assessments when models are updated or intended use is expanded.

The CAIO and the Company-wide AI Steering Committee should manage AI for recruitment screening as a high-risk use case through inventory management and clarify accountability for approval, monitoring, and review.

17.2 Generative AI for Customer Support

When using generative AI for customer support, the main objectives typically include improving response speed, self-service rates, and customer satisfaction. At the same time, risks include misinformation, leakage of confidential information, and insufficient emotional sensitivity. In the AIIA, clarify the channels in scope (chat, email, FAQs, voice assistance, etc.), role allocation between automated responses and human operators, and the scope of information the AI will handle.

Provide transparency notices on AI use and label responses so users can recognize when they are interacting with AI. As safeguards, incorporate red teaming, input validation, avoidance of sensitive topics, prompt injection countermeasures, and other measures to reduce misinformation and safety risks. As a general rule, limit personal data and confidential information to the minimum necessary, and separate confidential documents and personal data from the knowledge base. Preserve logs and interaction records in a centralized manner and clearly specify in terms of use and the privacy policy the scope of use for retraining and improvement.

In operations, aggregate data on complaints, misguidance, escalation rates, resolution rates, and customer satisfaction (CSAT/NPS, etc.) in dashboards, and define remediation measures—such as reviewing conversation flows, updating knowledge, and re-evaluating models—when thresholds are exceeded. As HITL, establish rules to always escalate to a human operator under certain conditions (high-value transactions, cancellations and complaints, consultations related to health or legal advice, etc.).

The CAIO and the AI Governance Office should establish standard operating procedures (SOPs) and training programs for generative AI for customer support and periodically confirm that frontline operators understand AI limitations and how to respond to unexpected behavior.

17.3 AI for Medical Support

AI for medical support is a high-risk use case that can directly affect patients' lives and health. Clarify whether medical device regulation (e.g., the EU MDR / the U.S. FDA) applies and, where applicable, develop and operate the system on the premise of regulatory compliance. Define the intended use and risk class and ensure: clinical evaluation and preparation of technical documentation; obtaining required approvals such as CE marking and 510(k); rigorous quality and risk management (ISO 13485 / ISO 14971); and conformity with medical device software safety requirements (IEC 62304). Establish performance metrics (sensitivity, specificity, PPV/NPV, calibration, etc.), validate using representative data, and conduct fairness evaluations by population group and disease category.

In the operational phase, coordinate with the healthcare provider's safety management arrangements and ethics committee, and continuously collect and evaluate performance metrics, adverse events, and near-miss incidents as part of post-market surveillance. AI recommendations should be limited to medical decision support for physicians. As a general rule, physicians should make the final decision on diagnosis and treatment. Document procedures for human double-checking and manual intervention (ignoring, modifying, or re-evaluating AI recommendations). Where model updates or adaptive retraining are performed, maintain version control and assess impacts of changes, and, where necessary, follow re-approval or recertification processes.

To ensure accountability to patients, clearly explain whether AI is used, its role, limitations, and risks, and explicitly state AI use in informed consent. Clarify available remedies and allocation of responsibilities in the event of misdiagnosis or malfunctions (re-examination, second opinions, complaint channels).

In coordination with the medical function, legal, and compliance, the CAIO should manage AI for medical support as a high-risk system through inventory management and oversee it so that regulatory compliance, risk management, and audit readiness remain aligned with the AIMS overall.

17.4 AI for Critical Infrastructure Operations

AI for critical infrastructure operations is a high-risk use case that directly affects safety, availability, and socioeconomic activity. Strengthen attack resilience, implement redundancy, conduct regular exercises, and clearly define oversight accountability, with "fail-safe" design as a basic principle—i.e., the system should "fail toward safety." Define intended use, system boundaries, and safety constraints, and incorporate separation of IT and OT, defense in depth including Zero Trust, and segregation of duties from the design stage.

In operations, standardize the use of audit logs and SBOM (Software Bill of Materials) or AI-BOM, as well as signed updates, and establish pre- and post-update validation and rollback procedures. To prepare for data poisoning, adversarial attacks, and sensor anomalies, implement consistency checks against physical and process models, fail-safe design, and behavior verification through backup control systems. Limit the scope of autonomous control, and ensure that, in hazardous conditions or when thresholds are exceeded, the system automatically executes a safe shutdown or switches to manual operation.

Depending on the deployment environment, confirm compliance with applicable regulations (e.g., U.S. NERC CIP, Europe's NIS2, the EU AI Act) and clarify coordination with and reporting channels to competent authorities. Ensure log retention and tamper detection, as well as rigorous data minimization, encryption, and access controls. Plan incident response for abnormal situations and business continuity planning (BCP) and disaster preparedness exercises in an integrated manner. Require a final decision by humans and document segregation of duties and accountability boundaries using RACI and similar tools.

In coordination with the CISO, the CIO, the CTO, and business functions, the CAIO should treat AI for critical infrastructure operations as Rights and safety impacting AI subject to approval by the Company-wide AI Steering Committee and continuously review risk levels and investment allocation based on results from the AIIA, resilience exercises, red teaming, external audits, and similar activities.

18. Template Key Points

18.1 AIIA Template

The AIIA should include the items below. Provide completion guidance for each item and include examples to improve the quality and consistency of entries. Require version control upon updates and require recording approvers and dates.

- **Basic Information**
  - Use case name, ID, responsible function and responsible owner
  - Date created, version number, status (draft/approved)
  - Related system/service names, target regions and jurisdictions
- **Use Case Overview**
  - Purpose and expected value (business value, efficiency gains, quality improvements, etc.)
  - Users and stakeholders affected (customers, employees, third parties)
  - Usage contexts (business processes, channels, time periods, etc.)
  - Intended use and non-intended use
- **Positioning Under Regulations and Internal Rules**
  - Regulations and guidelines the use case may fall under (AI Guidelines for Business, the EU AI Act, sector-specific laws and regulations, etc.)
  - Relationship to internal policies (AI policy, security policy, privacy policy, etc.)
  - Internal risk categorization (high/medium/low, whether it qualifies as Rights and safety impacting AI)
- **Overview of Data and Models**
  - Data types used (whether personal data/sensitive information/confidential information is included, logs, external data, etc.)
  - Data sources (in-house systems, vendor-provided data, open data, etc.) and legal basis/licenses
  - Model types (rule-based/statistical models/machine learning/generative AI, etc.) and delivery models (developed in-house/vendor-provided)
  - Links to relevant model cards and datasheets
- **Risk Identification and Assessment**
  - Fields by risk category (e.g., safety and health, fundamental rights and fairness, privacy, security, compliance, reputation, environmental impacts, etc.)
  - Assessment of likelihood and impact for each risk (qualitative and/or quantitative)

    - Examples of assumed scenarios and worst cases
- **Mitigation Measures and Residual Risks**
    - Controls and safeguards for each risk (design constraints, access controls, HITL, testing and monitoring, etc.)
    - Implementation status (planned/implemented), responsible owner, deadlines
    - Assessment of residual risks and judgments on acceptability
- **Operations and Monitoring Design**
    - Whether Human-in-the-loop/Human-on-the-loop is required and specific intervention points
    - Metrics to be monitored (performance, fairness, incidents and complaints, environmental impacts, etc.)
    - Thresholds and alert conditions, escalation contacts, interim measures (suspension of use, configuration changes, etc.)
- **Stakeholder Engagement**
    - Reviewing functions (legal, compliance, security, DPO, HR and labor, labor-management consultative bodies, etc.)
    - Policies for explanation to and consultation with external stakeholders (customer groups, experts, regulatory authorities, etc.), where required
- **Approval and Revision History**
    - Approvers (CAIO, Company-wide AI Steering Committee, relevant executives, etc.)
    - Approval date and effective date
    - Revision history (changes by version, rationale, scope of impact)

18.2 Model Card

The items that should be included in a model card are listed below. Where feasible, separate an internal detailed version from an external-facing summary version to adjust the level of disclosure. Also clarify operational notes and trigger conditions for retraining and updates.

- **Basic Information**
    - Model name, version, and type (classification, regression, generative, recommendation, etc.)
    - Model owner (organization and responsible owner) and contact information
    - Date created, last updated, and status (PoC / production, etc.)

- **Purpose and Use Scenarios**
  - Intended use and intended scope of use (target business processes, target users)
  - Non-intended use (prohibited use examples, non-recommended environments)
  - Business objectives (efficiency gains, quality improvements, risk reduction, etc.)
- **Input and Output Specifications**
  - Input data formats and preprocessing requirements (required fields, units, languages, etc.)
  - Output formats (labels, scores, text, images, etc.) and meanings
  - Dependent external systems and services (APIs, databases, etc.)
- **Overview of Training Data**
  - Main data sources (see "18.3 Datasheets" for details)
  - Data collection period, regions, and target populations
  - Sample size, labeling methods, and annotation quality
- **Evaluation Methods and Results**
  - Evaluation metrics used (Accuracy, Recall, F1, AUC, etc.) and their definitions
  - Test data conditions (representativeness, data splitting methods, cross-validation, etc.)
  - Performance by attribute and segment (including a summary of fairness metrics, where applicable)
- **Known Limitations and Risks**
  - Limitations for data, environments, and users that were not assumed
  - Known biases, error trends, and conditions sensitive to drift
  - Key considerations for use (e.g., prohibition on use as a standalone basis for the final decision)
- **Safety, Fairness, and Explainability**
  - Safety measures (output filtering, rounding toward safety, etc.)
  - Summary of fairness assessment results and whether mitigation measures are in place
  - Whether explainability methods (feature importance, example-based explanations, etc.) are available and how to use them
- **Operations, Monitoring, and Retraining**
  - Metrics to monitor and thresholds (performance, drift, fairness, etc.)
  - Trigger conditions for retraining and tuning (data changes, regulatory changes, performance degradation, etc.)

  - Procedures for model updates (validation, rollback, user notifications, etc.)
- **User Notes**
  - Intended users (experts / general users, etc.) and required prerequisite knowledge
  - Input-related considerations (prohibited inputs, handling of confidential information, etc.)
  - Guidance on interpreting outputs and avoiding misuse
- **Version Control and Links**
  - High-level summary of changes by version
  - Links to relevant datasheets, AIIA, and operational runbooks

18.3 Datasheets

The items that should be included in datasheets are listed below.

- **Basic Information**
  - Dataset name and version
  - Data owner (organization and responsible owner) and contact information
  - Date created, last updated, and start/end dates for use
- **Purpose and Scope of Use**
  - Purpose for creating the dataset (training, testing, evaluation, tuning, etc.)
  - Permitted scope of use (internal-only / whether external provision is permitted / whether reuse is permitted)
  - Related use cases and model names
- **Composition and Scope**
  - Number of records and overview of features/fields
  - Target period, target regions, and target populations (age groups, industries, product lines, etc.)
  - Data formats (structured/unstructured; text/images/audio, etc.)
- **Sources and Collection Methods**
  - Data sources (in-house systems, customer-provided data, public data, vendor-provided data, etc.)
  - Collection methods (log collection, surveys, crawling, etc.) and frequency
  - Whether consent and/or notices were provided for data collection
- **Quality and Preprocessing**
  - Status of missing values and outliers, and handling policies

  - Details of preprocessing, transformations, normalization, anonymization, and similar steps
  - Labeling/annotation methods and quality management
- **Representativeness and Bias**
  - Alignment with intended usage contexts and target populations (representativeness assessment)
  - Whether there is under- or over-representation of specific attributes (e.g., gender, age, region)
  - Summary of bias analyses conducted and identified biases and mitigation measures
- **Privacy and Legal Basis**
  - Whether personal data or sensitive information is included, and types
  - Legal basis for processing (consent, contracts, legitimate interests, etc.) and alignment with the privacy policy
  - Details of privacy protection measures (anonymization, pseudonymization, differential privacy, etc.)
- **Licenses, Rights, and Reuse Conditions**
  - License types for included third-party data/content (OSS licenses, commercial licenses, etc.)
  - Use conditions (whether commercial use is permitted, whether redistribution is permitted, attribution requirements, etc.)
  - Conditions for reuse and secondary use of the dataset (internal projects, external provision, etc.)
- **Security and Access Controls**
  - Storage location (data centers / cloud / on-premises, etc.) and whether encryption is used
  - Access permissions (view/edit/export permissions) and authorization processes
  - Policies for log collection and audits
- **Retention and Disposal**
  - Retention periods and review timing
  - Disposal/archiving methods (secure deletion, archiving after anonymization, etc.)
  - Whether automatic deletion and alerts are in place after retention periods elapse
- **Data Lineage and Diagrams**
  - Processing paths from main source systems to the dataset
  - Relationship with derived datasets and downstream models

  - Links to lineage diagrams (figures/tables)

18.4 Procurement Checklist

The items that should be included in a procurement checklist are listed below. Clarify vendor transparency and cooperation obligations and ensure that information provision required for post-contract operations is secured. It is also advisable for the checklist to include, at a minimum, columns such as “item”, “sample questions”, “expected answer”, “verification result”, “risk assessment”, and “response policy.”

- **Basic Information and Scope**
  - Target service/product name and version
  - Delivery model (cloud / on-premises / hybrid, etc.)
  - Intended use cases and risk categorization (high/medium/low)
- **Technical Conformity**
  - Alignment with required functional and performance requirements
  - Integration with existing systems and scalability
  - Roadmap and support model and SLAs
- **Transparency Documentation**
  - Whether model cards, datasheets, AIIA, and similar materials can be provided
  - Disclosure of known limitations, bias, and risks
  - Notification methods for updates and changes
- **Data Handling and Privacy**
  - Purpose of use and retention periods for training, testing, and log data
  - Whether and under what conditions customer data may be used for secondary use (including improving models for other customers)
  - Handling of personal data, sensitive information, and confidential information; whether anonymization, encryption, and cross-border transfers occur
- **Security**
  - Certification status (ISO/IEC 27001, etc.) and security policies
  - Vulnerability management and incident response processes
  - Whether the vendor can cooperate with attack resilience assessments (penetration testing, red teaming, etc.)
- **Compliance and Regulatory Alignment**
  - Status of alignment with the AI Guidelines for Business and legal and regulatory requirements (such as the EU AI Act)

  - Consistency with sector-specific laws and regulations and platform policies
  - Whether conformity assessment and certification are in place (e.g., for high-risk AI)
- **Fairness and Explainability**
  - Status of bias assessment and fairness assessment and whether results can be provided
  - Whether explainability features (mechanisms to show rationale for outputs, etc.) are available
  - User-facing notices and usage guidelines
- **SLAs and Continuity**
  - SLAs for availability, response time, support response time, and similar items
  - Recovery time objectives in the event of disasters or outages, and BCP/DR plans
  - Notifications and grace periods for service discontinuation or functional changes
- **Intellectual Property and Licenses**
  - Ownership of rights for training data, models, and generated outputs
  - Whether and to what extent indemnification is provided for third-party rights infringement
  - Scope of permitted use (commercial use, resale, redistribution, etc.)
- **Monitoring, Audit, and Cooperation Obligations**
  - Scope of log and report provision
  - Whether the vendor can support customer audits and security reviews
  - Cooperation obligations for regulatory authorities and customer audits
- **Vendor Lock-In, Termination, and Transition**
  - Whether data, models, and configuration information can be exported, and in what formats
  - Content, cost, and duration of transition support
  - Consultation processes and the right to terminate for price revisions or changes to contractual terms
- **Environment and Sustainability**
  - Disclosure of information on energy consumption and CO2 emissions
  - Data center environmental certifications and efficiency metrics
  - Efforts to reduce environmental impacts

19. Notes

This manual provides general practical guidance and is not intended as legal advice in any specific jurisdiction. Determinations regarding compliance with specific laws and regulations should be made under the oversight of the organization’s legal and compliance functions, supported, where appropriate, by external expert advice.

Organizations should tailor this manual in light of their risk tolerance, regulatory environment, business priorities, organizational size, and industry characteristics, and operate it after defining scope of application and priorities. As the accountable executive for AI governance, the CAIO should lead the use of this manual. However, it should be noted that ultimate management accountability rests with the board of directors and the senior management team as a whole, and that business owners are responsible for business decisions regarding individual use cases. It is also desirable to ensure integration so that this manual does not conflict with existing frameworks for information security, internal control frameworks, quality management, and similar areas.

In response to technological evolution and regulatory change, continuously updating policies, standards, and processes and strengthening organizational maturity is key to long-term competitiveness. This manual should also be maintained as a “Living Document”, rather than treated as a fixed “final version”, with periodic reviews led by the CAIO and the Company-wide AI Steering Committee, informed by results from internal audits, monitoring, and external assessments.

20. Conclusion

This manual presents a practical pathway for organizations, through the CAIO role, to achieve both AI value creation and responsible AI use. In practice, it is important not to apply the structures and processes described here mechanically, but to adapt them into an effective form in light of each organization’s challenges and constraints, through dialogue among stakeholders and iterative trial and error.

Establishing the CAIO function and strengthening AI governance is not a one-time project that ends upon completion. Rather, it is a long-term effort that should be continuously reviewed in response to changes in technology, regulation, and society. It is hoped that this manual will serve as a common language for internal discussion and consensus building around such efforts.

Appendix

A. Reference Frameworks

International standards and existing frameworks provide a common language and a primary reference framework for the CAIO when designing the overall architecture of AI governance. The frameworks and international standards listed below can serve as references for building AI governance, including the CAIO function. This section summarizes the overview and key characteristics of each framework. Policies for application within an organization and concrete approaches to legal compliance are addressed in “9. Compliance and Regulatory Response.”

A.1 AI Guidelines for Business (Japan)

Practical guidance for translating domestic AI principles into implementation. As non-binding soft law, it provides integrated, risk-based guidance covering the full lifecycle of AI development, provision, and use. It sets out common principles such as human-centeredness, safety, fairness, privacy, security, transparency, and accountability, and includes considerations by actor category (AI developers, AI providers, AI users), guidance for advanced AI in line with the Hiroshima AI Process, and agile governance.

This manual also uses the AI Guidelines for Business as a primary reference framework for implementing AI principles. The actor-based structure also provides a foundation for the CAIO when designing internal role allocation.

A.2 NIST AI Risk Management Framework (United States)

A voluntary and cross-industry framework for managing risks while realizing the benefits of AI. It defines seven characteristics of trustworthy AI—valid and reliable, safe, secure and resilient, explainable and interpretable, transparent and accountable, privacy-enhanced, and fair with managed harmful bias—and structures risk management into four iterative functions: Govern, Map, Measure, and Manage. It emphasizes sociotechnical TEVV (testing, evaluation, verification, and validation) and recommends: clearly defined roles and responsibilities; coverage across the full lifecycle; continuous monitoring; incident response; documentation; supply chain due diligence; human oversight; and development and use of context-specific profiles.

The CAIO can design an overall AI risk management architecture by centering on the Govern function and integrating the Map/Measure/Manage functions into the organization’s risk management processes (see “5.3 Risk Management” and “15. Key Risks and Mitigation Measures”).

A.3 EU AI Act (Europe)

A risk-based framework that governs the use of AI. Uses that threaten human rights (e.g., social scoring and indiscriminate biometric surveillance) are prohibited. For high-risk AI in critical domains, strict requirements apply, including compliance with risk management, data quality, technical documentation, cybersecurity, transparency, human oversight, conformity assessment, and CE marking. Additional obligations apply to general-purpose and foundation models, and even stricter regulation applies to those posing systemic risks. By mapping the organization’s AI use cases to the EU AI Act’s risk categories in an aligned manner and clarifying the correspondence with Rights and safety impacting AI (see “5.8 Additional Internal Controls for Rights and safety Impacting AI”), the CAIO can establish a unified baseline for internal controls.

A.4 OMB Memorandum M-24-10 (United States)

This memorandum sets out a government-wide policy to advance AI governance, innovation, and risk management in the United States. Agencies are to designate a CAIO, and agencies covered by the CFO Act are to establish an AI governance committee, publish AI use cases annually, and develop strategies to remove barriers related to infrastructure, data, security, and talent. For AI that impacts safety or rights, it sets minimum requirements, including impact assessments, operational testing, independent review, human oversight, ensuring fairness, and baseline standards for appeals and opt-out. It also provides procurement guidance.

While the memorandum is intended for public sector bodies, private sector organizations may also draw lessons from elements such as: (1) designating a CAIO, (2) establishing a company-wide AI governance committee, and (3) setting common minimum standards for AI that impacts safety or rights.

A.5 ISO/IEC 42001

ISO/IEC 42001 is a certification framework for an AI management system (AIMS) and provides guidance for operating an AI-specific management system through the PDCA cycle. In Plan, define scope, policies, objectives and KPIs, conduct risk assessments, and select controls. In Do, operate processes, deliver training, and maintain documentation and records. In Check, assess conformity through internal audits and management reviews. In Act, implement corrective and preventive actions and drive improvement plans to strengthen maturity. This helps ensure that governance does not become a formality and is continuously improved. The standard is structured to integrate readily with existing management systems such as an ISMS and enables AI-specific controls to be embedded into existing risk management processes.

It is desirable for the CAIO to position ISO/IEC 42001 as the “operating framework” for company-wide AI governance and to design an AIMS that is consistent with the roles and responsibilities defined in “5. Roles and Responsibilities” and the monitoring and reporting set out in “16. Audit, Monitoring, and Reporting.”

A.6 ISO/IEC 23894

ISO/IEC 23894 systematizes AI risk identification, analysis, evaluation, and treatment. In risk identification, it supports structured identification of threats, vulnerabilities, scope of impacts, and stakeholders. In risk analysis, it uses qualitative and quantitative assessments and scenario analysis to measure impacts. In risk evaluation, it sets priorities by comparison with acceptable levels. In risk treatment, it selects appropriate options among avoidance, mitigation, acceptance, and transfer, balancing costs and effectiveness. In designing the AIIA process, referencing this standard’s framework for risk identification, analysis, evaluation, and treatment also helps prevent gaps in assessment.

It is recommended that the CAIO adopt a risk management process based on ISO/IEC 23894 as a common format for company-wide AI risks and standardize the risk register and prioritization criteria.

B. Example First-Year Implementation Plan

The first year immediately after a CAIO assumes the role is a period that can significantly influence how well AI governance becomes embedded within the organization. It is therefore necessary to proceed with implementation carefully. This section explains, along a timeline, the order in which an organization that has newly established CAIO should take action and what it should prioritize.

B.1 Days 0–90: Organizational Setup and Current-State Assessment

During the first 90 days, the goal is to document and communicate, across the organization, the decision-making and authority framework—i.e., “who can decide what, and to what extent”—and to establish company-wide AI principles and a current-state assessment (an inventory). This provides a common baseline for subsequent design, procurement, and operations, as well as a clear view of the priority risks and opportunities to address.

- CAIO designation; definition of RACI for authority and responsibilities; and development of a Charter
- Drafting AI principles, prohibited uses, and guardrails; obtaining management approval; and announcing them internally as interim guidelines
- Establishing the Company-wide AI Steering Committee; determining its members, meeting frequency, and operating principles for agendas and minutes
- Taking stock of current use cases, models, data, and vendors (AI inventory v1) and assigning an initial risk categorization (high/medium/low)
- Conducting regulation mapping (personal data, copyright, sector-specific laws and regulations, the EU AI Act, etc.) and initiating development of an initial risk register
- Developing an internal awareness plan (internal portal site, FAQs, help desk channels) and designing mechanisms for handling “complaints and appeals”
- Collecting candidate use cases that could deliver quick wins through help desk channels and internal surveys

[Deliverables]

RACI matrix and Charter, draft AI principles, inventory v1 (with initial risk categorization), risk register v1, and a list of candidates quick-win use cases

B.2 Days 91–120: Standardization and Pilot Design

The objective of this period is to establish standard templates—such as AIIA/PIA and model cards—and risk-based approval gates, so that reviews can be conducted consistently without hesitation. At the same time, select representative pilot initiatives and design evaluation dimensions (KPI/ROI and legal compliance, safety, and fairness).

- Finalize AIIA/PIA templates, model cards and datasheets, and design guidance for HITL
- Define model risk categorization (high/medium/low), and define approval gate structures (Go/No-Go) and required approvers by category
- Establish standards for explanation, notices, and remediation (user notice templates, appeals flows, re-review SLA)
- Review procurement clauses (performance and limitations disclosure, data handling, continuous improvement, audit rights, environmental considerations, SBOM, etc.) and confirm alignment with "12. Procurement and Vendor Management" and "18.4 Procurement Checklist"
- Select pilot candidates and set success criteria (KPI/ROI, legal compliance, safety, and fairness)
- Finalize responsible owners (RACI) and schedules for each pilot initiative

[Deliverables]

A complete set of standard templates; an approval process diagram (gate definitions by risk category); draft procurement clauses; and a pilot plan (with use case–specific RACI and KPIs)

B.3 Days 121–180: Pilot Execution and Foundational Enablement

During this period, run the selected pilots to put end-to-end processes into operation—such as AIIA, HITL, and the complaint handling SLA—while introducing a minimum viable MLOps and governance foundation. The objective is to validate the "minimum scalable operating unit."

- Initial rollout of an MLOps/governance foundation (model registry, experiment tracking, data catalog, access controls, and similar components, starting with what is required to operate the pilots)
- Building an initial dashboard for fairness and robustness assessments, red teaming, and drift monitoring

- Designing audit logs and beginning collection; establishing and validating an incident response runbook; and developing and testing kill switch and rollback procedures
- Operational testing, within pilot initiatives, of user notice and consent, HITL workflows, and the complaint handling SLA
- Reviewing pilot outcomes (both value and risk) and clarifying decision criteria for "go-live / improve / discontinue"

[Deliverables]

Pilot evaluation reports (covering both value and risk); model cards and datasheets (for the pilots); operational runbooks (including incident response and kill switch procedures); and an initial dashboard

B.4 Days 181–210: Talent Development and Institutionalization

During this period, link role definitions, skill standards, and training programs to clarify "who should be able to do what, and to what level", and aim to incorporate them into evaluation and certification. This is the phase in which the approach described in "13. Training and Talent" is translated into specific job categories, training programs, and evaluation systems.

- Define roles and job categories by role (ML engineers, MLOps, risk/compliance, PMs, data management, product owners, etc.)
- Identify required skills and expected proficiency levels for each role (skill maps) and design methods for proficiency assessment
- Launch training curricula (fairness, security, privacy, regulatory response, explainability, etc.) for both onboarding of new personnel and practical, on-the-job application
- Introduce an internal certification program (to make skills at or above a certain level visible) and link it to evaluation and compensation systems
- Develop and publish, on the internal portal site, a set of best practices and patterns (safe design, notice templates, HITL patterns, sample AIIA entries, etc.)

[Deliverables]

Training plans and participation records; role and skill maps; internal certification program guidelines; and a best-practices repository (an updatable knowledge base)

B.5 Days 211–240 (Evaluation, Remediation, and Standards Readiness)

During this period, the organization evaluates initial operations against international standards and remediates gaps, marking the start of running the PDCA cycle for continuous improvement. Use ISO/IEC 42001 and ISO/IEC 23894 as “checklists” regardless of whether certification will be pursued.

- Finalize the scope for ISO/IEC 42001 and conduct a gap analysis; develop a remediation plan (even if certification is not planned in the near term, reference the requirements as internal standards)
- Document the risk assessment process based on ISO/IEC 23894 and apply it to representative use cases
- Begin KPI operations (AIIA completion rate, complaint handling SLA compliance rate, model approval lead time, fairness metrics, number of incidents, etc.)
- Establish a regular cadence for internal audits (a simplified version is acceptable) and begin operating corrective and preventive actions (CAPA) and change management
- Review evaluation results through the Company-wide AI Steering Committee and decide high-priority remediation themes

[Deliverables]

Gap analysis report; remediation plan; initial KPI dashboard; and an internal audit plan and CAPA records

B.6 Days 241–270: Scaling, Audits, and Exercises

During this period, gradually scale reviewed use cases and validate external dependencies and incident response capability through vendor audits and red team exercises. The objective is to demonstrate operations that remain robust even as they scale.

- Develop and execute a phased rollout plan for use cases that have passed approval gates. For high-risk use cases in particular, explicitly define steps such as shadow operation → limited rollout → full rollout.
- Conduct vendor audits and SLA reviews and validate secure updates for software and related components (signing, SBOM) and the supply chain risk management process.
- Confirm application of Zero Trust and segmentation and test countermeasures against prompt injection and data poisoning.

- Conduct incident response training through tabletop exercises and red team exercises and perform practical verification of regulatory authority and internal reporting channels.
- Integrate findings from audits and exercises into CAPA (Corrective Action Preventive Action) and develop a draft annual plan to continue these activities into the next fiscal year and beyond.

[Deliverables]

Rollout plan and progress records; vendor audit report; exercise reports and corrective action records; and a summary of supply chain management and Zero Trust implementation status

B.7 Days 271–360: Certification Readiness and Next-Year Planning

In the final quarter, the organization summarizes first-year outcomes and the status of risk response, advances preparations for external certification where appropriate, and aligns with management on next-year investment and governance structure. This phase transitions the mechanisms built in the first year from a “one-off project” to a “continuous management system.”

- Prepare documentation, records, and evidence for external certification readiness (e.g., ISO/IEC 42001). Even if certification is not pursued, organize documentation to a level that can withstand external assessment.
- Conduct management reviews and establish a management approval process for residual risks.
- Decide the external disclosure policy for a transparency report (intended purpose, performance/limitations, fairness assessment, update history, incident overview, etc.).
- Summarize ROI and risks and develop an improvement roadmap (carry-over issues from the first year, priority areas for year two, required budget, staffing, and external partners).
- Retire models that are no longer needed, identify models to be improved, and develop response plans (reflected as model inventory v2).
- Update schedules for continuing education and recertification (define required training hours and content annually and by role).

[Deliverables]

A certification readiness package (or a set of documents for external assessment); management review records and residual risk approval documentation; a draft transparency report; an annual roadmap (investment and governance structure plan); model inventory v2; and a continuing education and recertification plan