



Guidelines for Third-Party Compliance with Responsible Artificial Intelligence (AI) Principles

AIS Group

Version: 2.1

Document Owner: AI Governance

Revision Date: 06 July 2026

Revision History

All revisions to this document are recorded in the table below.

Version	Date	Prepared by	Description	Reviewed by	Approved by
1.0	19 December 2025	AI Governance	Initial document	Intelligent Process Automation Manager	CoE
2.0	20 April 2026	AI Governance	Added: - Intellectual property rights and use of AI technology	Intelligent Process Automation Manager	CoE
2.1	06 July 2026	AI Governance	Updated terminology and adjusted content layout.	Intelligent Process Automation Manager	CoE

Table of Contents

1. INTRODUCTION	4
1.1 PURPOSE	4
1.2 SCOPE	4
1.3 DEFINITIONS	4
2. RESPONSIBLE ARTIFICIAL INTELLIGENCE (AI) PRINCIPLES	5
2.1 FAIRNESS	5
2.2 HUMAN AGENCY AND OVERSIGHT	5
2.3 PRIVACY AND SECURITY	5
2.4 SAFETY AND ROBUSTNESS	6
2.5 TRANSPARENCY AND EXPLAINABILITY	6
2.6 ACCOUNTABILITY	6
2.7 ENVIRONMENTAL IMPACT	6
3. SAMPLE RESPONSIBLE AI DOCUMENTATION FOR THIRD-PARTY VENDOR ASSESSMENT	7
4. INTELLECTUAL PROPERTY RIGHTS AND USE OF AI TECHNOLOGY	9
4.1 OWNERSHIP OF DELIVERABLES AND INTELLECTUAL PROPERTY RIGHTS	9
4.2 USE OF THE PRINCIPAL’S DATA SOLELY FOR PROJECT PURPOSES	10
4.3 NON-INFRINGEMENT OF THIRD-PARTY INTELLECTUAL PROPERTY AND CONTRACTOR LIABILITY ..	10

1. Introduction

1.1 Purpose

This AI Ethics Principles document is intended to serve as a clear standard and practical guideline for contractors and third parties in designing, developing, and operating artificial intelligence (AI) systems. Its primary objective is to ensure alignment with the Company's ethical, social, and legal principles, while establishing a shared understanding of key expectations regarding fairness, accountability, safety, sustainability, and privacy. Compliance with these principles serves as a mechanism to mitigate risks that may arise from negligence or inappropriate use of AI systems.

1.2 Scope

This document applies to all work performed by contractors and third parties that involves the organization's artificial intelligence (AI) systems. The scope covers the end-to-end project lifecycle, including design, development, testing, deployment, and maintenance.

1.3 Definitions

No	Term	Definition
1.	Third-Party	An external company or service provider engaged to develop, procure, or provide services related to artificial intelligence (AI) systems, and required to comply with the organization's Responsible AI principles.
2.	Responsible AI	An approach to developing and using artificial intelligence (AI) systems in an ethical, transparent, secure, fair, and auditable manner.
3.	Bias	Prejudice or imbalance in data or algorithms that may affect the fairness of an artificial intelligence (AI) system.
4.	Explainability	The ability to explain the rationale behind decisions or outputs generated by an artificial intelligence (AI) system.
5.	Human-in-the-Loop (HITL)	Human involvement in the decision-making process of an artificial intelligence (AI) system in order to control and reduce risks.
6.	Privacy	The protection of personal data and user confidentiality within artificial intelligence (AI) systems.
7.	Transparency	The disclosure of information, processes, and operating methods of artificial intelligence (AI) systems in a manner that enables review and auditability.

2. Responsible Artificial Intelligence (AI) Principles

To ensure that the development or use of artificial intelligence (AI) systems performed for the organization meets the highest standards of ethics, trustworthiness, and responsibility, third parties must adhere to and comply with the following seven principles.

2.1 Fairness

The artificial intelligence (AI) systems that you develop or operate must strictly ensure fairness, with clear mechanisms to prevent discrimination against any individual or group. The primary objective is to avoid decisions or inferences that create adverse or biased outcomes for certain user groups. Systems must be designed and developed with due regard for equality and equitable treatment, so that all user groups can access and benefit from AI technology fairly. Building fair systems not only reduces potential inequality but also fosters long-term trust and confidence in the responsible use of technology.

2.2 Human Agency and Oversight

Artificial intelligence (AI) systems should be designed and deployed in a manner that supports human decision-making, allowing humans to appropriately control and oversee the system. This is intended to prevent negative impacts on individuals and society and to promote fundamental values and human rights. Over-reliance on AI systems should be mitigated, particularly in contexts involving life, safety, and human dignity.

Additional oversight requirements: You must define an appropriate level of human involvement based on the risk level of the system.

Human-in-the-Loop (HITL): Applied to high-risk systems or systems with severe impact on individuals, requiring human intervention or final human decision-making whenever the system provides recommendations or decisions.

Human-on-the-Loop (HOTL): Applied to medium- to low-risk systems, requiring humans to periodically monitor and oversee system operations, with clear alert mechanisms that allow humans to intervene or correct errors or unexpected outcomes.

2.3 Privacy and Security

Artificial intelligence (AI) systems should be designed and developed with privacy and information security as key considerations, in order to prevent violations of individual rights and protect data from unauthorized access. They must also be aligned with clear and beneficial purposes for users and society, taking into account ethical, social, and legal impacts. This ensures that AI systems create maximum benefit for humans and society while reducing potential risks and negative impacts.

2.4 Safety and Robustness

Artificial intelligence (AI) systems should be designed and deployed with confidence in their safety, with the ability to prevent and reduce risks that may arise in different usage scenarios.

AI systems should be robust; that is, they should be able to avoid operational issues or vulnerabilities when deployed. Robustness is essential to the development of trustworthy AI systems. AI systems should be able to withstand attacks and operate as expected under various conditions without interruption or defects. In addition, AI systems should operate efficiently within specified timeframes and use resources effectively.

2.5 Transparency and Explainability

Artificial intelligence (AI) systems should be designed to provide clear, accurate, and accessible information about their processes, decisions, and actions, including disclosure of data sources, algorithms, and methods used, so that users can understand how AI operates. AI systems must be capable of explaining their reasoning, outputs, and decision-making processes in an understandable manner. This includes providing information on how conclusions are derived, enabling users to understand and trust the outputs of AI systems.

2.6 Accountability

Developers and users of artificial intelligence (AI) systems must be accountable for the outcomes resulting from the use of such systems.

This accountability includes providing clear, accurate, and accessible information regarding processes, decisions, operations, and oversight.

AI systems should have clear governance, with defined structures and responsibilities that support the use of AI in accordance with fairness and effectiveness. Such governance includes review, monitoring, and continuous improvement to ensure that AI operates as expected.

2.7 Environmental Impact

Artificial intelligence (AI) systems should be designed and developed with emphasis on reducing environmental impact throughout all stages of system development, including design, model and algorithm selection, deployment, and operational management.

3. Sample Responsible AI Documentation for Third-Party Vendor Assessment

This list of documents serves as guidance for vendor assessment and does not constitute a mandatory requirement to submit every item listed. Vendors may provide equivalent documentation in other formats, provided that such documentation demonstrates compliance with the organization's Responsible AI principles.

Principle	Document List
1. Fairness	Fairness Metrics & Thresholds: Clearly specify the metrics and thresholds used to assess model fairness, including the rationale for selecting such metrics. Dataset Card: Detailed information on datasets used for model training and testing, such as data sources, demographic statistics, limitations, collection methods, and handling of sensitive groups. Bias Mitigation Strategy: Methods and results of mitigating detected bias, including any residual bias remaining after mitigation. Post-Deployment Fairness Monitoring Plan: A plan to monitor and assess bias after production deployment, including drift detection methods, assessment frequency, and corrective action criteria when unfairness is identified.
2. Human Agency and Oversight	Risk Classification & HITL/HOTL Plan: Classification of the risk level of the artificial intelligence (AI) system and a plan identifying points where humans must intervene (Human-in-the-Loop) or monitor and control outputs (Human-on-the-Loop), including the rationale for the oversight design. Formal Oversight Sign-off: Records of approval, review, or final decision-making by a human controller for cases involving high risk or significant impact on users or the organization. Oversight Runbook: An operational manual specifying procedures, review methods, warning signs to monitor, and decision-making guidance when an artificial intelligence (AI) system exhibits abnormal behavior or uncertain outputs. Escalation Path & Criteria: Criteria and escalation paths for situations where reviewers identify high-risk or uncertain conditions, including defined points for escalation to specialists, management, or temporary system suspension.
3. Privacy and Security	Access Control Matrix: A table specifying access rights to training data, validation data, and live data, as well as access rights to models, model versions, and related systems, together with the access control principles and roles of each user group.

Principle	Document List
	AI-Specific Security Test Report: A security testing report specifically designed for artificial intelligence (AI) systems, such as assessments of risks related to model inversion, membership inference, model extraction, data poisoning, and adversarial attacks, including a summary of identified vulnerabilities and remediation measures. Data Quality Assessment: A report assessing the completeness, accuracy, consistency, timeliness, and cleanliness of data used throughout all stages of the model development process, including limitations and potential impacts of data on system performance or risks.
4. Safety and Robustness	Performance Testing Report: Results of accuracy and performance testing, including metrics such as Accuracy, F1-Score, and AUC, under normal and challenging conditions, together with an assessment of impacts when data is inaccurate or incomplete. Operational Design Domain (ODD): A document specifying the scope and operating conditions for which the artificial intelligence (AI) system has been designed and tested to operate safely, such as supported data types, operating environment, limitations to be observed, and conditions under which the system should not be used. Adversarial Testing Report: A test report assessing the system's ability to withstand attacks that may cause errors, including adversarial inputs, model deception, signal manipulation, and remediation results for identified vulnerabilities. Stress & Load Robustness Test: Results of robustness testing when the system is subject to high processing loads or unusually large request volumes, including model behavior under resource constraints or high latency.
5. Transparency and Explainability	Model Card: Technical details of the model, including its purpose, limitations, performance metrics, appropriate conditions of use, and factors that may cause inaccurate model outputs. User Disclosure Statement: A clear notice informing users that they are interacting with an artificial intelligence (AI) system, including the level of automation relied upon for outputs and any precautions users should be aware of. Explainability Report: A report explaining the model's decision-making rationale by presenting key factors, such as feature importance or Explainable AI (XAI) methods, and examples of output explanations for high-risk cases or cases that may affect individuals.
6. Accountability	Responsible AI Governance Policy: The vendor's Responsible AI governance policy or documentation confirming compliance with the principal organization's Responsible AI policy, including the governance framework, roles, and decision-making processes related to artificial intelligence (AI) systems.

Principle	Document List
	RACI Matrix (for AI-related functions): A matrix specifying responsibilities at each stage of the artificial intelligence (AI) system lifecycle, clearly identifying who is Responsible, Accountable, Consulted, and Informed. Audit Trail & Evidence Index: Records of evidence and key events occurring during the development, testing, improvement, and operation of artificial intelligence (AI) systems, organized as an index to support transparent and complete traceability. Incident Response Plan: A response plan for AI-related risk events or system errors, covering detection methods, severity assessment, incident notification, remediation, and prevention of recurrence.
7. Environmental Impact	Energy & Carbon Report: A report on energy consumption and carbon emissions resulting from model training, testing, and operation, specifying the calculation method for resources used and the overall environmental impact. Efficiency Optimization Plan: A plan to optimize model resource efficiency, such as model size reduction, architectural optimization, or the use of efficiency-enhancing techniques to reduce energy consumption and processing costs. Hardware Efficiency Disclosure: Information on hardware used, such as GPU/TPU type, performance per watt, and factors affecting energy consumption, to enable assessment of the environmental efficiency of infrastructure options.

4. Intellectual Property Rights and Use of AI Technology

To protect intellectual property interests and ensure the principal's full ability to utilize project deliverables, the preparation, development, and delivery of work under projects involving artificial intelligence (AI) technology must be governed by principles relating to intellectual property and lawful use of AI systems. The following principles serve as a reference framework for contractual requirements and for the consideration, interpretation, and oversight of project execution throughout the term of the agreement.

4.1 Ownership of Deliverables and Intellectual Property Rights

All deliverables arising from the project, whether resulting from human development, automated systems, or artificial intelligence (AI) technology at any stage, shall be deemed work specifically created for the principal and shall be the exclusive intellectual property of the principal. This principle is intended to ensure that the principal may fully use, reproduce, modify, further develop, publish, or commercially exploit the deliverables without restriction arising from the rights or obligations of any other party. Any use of pre-existing assets must not affect the completeness of the principal's rights in the delivered work.

4.2 Use of the Principal's Data Solely for Project Purposes

Data, prompts, and outputs generated in the course of project execution shall be regarded as assets under the control of the principal. This principle requires that the use of such data be strictly limited to the purposes of the project in order to protect the principal's confidential business information and intellectual property rights. Such data or outputs may not be used to train, improve, or develop any other AI model, whether owned by the contractor or a third party, unless prior written authorization has been obtained from the principal. This principle is intended to prevent out-of-scope use of data and reduce long-term legal risks.

4.3 Non-Infringement of Third-Party Intellectual Property and Contractor Liability

Project execution must not involve the use of any work, data, model, software, content, or intellectual property belonging to another party without lawful authorization. This principle requires the contractor to carefully verify the legality of all resources, technologies, and content used and to provide evidence of usage rights upon request. If it is found, or there is reasonable suspicion, that any deliverable under the project infringes the rights of another party, the contractor shall be responsible for all damages, impacts, and legal liabilities arising therefrom in order to protect the principal from potential disputes and risks.