SUSTAINABLEIT.ORG

Responsible AI Data Governance Principles and Runbook

A Practical Guide

IT Toolkit for Responsible & Sustainable AI: A Field Guide for Implementation at Scale

Responsible AI can only be achieved if governance extends to, and indeed prioritizes, the ethics and sustainability of the data that powers these systems.

As AI rapidly infiltrates the core engines of business transformation and decision-making, it is critical to guarantee that its data is as accurate, purposeful, ethically sourced, and environmentally managed as possible.

Without such oversight, organizations risk building AI on an opaque, biased, and inefficient foundation that will compromise performance, ROI, trust, compliance, and the environment.

A responsible and sustainable AI portfolio requires disciplined stewardship of data quality, transparency of data sources, and safeguards against bias and misuse. While designing data architectures for scalability, it also avoids unnecessary data duplication, which mitigates energy-intensive storage needs. By defining and tracking data purpose and intended usage, organizations also ensure their AI models remain auditable, explainable, and aligned with sustainability goals.

This Runbook introduces five foundational principles—Validity & Reliability, Transparency & Explainability, Fit for Purpose & Scalability, Environmental Sustainability, and Accountability—that collectively guide IT and business leaders in embedding ESG-aligned stewardship into AI data practices. It also provides step-by-step actions, KPIs, and documentation suggestions for operationalizing the principles in AI data governance. This document, part of the IT Toolkit for Responsible and Sustainable AI, was developed and vetted by IT leaders and AI experts volunteering time in SustainableIT.org’s 50+ member Responsible AI Working Group.

This report is only one element of the **IT Toolkit for Responsible and Sustainable AI** from **SustainableIT.org.** Other resources include:

- A Toolkit overview that makes the case for responsible and sustainable IT and identifies the people who contributed to the resources
- A Sustainability Runbook to operationalize climate- and resource-related governance
- Data Governance Principles and Runbook focused on AI data quality, ethics, and compliance
- A Responsible AI Governance Lifecycle Model mapping critical principles and actions from development to post-deployment
- A comprehensive AI glossary and reference guide to global standards, frameworks and tools

Format

This report begins by presenting a set of AI Data Principles and Goals for responsible AI data governance. Principles provide a common framework for governance, enabling consistent oversight, risk mitigation, and responsible innovation. A hypothetical example of practical application is included with each principle.

The balance of this guide, the "Runbook" portion, helps IT leaders and their business peers operationalize the AI data governance principles through step-by-step guidance. The Runbook is organized in two main parts: Preparation and Principles Operationalization.

Part 1, Preparation, addresses establishment of governance roles and responsibilities and the prerequisites of processes and capabilities fundamental to data management and governance.

Part 2, Principles Operationalization, provides checklists for actions, key performance indicators, documentation, and tools. Each of these checklists are grouped into development and implementation sections.

Future iterations of this document will evolve alongside the corporate sustainability, AI and data policy landscape, with periodic reviews to align with emerging best practices and stakeholder expectations. We welcome feedback and suggestions at **info@sustainableIT.org**.

AI Data Principles and Goals

A principle in AI governance is a foundational guideline that establishes expectations, boundaries, and values for the responsible development and use of AI technology. Principles serve as anchors for decision-making, aligning AI data governance practices with legal, ethical, operational, and environmental objectives.

Without well-defined principles, AI systems may operate in ways that are opaque, biased, or misaligned with corporate and regulatory expectations, increasing risk both from a security and brand perspective as well as undermining public trust. Principles provide a common framework for governance, enabling consistent oversight, risk mitigation, and responsible innovation.

A hypothetical example of practical application is included with each principle.

1. **Validity & Reliability**

   Data used for AI systems will be accurate, consistent, and complete, with processes in place to identify and minimize errors, gaps, and bias – especially those that may impact protected groups (e.g., race, gender, age, or disability). Data must be relevant for its intended purpose, ensuring reliability(i.e., data yields consistent results across time, contexts, and applications).

   Example: If an AI hiring tool is trained on historical employee data that underrepresents women in technical roles, the system may perpetuate bias in hiring recommendations. Ensuring validity and reliability would involve identifying this imbalance and correcting it through rebalancing or supplementary data.

2. **Transparency & Explainability**

   The organization will maintain clear, traceable records of where data comes from, how it is collected, processed, and used—including any transformations, filtering, or enrichment. Explainability means designing AI systems in a way that makes their output understandable and interpretable by humans, including why certain data influenced a decision. Establishing clear ownership roles and enforcement mechanisms will facilitate responsible data stewardship.

   Example: In a loan approval system, transparency requires documenting that credit history and income data were sourced from verified repositories. Explainability ensures that if a loan is denied, the applicant or reviewer can understand that the decision was based on specific data thresholds and not unrelated factors.

3. **Fit for Purpose & Scalability**

   The high volume of data required to train and operate AI systems—especially in large-scale or generative applications—must be managed in a way that ensures data is both fit for purpose and scalable by design. Only data that is necessary, proportionate, and relevant to a clearly defined, lawful, and approved AI use case should be collected, stored, and processed. AI systems should be designed to scale responsibly, with data architectures that accommodate increasing volumes, varieties, and velocities without compromising security, compliance, or sustainability. Retention must be limited to what is strictly necessary, with automated lifecycle rules enforcing timely archiving or deletion.

Example: An insurance company considered training its claims AI system on 20 years of policyholder records but found older data misaligned with current products and regulations. Using it risked legal and operational issues. Instead, the company opted for a smaller, higher-quality dataset better fit for purpose, resulting in effective and efficient model performance and reduced environmental impact of storage and processing. This fit-for-purpose approach not only improved model accuracy and reduced processing load but also enabled the AI platform to scale more efficiently as new policy types and data streams were added.

4. Environmental Sustainability

Data practices are governed to minimize environmental impact, including energy use, carbon emissions, water consumption for cooling, and e-waste across the AI lifecycle. This involves establishing the appropriate data architecture inclusive of understanding the type of model to use, limiting unnecessary data generation and storage, optimizing data processing for efficiency, and selecting infrastructure that supports sustainability goals – such as renewable-powered data centers, water-efficient cooling systems, and hardware strategies that avoid premature obsolescence.

Example: To reduce environmental impact, an AI team reconfigures a model to run on existing hardware rather than upgrade to more power-hungry servers, reducing e-waste. They also shift model training to a cloud region with renewable energy and water-efficient cooling, cutting both emissions and water use.

5. Accountability

Clear ownership and responsibility must be assigned for managing data integrity and risk across the AI lifecycle. This includes compliance with internal policies and external regulations, supported by audit trails, review processes, and escalation procedures in case of issues or violations. Data security, privacy, and safety-related risk assessment, mitigation, and monitoring must be continuously updated to address the evolving demands of AI systems such as dynamic model behavior, autonomous decision-making, and heightened regulatory scrutiny.

Example: In a healthcare AI application, accountability might involve assigning data stewards responsible for ensuring that patient data is anonymized and reviewed for quality before use. An AI ethics committee may periodically audit these processes, and a defined escalation path would be triggered if sensitive data is improperly accessed.

These principles and the effectiveness of data governance for AI directly impact common business goals. The following table maps business goals to sustainable AI principles and the impact that strong and weak data governance should have on each.

Enterprise Goal	Impact of Strong AI Governance	Impact of Weak AI Governance	Relevant Principles
Ensure accurate, fair, and consistent AI outcomes	Data is validated, complete, and free of systemic bias, leading to reliable, equitable results.	Incomplete or biased data skews results, leading to discrimination or reputational damage.	• Validity & Reliability
Compliance with laws and regulations (e.g., GDPR, AI Act, ESG reporting)	Clear data provenance and transparent, accountable, governance streamlines ESG, compliance, and risk reporting.	Untraceable or excessive data use can lead to legal violations, fines, or audits.	• Transparency & Explainability • Fit for Purpose & Scalability • Accountability
Support stakeholder trust and brand credibility	Transparent, explainable AI decisions and clear data governance reassure employees, customers, partners, and regulators.	Lack of transparency and unclear data handling erodes confidence in the organization's use of AI.	• Transparency & Explainability • Accountability
Improve operational efficiency and decision-making	High-integrity data enables dependable analytics and efficient AI operations.	Dirty, redundant, or low-quality data slows AI pipelines and may lead to bad decisions.	• Validity & Reliability • Fit for Purpose & Scalability
Reduce environmental footprint	Governance ensures efficient data use, responsible hardware management, and low-impact processing.	Oversized datasets and inefficient pipelines increase energy use, water consumption, and e-waste.	• Environmental Sustainability • Fit for Purpose & Scalability
Lower costs of AI ownership and risk mitigation	Purposeful data management reduces storage, compute, compliance, and remediation costs.	Overcollection and unmanaged data increases infrastructure spend and risk exposure.	• Fit for Purpose & Scalability • Accountability • Environmental Sustainability

Part 1: Preparation

Establish Roles and Responsibilities

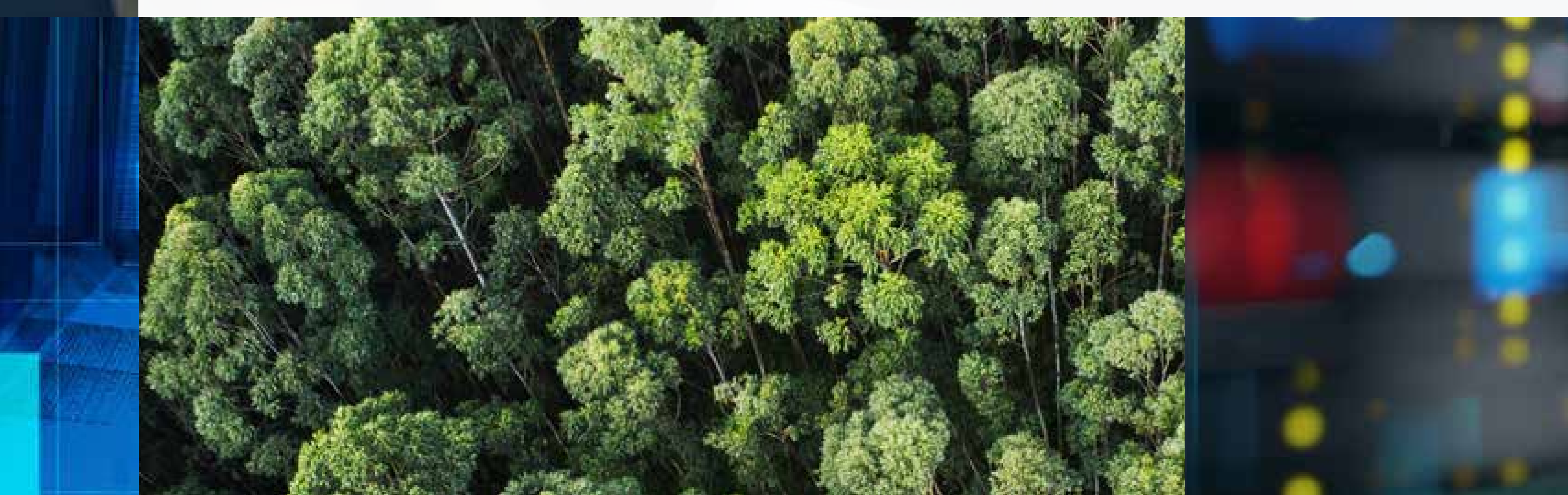
Organizational leaders should identify the teams and individuals responsible for AI data governance, including those directly managing data, as well as those involved in approval, oversight, notification, and general awareness. It is essential to include enterprise architects to align data usage with established design principles and solution architecture patterns. Business data and systems owners, of course, must be engaged, as well as legal representatives to address data's potential liability and compliance concerns. Finally, sustainability offices and AI governance committees should treat data governance as a core component of their responsibilities.

Role	Responsibility	Contact (Name & Contact Details)
AI Governance Team	Define and enforce data governance policies across AI system lifecycles, ensuring traceability, auditability, and alignment with technology and data strategy, and responsible AI data principles.	
Enterprise Architecture Review Board	Work with EAs in each discipline to create high-level solution design of options as well as contribute to business cases. Ensure the AI/ML engineering team is leveraging architectural patterns in the detailed logical and physical design of the solution. Post deployment, captures metrics on solutions alignment to technology and sustainability goals.	
Sustainability Officer	Evaluate and report on the environmental and social impacts of data used in AI systems, including emissions from data storage and processing.	
Compliance Officer	Monitor AI data practices to confirm their legal and regulatory compliance related to data privacy, security, and anti-discrimination.	
Chief Data Officer	Establish enterprise-wide data governance policies to ensure validity, reliability, and ethical use of AI data.	
Legal Counsel	Review AI data usage against legal standards related to data privacy, IP rights, and discrimination laws.	
Data Security & Privacy Officer	Safeguard sensitive data with access controls, encryption, and monitoring for misuse, address loss prevention, and maintain compliance with industry-specific data privacy laws.	
Data Steward	Monitor data quality, lineage, and integrity; implement standards for data collection and usage to mitigate bias.	
Data Administrator	Manages data systems architecture, data volumes, resiliency, security, archiving, retention, storage, distribution, and availability.	
AI/ML Engineering Lead	Implement model documentation, logging, and audit trails to enable explainability and transparency.	
IT Infrastructure Lead	Ensure infrastructure planning and support for data volume, velocity and scalability, including network, storage, availability, monitoring, and security. Also, enforces and enables traceable energy consumption reporting.	
Risk & Audit Manager	Conduct audits and risk assessments to ensure accountability and identify non-compliance in AI data practices.	
Ethics & Responsible AI Committee	Provide cross-functional oversight and escalation pathways for ethical data issues in AI.	
Business Unit Lead	Define purpose-specific use of AI data aligned with business outcomes and legal constraints.	

Assemble Data Governance Prerequisites

Obtaining the following information in structured, accessible formats will help launch or accelerate the actions to follow.

- **Enterprise Data Inventory with Metadata** – A structured inventory of datasets across the organization, including metadata for source, ownership, sensitivity, and use, enables accurate data mapping, lineage tracking, and purpose classification
- **Data Classification and Tagging Framework** – Standardized schema to tag data based on its purpose, sensitivity level, regulatory requirements, and retention policy
- **ModelOps and MLOps Integration Capabilities** – Technical infrastructure to integrate data governance checkpoints into AI development and deployment pipelines
- **Access Management and Logging Infrastructure** – Systems that control and log user access to datasets and AI models, linked to declared purposes
- **Audit-Ready Data Provenance and Lineage Tools** – Mechanisms to document data flow from source to model input/output, including transformations
- **Environmentally Aware Infrastructure Monitoring** – Capability to track energy usage, emissions, and data center efficiency (e.g., Power Usage Effectiveness – PUE)
- **Legal and Regulatory Interpretation Capability** – Internal or external expertise to interpret data-related obligations from laws like GDPR, CCPA, AI Act, etc.
- **Privacy Impact Assessment (PIA)** – Standardized workflows to assess risks and compliance before deploying new AI systems involving data collection, typically done as part of solutioning with enterprise architecture
- **Pre-production Testing** – Process to ensure solutions are aligned with software development lifecycle best practices such as DevSecOps (Development, Security, and Operations)
- **Incident Reporting and Feedback Mechanisms** – Channels and escalation pathways to capture, investigate, and resolve AI-related data governance issues



Part 2 – Principles Operationalization

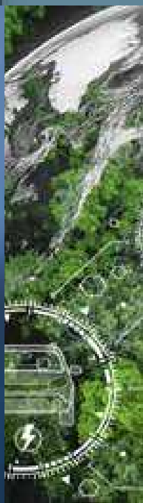
The recommended actions in each of the following sections are not the only actions necessary to establish and embed responsible and sustainable AI data governance into operations. The KPIs and documentation are suggested possibilities, and readers are not expected to adopt all of these, only those best aligned to the organization's priorities, i.e., materiality, regulatory reporting requirements, risk appetite, sustainability commitments, and other factors.

Principle 1: Validity & Reliability

Data used for AI systems will be accurate, consistent, and complete, with processes in place to identify and minimize errors, gaps, and bias—especially those that may impact protected groups (e.g., race, gender, age, or disability). Data must be relevant for its intended purpose, ensuring reliability (i.e., data yields consistent results across time, contexts, and applications).

Development Actions

- ✓ Define and enforce data quality metrics for completeness, accuracy, timeliness, and consistency across datasets. Set thresholds for acceptable variation and flag anomalies.
- ✓ Conduct cross-domain validity testing - Validate models across multiple demographic, geographic, and contextual segments to detect overfitting to dominant or majority data groups.
- ✓ Simulate edge cases or biased input scenarios to evaluate if AI explanations remain consistent, accurate, and meaningful under stress conditions.
- ✓ Investigate and review third-party AI systems and model training data for compliance with validity and inclusiveness standards and expectations.
- ✓ Develop and implement strategies to mitigate identified biases, inaccuracies, and gaps.
- ✓ Begin with the least amount of data needed to create a viable solution. Prioritize use of the organization's own data versus augmentation with outside sources.

KPIs

- **Data Quality Score** – Composite metric combining completeness, accuracy, timeliness, and consistency across datasets used for AI training and inference
- **Anti-bias Methodologies Applied to Model Training** – Percentage AI algorithms or ML models trained using anti-bias methodologies
- **Fairness Metric Deviation** – Difference in model performance (e.g., precision, recall) across protected demographic groups
- **Edge Case Handling Accuracy** – Performance of AI systems when processing rare, extreme, or stress-test input conditions
- **Cross-Domain Model Performance Variance** – Difference in model accuracy or reliability across different demographic or operational segments

Documentation

- **Data Quality Report** – Details the accuracy, completeness, and consistency benchmarks (see Appendix for Template 1)
- **Bias Audit Log** – Captures results of fairness testing and bias detection methods
- **Data Filtering & Exclusion Log** – Document exclusion criteria and data filters, recording what data was excluded during preprocessing and why
- **Logical/Physical Architecture Design** – Diagrams and specifies how data flows through the AI system
- **Cross-Domain Evaluation Report** – Highlights how models performed across diverse segments
- **Training Data Summary** – Describes sources, demographics, and known limitations of training data

Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

- **Great Expectations** – An open-source tool for data validation
- **Talend Data Quality** – A platform for ensuring data integrity and accuracy
- Fairness Indicators (**TensorFlow**), Fairness Indicators (**Google**), **AI Fairness 360** (IBM)
- **Cleanlab** – Python library for detecting label errors, outliers, and data quality issues in training datasets
- **Aequitas** – Open-source bias and fairness audit toolkit that helps stakeholders understand model risk from a social impact perspective

Implementation Actions

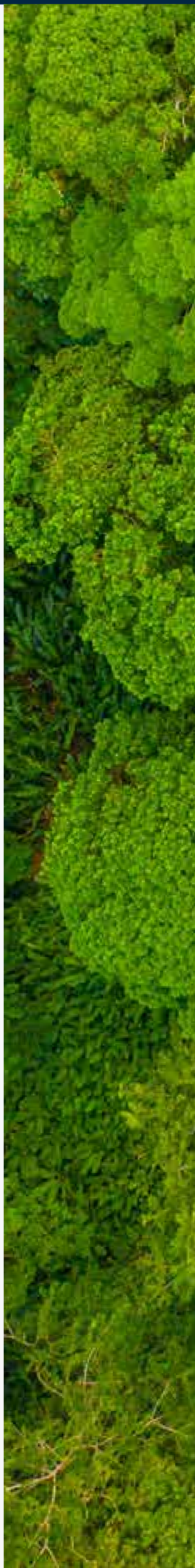
- ✓ **Operationalize model validation checkpoints** – Embed checkpoints in the AI lifecycle to verify model performance under real-world data distributions and changing environments.
- ✓ Implement Human-in-the-Loop (HITL) review for high-risk use cases, integrating expert oversight to validate AI outputs.
- ✓ **Maintain a bias and reliability risk register** – Track known and emerging risks related to model validity and data quality, and link these to mitigation strategies and responsible teams.
- ✓ **Benchmark against representative population baselines** – Compare model input output distributions against known real-world benchmarks to validate representativeness and equity.
- ✓ **Deploy continuous drift monitoring tools** – Monitor for data drift (changes in input data over time) and concept drift (changes in underlying patterns) to maintain model reliability over time.
- ✓ **Assess infrastructure architecture and topology** – Check that it is appropriate for sustainable environments.

KPIs

- » **Bias Mitigation Implementation Rate** – Percentage of detected biases that have documented and applied mitigation strategies
- » **Error Rate in AI Outputs** – Percentage of incorrect or biased AI-generated outputs
- » **Model Drift Detection Rate** – Frequency and magnitude of shifts between training and real-world data distributions
- » **Equity Gap in Model Performance** – Absolute difference in key performance metrics (e.g., false positives/negatives) between demographic subgroups
- » **Discrepancy Rate in HITL Review** – Percentage of AI outputs that are modified or overturned by human reviewers

Documentation

- » **Data Quality Metrics Dashboard** – Tracks ongoing adherence to defined quality thresholds
- » **Validation Checkpoint Report** – Summarizes model performance under real-world testing
- » **HITL Review Log** – Records decisions reviewed by humans, outcomes, and override rates
- » **Issue-Resolution Workflow** – Tracks how anomalies and data quality issues were addressed
- » **User Feedback Log** – Captures end-user feedback related to model performance or errors
- » **Technology Architecture Alignment Log** – Documents how deployed AI systems align with approved architectural patterns, principles, and sustainability requirements (including capture of component reuse and system run-rate metrics)



Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

» **DeepChecks** – Open-source tool for testing and validating data integrity and model reliability during development and deployment phases

» **Data Sheets for Datasets** – A standardized documentation method for ensuring transparency and accountability in dataset collection and use, hosted by Cornell University

» **WhyLabs AI Observatory** – A platform for monitoring data and model quality in production environments, helping detect drift, anomalies, and reliability issues

» Arize AI – An ML observability platform offering tools to monitor performance degradation, data drift, and fairness breakdowns across diverse segments

» Evidently AI – A toolkit for evaluating, monitoring, and visualizing machine learning model performance, including drift and integrity metrics

Principle 2: Transparency & Explainability

Organizations will maintain clear, traceable records of where AI data comes from, how it is collected, processed, and used—including any transformations, filtering, or enrichment. AI systems are designed to make their outputs understandable and interpretable by humans, including why certain data influences a decision.

Development Actions

✓ Institute a formal data provenance and lineage framework—Track and document where data originates, how it was transformed, and how it moves through the AI system to ensure traceability and quality assurance.

✓ Document exclusion criteria and data filters—Record what data was excluded during preprocessing and why—supporting transparency, reproducibility, and fairness audits.

✓ Require third-party or vendor AI systems to include explainability and transparency documentation as part of procurement or risk assessment processes.

✓ Develop “tiered explainability” by providing different levels of explanation tailored to different audiences—technical teams, executives, regulators, and end users—balancing detail with accessibility.

✓ Seek validation through third-party AI governance or ethics assessments (e.g., ISO/IEC 42001, NIST AI RMF) to demonstrate transparency commitments.

KPIs

- **Data Provenance Completeness Rate** – Percentage of datasets with fully documented origin, lineage, and transformation history
- **Data Exclusion Rationale Coverage** – Percentage of datasets where exclusion rules and rationale are explicitly documented and reviewed
- **Model Documentation Coverage Rate** – Percentage of AI models accompanied by complete documentation, including enterprise and solutions architecture, training data description, decision rationale, and known limitations
- **Number of Models Using Standardized Transparency Artifacts** – Count or percentage of models using tools like Model Cards, Data Sheets for Datasets, or Fact Sheets for AI

Documentation

- **Data Provenance Map** – A visual or structured record (e.g., flow diagram or table) documenting the origin of each dataset, key transformations (e.g., labeling, aggregation), and how the data flows through the AI pipeline
- **Exclusion Criteria Record** – A documented log detailing what data was removed or filtered during preprocessing, along with justifications
- **Model Cards** – Standardized documentation outlining each AI model's intended use, architecture, performance metrics, known limitations, ethical considerations, and decision rationale (see Appendix for Template 2)
- **Training Run Log** – A chronological record of each model training session, including dataset versions, hyperparameters, training duration, performance benchmarks, and environmental variables, to facilitate traceability and support debugging
- **Data Sheets for Datasets** – Structured documentation that includes metadata about training datasets, such as collection methods, demographics, intended use, licensing, known biases, and maintenance procedures. Data Sheets help users evaluate the dataset's fitness and risks for specific AI applications.

Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

- **IBM's AI Explainability 360** – A toolkit to support the interpretability and explainability of AI models
- **Explainability Tools** – **DARPA XAI**, **SHAP**, **LIME**
- **Microsoft InterpretML** – An open-source library for training interpretable models and explaining black-box systems using methods like SHAP and LIME
- **Data Nutrition Project** – Provides "nutrition labels" for datasets to promote better understanding and responsible use in AI model training
- **Model Cards for Model Reporting** – Standardized template for documenting model behavior, use cases, and ethical considerations, aiding in stakeholder transparency

Implementation Actions

- ✓ Conduct regular collection and analysis of feedback from users, technical teams, and regulators regarding the clarity and comprehensibility of AI system outputs.
- ✓ Conduct periodic assessment of AI systems' performance against transparency requirements, identifying gaps and areas for improvement. Establish and utilize a feedback loop to infrastructure and GreenOps architects.
- ✓ Provide summaries or visualizations for stakeholders of how key decisions were made by AI systems (e.g., via dashboards or interactive reports).
- ✓ Maintain a documentation inventory of AI system components including data inputs, preprocessing steps, model types, training runs, and deployment environments to support traceability.

✓ Make available clear data usage policies to stakeholders, ensuring they are easily understandable and accessible.

KPIs

» **Lineage Governance Assignment Rate** – Percentage of AI projects with clearly assigned owners responsible for data lineage, versioning, and change tracking

» **Use of Interpretable Model Architectures** – Percentage of models designed for interpretability (e.g., decision trees, GAMs, attention-based networks) in use where explainability is critical

» **AI Decision Explainability Score** – AI output interpretability and transparency (measured using tools like SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations)

» **Explanation Satisfaction Score** – Average score from internal or external users rating the clarity and usefulness of AI model explanations (via surveys or UX feedback)

» **Frequency of Explanation Review Audits** – Number of scheduled explanation audit reviews conducted per year to evaluate transparency effectiveness and alignment with user needs

Documentation

» **Tiered Explanation Summaries** – Curated explanation outputs tailored for different audiences highlighting relevant details such as key features influencing decisions, model confidence levels, and limitations in plain or technical language

» **Public-Facing AI Use Statement** – Describes purpose, data used, and limitations to inform and improve stakeholder understanding

» **System ROI Justification Review** – Documented value assessments (quantitative and qualitative) to promote transparency of outcomes

» **Feedback Collection Form** – Captures user perspectives on the clarity of AI outputs to inform transparency effectiveness or areas for improvement

» **Third-Party Transparency Assessment Record** – Reports or certificates from independent audits or reviews (e.g., ISO/IEC 42001, NIST AI RMF conformity) verifying that transparency and explainability standards are met

Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

- **FactSheets for AI (IBM Research)** – Structured transparency documentation approach covering AI system purpose, performance, safety, fairness, and explainability
- **Model Cards for Model Reporting** – Standardized template for documenting model behavior, use cases, and ethical considerations, aiding in stakeholder transparency
- **NIST AI Risk Management Framework (AI RMF 1.0)** – Provides structured guidance for improving explainability, trustworthiness, and transparency in AI systems
- **ISO/IEC 42001 – AI Management System Standard (2023)** – A formal international standard helping organizations manage AI risk, including explainability and transparency principles
- **FAT Forensics** – Python toolkit for auditing fairness, accountability, and transparency of predictive systems with built-in explainability tools

Principle 3: Fit For Purpose & Scalability

AI data must be purpose-driven and scalable by design. Organizations should collect, store, and process only data necessary, proportionate, and relevant to a clearly defined, lawful, and approved AI use case. At the same time, AI systems architecture must be designed to scale responsibly—able to manage increasing volumes, varieties, and velocities of data without compromising security, compliance, or sustainability.

Development Actions

- ✓ Define authorized purposes, clearly documenting legally authorized and business-approved purposes for data use in AI systems, aligned with compliance and regulatory obligations.
- ✓ Perform AI data mapping and classification, identifying what data is collected, where it resides, and how it is used, tagged with purpose, sensitivity, and retention requirements.
- ✓ Implement data minimization controls in AI systems design and processes, specifying collection of only data strictly required for the authorized purpose and discouraging “just-in-case” data hoarding.
- ✓ Design AI data pipelines to be modular and scalable, supporting flexible integration of new data sources while maintaining alignment with purpose and compliance requirements.
- ✓ Develop or adapt data system architecture requirements to define system scalability and management guidelines (including volume, velocity, sources, security, compliance and retention).
- ✓ Establish system capacity planning protocols that anticipate scale in data volume, variety, and velocity, including stress testing under future load scenarios.

KPIs

- **AI Data Minimization Score** – Percentage of datasets where only the minimum necessary data is collected based on the defined purpose
- **Purpose-Mapped Data Coverage** – Percentage of datasets that are explicitly tagged with a defined and documented purpose
- **Systems with Data Lifecycle Policies Implemented** – Percentage of enterprise systems with automated lifecycle rules (e.g., retention, archival, deletion) in place
- **Scalability Readiness Score** – Percentage of AI systems with documented scalability plans addressing data volume, velocity, and integration growth

Documentation

- **Purpose and Use Definition Record** – Includes formally approved purpose definitions for each dataset or AI system and data catalog annotations tagging data to specific, lawful uses
- **Privacy and Fit-for-Purpose Assessment** – Structured assessments (e.g., PIAs) verify data collection, use, and retention align with privacy requirements and documented purposes
- **Data Minimization and Lifecycle Design Notes** – Combined technical and policy documentation show how systems restrict data collection to what is necessary, and how they implement automated rules for retention, archival, and deletion
- **Scalable Architecture and Capacity Planning Document** – Includes system architecture blueprints, integration patterns, and future-state growth forecasts to ensure responsible scaling across data volume, velocity, and variety

Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

- Tools that identify and mask sensitive data and support "compact" data (e.g., **Perforce** and **Granica**)
- IBM AI Privacy **Toolkit**
- NIST AI Minimization **Toolkit**
- Collibra Data **Catalog** – Enables purpose-tagging, data lineage mapping, and stewardship

tracking in development stages

Implementation Actions

- ✓ Train staff on purpose-driven AI data use and how to apply minimization principles in daily decision-making.
- ✓ Automate AI data retention and deletion with lifecycle management tools to enforce time-bound data retention and automated deletion when data is no longer needed.
- ✓ Audit AI data use against purpose by regularly reviewing system logs and access patterns to ensure data is only used for its intended purpose. Flag deviations for review.
- ✓ Monitor and report on AI data lifecycle compliance, showing data volumes, retention status, deletion rates, and exceptions by purpose.
- ✓ Require justification and documentation for any data scope expansion.

KPIs

- » **Staff Training Rates** – Percentage of relevant employees trained annually on Fit for Purpose principles and data handling responsibilities
- » **Data Purpose Alignment Audit Score** – Internal/external audit rating of alignment between data use and declared purpose over period of usage
- » **Unauthorized Purpose Access Rate** – Number of instances where data was accessed or used for purposes outside of the documented authorized use
- » **Average Data Retention Duration** – Average length of time data is retained (segmented by data type)
- » **Data Volumes** – Total GB/TB of data, as well as redundant, obsolete, or trivial data–identified as misaligned with any current purpose or expired, and rate of change over time

Documentation

- **Purpose Compliance Review Report** – Internal or third-party assessment for tracking whether deployed AI systems use data in alignment with declared purposes (see Appendix for Template 3)
- **Training Attendance and Completion Log** – Evidence that relevant employees (e.g., analysts, engineers, product managers) have been trained in Fit-for-Purpose policies
- **Exception Handling Log for Data Use** – Documents how deviations from purpose-aligned use (e.g., temporary overrides or emergency use) were managed and approved
- **Data Retention and Deletion Metrics Report** – Dashboards and records showing how long data is held and when it is purged according to policy
- **Scaling Decision Log** – Records documenting the rationale, timing, and impact analysis of decisions to scale AI systems (e.g., increase model inputs, broaden geographic data intake, add real-time sources)
- **Performance Monitoring Dashboard** – Real-time tracking of throughput, latency, and resource utilization as data volume and variety increases, confirming the system's scalable performance

Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

- **APP Supporting Data Minimization in AI** – Guidance for applying federated learning principles to minimize data use during implementation
- **AI Now Institute Data Minimization as a Tool for AI Accountability** – Frameworks for deploying systems that enforce minimization during real-world use
- **Balancing GDPR & Research Insights** – Best practices for implementing legally compliant, purpose-limited data usage in AI
- **BigID** – Lifecycle automation for data discovery, purpose-based retention rules, and deletion enforcement
- **OneTrust Data Governance** – Supports purpose-based data retention scheduling, policy automation, and compliance training

Principle 4: Environmental Sustainability

AI data practices are governed to minimize environmental impact, including energy use, carbon emissions, water consumption for cooling, and e-waste across the AI system lifecycle. This involves limiting unnecessary data generation and storage, optimizing data processing for efficiency, and selecting infrastructure that supports sustainability goals—such as renewable-powered data centers, water-efficient cooling systems, and hardware strategies that avoid premature obsolescence.

Development Actions

✓ Evaluate environmental impact of data acquisition, labeling, preprocessing, and storage workflows, including compute load and volume of data required for training.

✓ Design data ingestion and preprocessing pipelines to reduce duplication, data sprawl, and unnecessary data transformations that increase energy use.

✓ Embed sustainability goals into data sourcing and data governance frameworks—prioritizing lower-footprint datasets, synthetic data when appropriate, and efficient data formatting.

✓ Periodically review training data volume, transfer rates, and retention timelines to eliminate stale or unnecessary data that increases storage and compute burden.

✓ Evaluate sustainable alternatives to LLMs for both training and application phases based on data volume needs, performance trade-offs, and environmental impact:

- Modular AI architectures that allow selective model retraining instead of full reprocessing
- Retrieval-augmented generation (RAG) systems that rely on external databases to reduce training data loads
- Fine-tuned or distilled models that reuse pre-trained knowledge with minimal new data
- Edge-optimized SLMs that reduce reliance on centralized data centers
- Alternatives to AI, including viable existing analytic and automation tools

KPIs

- **Estimated Training Energy Consumption per Model** – Total kilowatt-hours (kWh) consumed during training of a specific model, calculated based on dataset volume, quality, and preprocessing complexity; used to benchmark and reduce energy-intensive training practices
- **Low-Carbon Dataset Processing Rate** – Percentage of training datasets stored, processed, or preprocessed using energy-efficient or renewable-powered infrastructure (e.g., green data centers, low-carbon cloud services)
- **Dataset Utilization Efficiency Score** – Ratio of model performance (e.g., accuracy, F1 score) to dataset size or volume, highlighting how effectively data is used to achieve outcomes
- **Preprocessing Energy Intensity Metric** – Measured energy consumption (kWh) per gigabyte (GB) of data preprocessed, helping identify and reduce high-energy preprocessing tasks

Documentation

- AI Design Notes with Green Optimization Strategies – Record design choices to reduce environmental impact, such as data sampling, pruning, and efficient formatting
- Data Lineage and Provenance Log with Environmental Metadata – Track data origin and processing with metadata on storage location, energy source, and carbon intensity
- Environmental Impact Assessment Report – Evaluates emissions, energy, and water use across AI data workflows (see Appendix for Template 4)
- Compute and Storage Budgeting Log – Documents resource needs per dataset or model to prevent overprovisioning and reduce waste

Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

- **Hugging Face Optimum** – Toolkit for evaluating model size, deployment efficiency, and energy trade-offs
- **OpenDC** – Simulator for data center energy and cooling needs
- **WattTime** – Provides real-time data on the carbon intensity of electricity grids for optimization
- **Microsoft Sustainability Manager** – ESG dashboard for tracking sustainability metrics, including AI and cloud workloads
- **WhyLabs + Prometheus** – AI observability and monitoring tools that can track resource usage
- **DataHub** – Open-source metadata platform to manage and monitor data lifecycle impacts

Implementation Actions

- ✓ Select smaller, less resource-intensive model—such as Small Language Models or compressed variants—for inference tasks to lower energy consumption, reduce emissions, and improve scalability without compromising essential functionality.
- ✓ Equip stakeholders with the knowledge to implement efficient data strategies such as deduplication of inputs, compression of outputs, and responsible data retention to help curb unnecessary data proliferation and mitigate environmental impact.
- ✓ Implement automated or manual output compression/tokenization techniques that minimize log file sizes and storage requirements.
- ✓ Purge or archive outdated input/output datasets and logs from completed AI tasks to limit persistent storage energy costs and prevent data buildup.
- ✓ Monitor environmental impact of inference output size and logging practices to identify inefficiencies, enforce environmentally sound logging practices, and ensure inference processes remain aligned with sustainability targets.

KPIs

- » **Total CO_2e Reduction from Model Optimization** – Estimated carbon emissions (kg or tons CO_2e) avoided through model refinement techniques such as output pruning and data-efficient sampling, indicating effectiveness of sustainability-driven model design choices.
- » **AI Environmental Drift Score** – A composite score that tracks changes in environmental resource use (compute time, memory, storage, logging volume) over time due to evolving input/output data patterns
- » **Output Storage Efficiency Ratio** – Average number of bytes stored per inference, factoring in compression, tokenization, and logging policies; used to evaluate how efficiently inference outputs are managed for storage-related energy impact mitigation
- » **Water Consumption Score per Inference** – Estimated water usage (liters or gallons) required for cooling per AI inference, calculated using backend infrastructure water efficiency and tied to data volume, runtime, and system location; promotes awareness of water-intensive compute loads
- » **Carbon Intensity per AI Workflow Execution** – Actual emissions output per execution (e.g., gCO_2e/inference or per task completed), reflecting the combined environmental cost of data processing, inference, and logging

Documentation

- **Input/Output Data Retention Logs** – Storage time, purpose, and environmental impact
- **Green Inference Design Notes** – Focused on minimizing emissions from output handling
- **Optimization Change Logs** – Related to data pruning or compression in outputs
- **Environmental Performance Monitoring Reports** – With breakdown by data handling stage
- **Environmental Drift Reports** – Specifically related to data handling and retention practices
- **Input/Output Data Retention Log** – Records the duration, purpose, and retention policies of input and output datasets, including associated storage energy estimates and justifications for continued retention
- **Green Inference Design Note** – Documents inference phase efforts to reduce environmental impact, including output format selection, real-time compression, and infrastructure choices
- **Optimization Change Log** – Tracks modifications to data or model outputs such as pruning, tokenization, or compression; include justifications and impact metrics
- **Environmental Performance Monitoring Report** – Evaluates environmental performance across the AI system lifecycle, segmented by data handling stage (e.g., preprocessing, storage, inference), with metrics on emissions, energy, and water use.
- **Environmental Drift Report** – Identifies and assess increases in environmental impact due to data accumulation, logging growth, or changes in inference behavior over time; support corrective actions when system operations begin deviating from sustainability targets.

Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

- **WattTime** – Tracks carbon intensity of power used for data inference tasks
- **Whylogs** – Data logging for ML systems
- **DataHub** – Metadata platform for tracking environmental impact of data lifecycle
- **Microsoft Sustainability Manager** – ESG dashboard that includes AI data use metrics

» **MLCO2 Impact Calculator** – Tool for estimating emissions from model training and inference

Principle 5: Accountability

Roles and responsibilities for complying with ESG governance principles, policies, and regulations are established and clear for AI system data. Accountability is supported by monitoring and audits throughout the AI lifecycle and governance structures such as AI ethics committee, green IT architecture standards, a risk assessment framework and an escalation process for non-compliance.

Development Actions

- ✓ Create a cross-functional AI governance council (enterprise data architect, data scientists, legal, ethics, IT, sustainability, and business leaders) to provide ongoing data oversight and prioritization of risks
- ✓ Assign AI data stewardship roles responsible for determining data relevance, necessity, and lifecycle requirements.
- ✓ Develop or adapt data and information policies that define ethical guidelines (including equity, fairness, IP protection), data privacy standards, and security protocols.
- ✓ Embed responsible and sustainable AI data and information requirements into enterprise data, application, integration and security architectures.
- ✓ Integrate AI data governance checkpoints into Architecture Review Boards, Technical Review Board (Solution Architects and Engineers), DevOps or MLOps, requiring validation gates for ethics, compliance, and sustainability before deploying AI models to production.

KPIs

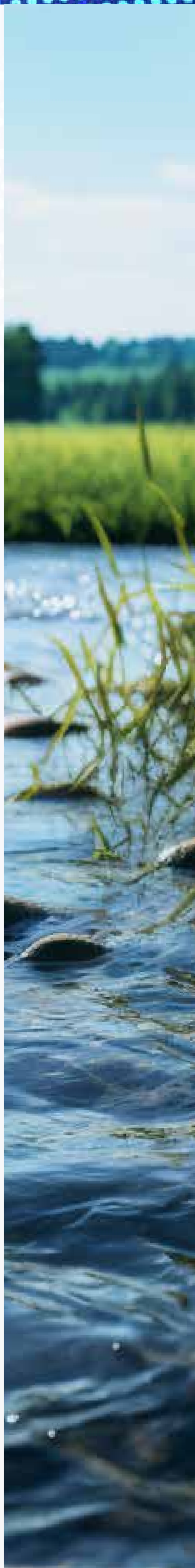
» **AI Vendor Data-Ethics Review** – Percentage of AI spend on vendors reviewed for AI data ethics & inclusiveness

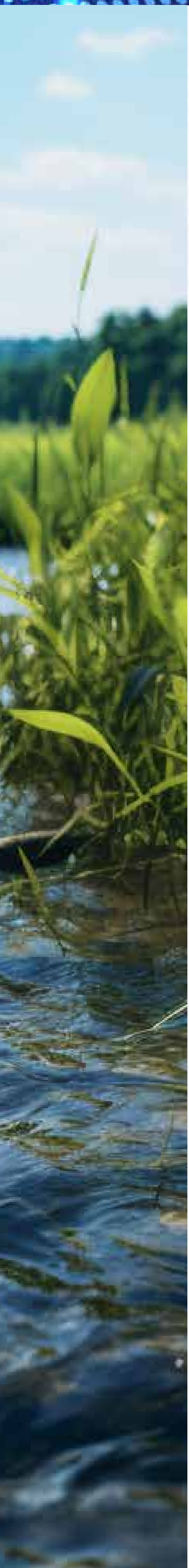
» **Third-Party AI Review Rate** – Percentage of high-risk AI models reviewed by independent external experts or ethics boards

» **Alignment Score with Net-Zero AI Principles** – KPI measuring consistency of AI system design and data governance with enterprise net-zero or climate action goals

» **Ethical Audit Frequency and Ethical AI Data Compliance Score** – AI model data management adherence to ethical AI principles (corporate & 3rd-party)

» **AI Governance Gate Pass Rate in MLOps** – Percentage of AI models that pass all governance





checks (e.g., ethics, fairness, sustainability) before deployment

Documentation

- **Bias Audit Log** – Provides evidence that fairness testing was conducted and outcomes were evaluated
- **Benchmark Comparison Sheet** – Helps compare performance against standards, supporting decisions around model readiness and fitness
- **AI Vendor Review Record** – Supports verification that third-party models or data meet ethical and governance standards
- **Remediation Pathway Log** – Enables demonstration that a defined and followed process exists for bias and harm mitigation
- **Architecture Compliance Log (LCA report, and ARB Decision of Record)** – Enables confirmation that responsible and sustainable design has been implemented early in technical planning and architecture

Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

- ITI's AI Accountability **Framework**
- AI Governance Best Practices and Importance by **Informatica**
- **OECD's AI Principles** – Guidelines promoting transparency and accountability in AI
- Atlan AI Governance Framework
- Enterprise Architecture Frameworks (**TOGAF**, **IASA BTABoK**) for sustainability and responsible AI frameworks

Implementation Actions

- ✓ Monitor and report the environmental footprint of AI workloads, tracking compute usage, data center efficiency (PUE), and cloud provider emissions for training and inference tasks
- ✓ Mandate data-oriented AI sustainability lifecycle assessments (LCAs) to evaluate data's contribution to energy consumption, compute intensity, and emissions of AI systems.
- ✓ Engineer a "kill switch" for rapid shut-down of AI systems to avoid proliferating biased or inaccurate data and potentially harmful automated actions or decisions.
- ✓ Conduct mandatory compliance training for employees in management and use of data and information with AI tools.
- ✓ Establish or adapt whistleblower and feedback channels.

KPIs

- **AI Report Channels and Tracking** – Number of active reporting channels for AI data governance concerns and reports over time, and percentage resolved by corrections
- **Ethical Audit Frequency and Ethical AI Data Compliance Score** – AI model data management adherence to ethical AI principles (corporate & 3rd-party)
- **AI Environmental Impact Audit Frequency** – Number of audits conducted on environmental impact of AI models (e.g., energy usage per training run)
- **Remediation Completion Rate for AI Failures** – Percentage of identified biased or faulty AI data-related outcomes with completed corrective actions and documented learnings
- **AI Data Literacy** – Hours and percentage of employees trained in AI dataset management and usage compliance with regulations, standards, ethics guidelines, privacy and security

Documentation

- **Training Completion Log** – Proof of data literacy training for staff, reducing risk of violations and supporting audits
- **Whistleblower Feedback Records** – Validates responsiveness to ethical or legal concerns
- **Data Quality Issue Log** – Facilitates tracing of data problems and resolutions
- **Compliance Incident Log** – Demonstrates that AI data breaches were identified, tracked, and resolved per policy
- **ARB Decision of Record** – Decision log and quarterly metrics tied to architectural sustainability and responsible AI metrics

Tools (representative sample, not endorsed by SustainableIT.org, its members or partners)

- ISO **37002** Whistleblowing Management
- **MLFlow with Role-Based Access and Emergency Flags** – Allows runtime interruption or model deactivation through API controls

Conclusion

The Responsible AI Data Governance Principles Runbook provides a practical, action-oriented guide that enables data and AI leaders to embed environmental, social, and governance considerations into every stage of the AI system lifecycle. As AI systems evolve and regulatory expectations grow, this Runbook serves as a living reference to support continuous improvement. Organizations are encouraged to adapt its recommendations to their unique contexts, share lessons learned, and contribute to the broader movement toward sustainable digital transformation.

By implementing these principles with discipline and integrity, organizations not only safeguard against ethical, legal, and environmental risks—they also strengthen resilience, earn stakeholder trust, and lead responsibly in the age of AI.

Appendix

The following are starter templates for some of the key documentation suggestions in the Runbook. These are suggestions and should be tailored to fit an organization's operating model and priorities.

Template 1: AI Data Quality Report

Documents benchmarks for accuracy, completeness, consistency, and timeliness of datasets used in AI model training and inference.

Dataset Name	Accuracy %	Completeness %	Consistency	Timeliness	Reviewer	Notes
Claims_ Train_2023Q3	97.8	98.5	95.2	Current	Name	Minor discrepancies in state codes normalized during preprocessing
Credit_ Score_2022	95.2	88.9	92.1	2022-11-15	Name	Some fields incomplete due to third-party provider integration
ChatLogs_2024_ Prod	99.1	99.7	98.9	Live feed	Name	Real-time feed monitored with anomaly detection enabled

Template 2: AI Model Transparency Tracker

Tracks how transparency is maintained across AI models, including documentation, interpretability, and provenance completeness.

Model Name	Model Card Complete	Lineage Documented	Explanation Tool Used	Reviewed By	Notes
CreditRiskML_v2	Yes	Yes	SHAP	AI Risk Team	Model card available to external auditors
ProductRankAI	No	Partial	None	In progress	Backlog item for Q3 documentation push
MedMatch_NLP	Yes	Yes	LIME + GAM	Compliance & Legal	Explainer dashboard available to clinicians

Template 3: AI Fit-for-Purpose Data Approved Usage Log

Documents how datasets are mapped to approved purposes and verifies lifecycle compliance.

Dataset Name	Approved Purpose	Retention Policy	Usage Verified	Last Access Date	Notes
Employee_History	Workforce planning only	3 years max retention	Yes	2025-04-20	Access logs audited quarterly
Sensor_Logs_ Vehicles	Predictive maintenance	Auto-delete after 12m	Yes	2025-05-15	Scaling concerns flagged; moving to compact format
Legacy_Client_Data	Not approved for AI use	Archive only	No	2023-10-10	Flagged in compliance audit – to be purged Q2

Template 4: Environmental Impact Report for AI Data Workflows

Captures sustainability metrics related to data collection, storage, and preprocessing.

Workflow Name	Estimated kWh Used	Carbon Intensity (gCO_2e)	Water Use (L)	Optimization Applied	Notes
NLP_Model_ Training	1,120	420,000	780	Trained on pruned dataset	Shifted to cleaner cloud region in final training phase
Vision_Pipeline_V1	2,340	680,000	1,100	RAG model + batch preprocessing	Dataset size reduced by 38% via format standardization
Predictive_Maint	680	150,000	290	Deployed as compressed model	Embedded logging limits storage overhead

About SustainableIT.org

Vision

Advancing global sustainability through technology leadership.

Mission

Our mission is to unite the world's largest community of technology and sustainability leaders to define sustainability transformation programs, author best practices and frameworks, set standards and certifications for governance, provide education and training, and raise awareness for IT-centric ESG programs that make their organizations and the world sustainable for generations to come.

Mandates

Best Practices, Research and Standards

Identify sustainable digital transformation programs by industry. Research and define best practices, frameworks, and standards for all three pillars of sustainability (environmental, societal, and governance) for IT departments and organizations.

Global Awareness and Recognition

Promote sustainable digital transformation programs and advances in sustainability. Raise awareness through local, regional, and global awards, as well as through social media, publications, and public relations.

Community, Education and Training

Build local and regional communities for technology leaders to advance sustainability. Develop education and training programs for IT leadership and professionals for all three pillars of sustainability.

Transparency and Accountability

Set standards for metrics and reporting to enable transparency and accountability. Create certification programs for individuals and organizations with rights to use our sustainability emblem.

SUSTAINABLEIT .ORG

About SustainableIT.org

SustainableIT.org is a Delaware 501(c)(6) nonprofit, non-stock legal entity led by technology executives who will advance global sustainability through technology leadership. Our mission is to define sustainable transformation programs, author best practices and frameworks, set standards and certifications, provide education and training, and raise awareness for environmental and societal programs that make our organizations and the world sustainable for generations to come.