SUSTAINABLEIT
.ORG

The AI Sustainability Runbook

A Practical Guide

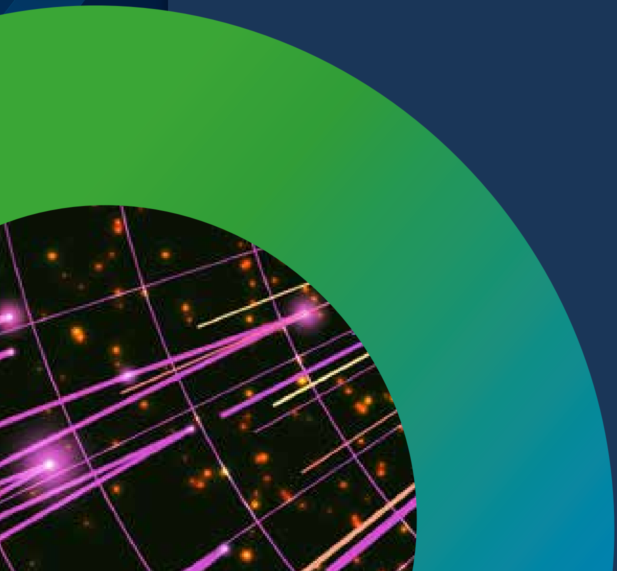
IT Toolkit for Responsible & Sustainable AI: A Field Guide for Implementation at Scale

As AI evolves to be the foundational enabler of business transformation, it is imperative that its governance extends beyond performance, ethics, and compliance to include comprehensive environmental, social, and governance sustainability.

Yet despite the accelerating deployment of AI systems across industries, there remains a striking lack of practical guidance for how enterprises—especially IT and business leaders—can design, deploy, and manage AI responsibly, especially from an environmental perspective. This gap creates risks not only for energy use and emissions but also for hardware overprovisioning, water consumption, and electronic waste. IT leaders are uniquely positioned to lead the charge: they design and manage the infrastructure, understand the lifecycle dynamics of digital systems, and serve as key architects of AI strategy and operations.

This Runbook responds to that leadership imperative, providing actionable, lifecycle-based guidance to help operationalize environmental sustainability across AI systems—from goal setting and architecture to monitoring and continuous improvement. It was developed and vetted by volunteer AI experts and IT practitioners from SustainableIT.org's 50+ member Responsible AI Working Group. It empowers IT to integrate ESG priorities into AI governance and guarantee that innovation advances not only performance, but planetary resilience.

Given the relative lack of environmental AI guidance for practitioners, this Runbook focuses solely on this aspect of ESG sustainability, with an acknowledgement that social and governance sustainability are a must-have for a fully responsible AI portfolio.



The Case for AI Environmental Sustainability

AI in the enterprise will have both positive and negative sustainability impacts. Starting with the environmental aspect of ESG, intelligent data analysis and automation can streamline supply chains, minimize transportation emissions, and improve energy management in industries ranging from manufacturing to logistics. Predictive maintenance powered by AI extends the life of equipment, reducing the need for premature replacement and lowering embodied carbon impacts. AI also accelerates the design of more sustainable products and materials, supports smart energy grids and materials circularity, and enables dynamic optimization of building systems for reduced electricity consumption.

However, AI's negative environmental impacts can arise from the model's frequent inferences, which increase demand for high-performance hardware, such as GPUs or AI accelerators, which draw more power than standard CPUs. Water consumption for traditional data center cooling is typically increased and can add strain to local resources. Performance demands will shorten device lifecycles, contributing to higher e-waste and embedded emissions. The net sustainability impact of any AI system depends on how efficiently its workloads and data are managed and whether infrastructure upgrades are guided by both performance needs and environmental factors.

An AI portfolio that prioritizes sustainability will yield these key benefits:

» **Operational & Environmental Efficiency** – A sustainability perspective encourages low-carbon, resource-efficient AI solutions and AI data stewardship that reduce environmental impact and operational costs over time.

» **Risk Mitigation** – Proactively identifying and addressing sustainability risks protects the enterprise against reputational harm, regulatory penalties, and stakeholder pushback.

» **Regulatory Readiness** – Sustainable AI governance addresses current and forthcoming regulations related to ESG and AI (e.g., CSRD, SEC climate disclosure rules, California's SB253).

» **Stakeholder Trust** – Demonstrating a commitment to sustainable AI enhances trust among investors, customers, and employees.

» **Infrastructure Resilience and Longevity** – Embedding sustainability into AI governance extends the useful life of infrastructure by minimizing hardware overuse, avoiding premature obsolescence, and aligning with circular economy principles.

» **Sustainable Growth** – Governing AI sustainably enables the development and scaling of applications that drive innovation that supports long-term business growth and positive societal outcomes.

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

The Runbook is organized into four parts: **Goals Setting and Governance**, **Impact Assessment, Impact Acceleration/Mitigation,** and **Monitoring and Improvement**. Each part includes a series of recommended action with associated documentation and key performance indicators. It is important to note that the recommended action steps in each of the following sections are not the only actions necessary to establish and embed sustainable AI 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.

Part 1 – Goal Setting and Governance

The foundation of sustainable AI begins not with technology, but with intention: a shared organizational commitment to align AI development and deployment with environmental, social, and governance (ESG) goals. This Runbook is focused solely on environmental sustainability (social and governance Runbooks will follow). Part 1 of this Runbook outlines the essential structures required to turn that intention into repeatable action—starting with clear sustainability objectives and extending into enterprise-wide accountability and oversight.

To ensure AI systems support long-term business resilience and stakeholder trust, sustainability goals must be mapped to the entire AI lifecycle, from model design to retirement. These goals then require translation into governance mechanisms that guide decision-making, enforce policy compliance, and enable transparent measurement of environmental impacts. This section covers the fundamental steps to build that foundation.

Step 1: Establish AI Sustainability Goals

It is essential to have clear, articulated sustainability goals when approaching AI governance. In most cases, the environmental, social, and governance (ESG) commitments of the enterprise can be mapped to the lifecycle of AI systems, just as they are applied to information technology in general. AI's impact (positive and negative) on the enterprise's targeted goals must be evaluated, applied and/or mitigated. The following table lists common ESG goals and AI's potential positive and negative impacts:

ESG Goal	Positive AI Impact	Negative AI Impact
Net-Zero Carbon Emissions	AI can optimize energy use, improve building management, streamline logistics, and help design lower-carbon products.	Training and operating large AI models (especially GenAI) can significantly increase energy consumption and carbon emissions if not managed carefully. Embodied emissions from AI hardware (servers, chips) are often overlooked.
Circular Economy and Waste Reduction	AI can improve waste tracking, recycling, and product lifecycle analysis, enabling design for reuse, refurbishment, and recycling. Predictive maintenance powered by AI extends product life.	AI-accelerated consumption patterns (e.g., faster product development cycles) can worsen short hardware lifecycles and increase the amount of electronic waste.
Responsible Water and Resource Use	AI can optimize water use in agriculture and manufacturing, detect leaks, forecast water demand, and enhance recycling processes.	Large AI data centers often have water footprints for cooling, contributing to local water stress, especially in drought-prone regions.
Climate Risk Resilience	AI can forecast climate risks better (e.g., flooding, supply chain disruption), optimize agricultural outputs, and enhance disaster response planning.	AI models trained on biased or incomplete climate data may mislead planning, creating overconfidence or misallocation of resources.
Biodiversity Protection and Land Stewardship	AI tools (e.g., satellite imaging, acoustic monitoring) can help detect illegal deforestation, protect endangered species, and optimize land management.	Large-scale AI infrastructure (e.g., server farms, mining for rare earth minerals for AI hardware) can contribute to land degradation and biodiversity loss.

Step 2: Embed Sustainability in Governance

Once sustainability goals are defined, organizations must embed these goals into a governance structure that ensures consistent oversight, accountability, and compliance throughout the AI lifecycle. AI sustainability governance provides the connective tissue between high-level environmental ambitions and the day-to-day decisions made during AI development, deployment, and operations.

This requires integrating sustainability requirements into the broader enterprise AI governance framework, not treating them as parallel or standalone processes. Doing so helps ensure that all AI initiatives are evaluated, approved, and monitored through a consistent environmental sustainability-aligned lens.

Actions

✓ **Policy Integration** – Codify sustainability principles into AI governance policies, covering areas such as energy efficiency, data minimization, fairness, transparency, and lifecycle management. These policies should align with enterprise environmental goals and be referenced during model approval, deployment, and scaling.

✓ **Environmental Evaluation** – Require ESG impact evaluation in business case development and technical feasibility studies.

✓ **Sustainability Risk Management Frameworks** – Extend the enterprise's AI risk register to include sustainability risks—e.g., excessive emissions and water consumption, premature hardware obsolescence. Use materiality assessments and thresholds to prioritize risk mitigation.

✓ **Workflow Checkpoints** – Mandate sustainability checkpoints within model approval and deployment workflows. The following are checkpoint examples for the pre-deployment stage of the AI lifecycle.

- **Sustainability Impact Screening** – Flag high-energy models, sensitive use cases, or hardware-intensive designs.
- **Architecture Sustainability Review** – Technical reviewers evaluate whether the model architecture aligns with low-carbon design principles (e.g., SLM, pruning, retrieval-augmented generation (RAG)).
- **Storage and Transfer Efficiency Review** – Assess if training/inference data is optimized to reduce movement, duplication, and high-intensity storage (e.g., "cold" tier use).
- **Infrastructure Alignment Verification** – Ensure workload is assigned to a cloud/data center with verified low-carbon energy sourcing. Verify the model will run on hardware with sufficient utilization rates (>50%) to avoid overprovisioning.

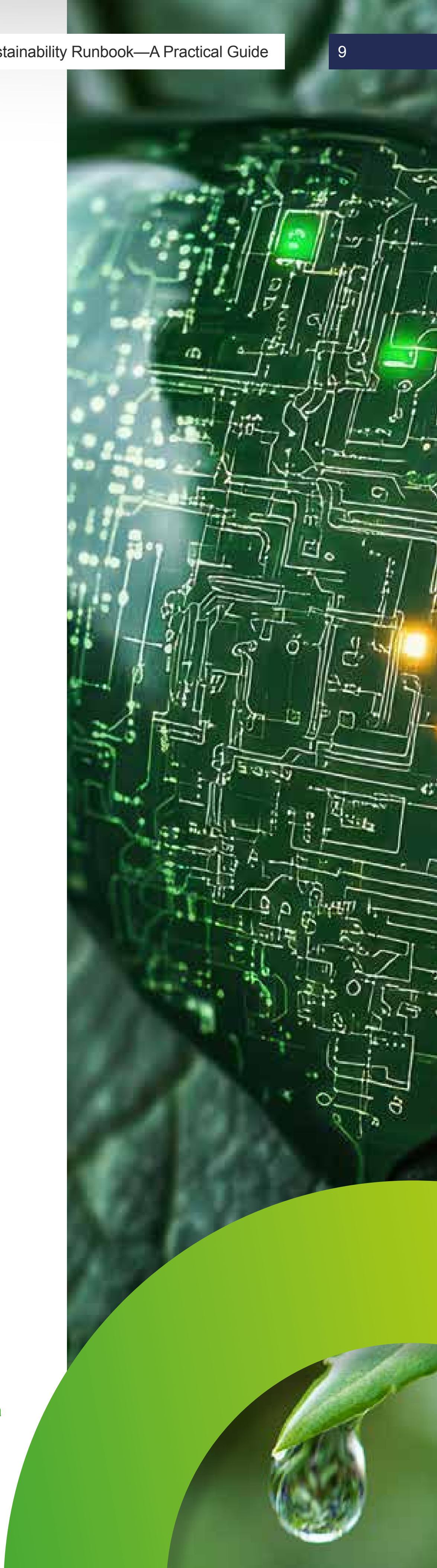
- ✓ **Oversight Mechanisms** – Assign sustainability oversight roles to existing AI governance bodies (e.g., AI councils) or establish new working groups focused on environmental responsibility. Ensure these bodies review sustainability impact assessments, mitigation plans, and post-deployment audits.
- ✓ **Integration with Compliance and Audit Functions** – Coordinate with legal, audit, and sustainability reporting teams to ensure compliance with emerging regulations (e.g., CSRD, EU AI Act, SEC climate disclosures) and alignment with third-party sustainability standards (e.g., ISO 14001, GHG Protocol, ISO/IEC 42001).
- ✓ **Documentation and Decision Logs** – Establish standardized documentation requirements (e.g., AI Sustainability Risk Register, Sustainability SLA for Vendors) to support governance transparency, audit readiness, and continuous improvement. (See Appendix for templates.)
- ✓ **Continuous Governance Feedback Loop** – Use sustainability performance data from post-deployment monitoring (see Part 4 of this Runbook) to refine governance policies and risk thresholds over time. Incorporate lessons learned into updated governance playbooks and training for cross-functional teams.

For long-term success, the governance setup should follow a cycle similar to environmental or energy management systems, like ISO 14001 (Environmental Management Systems) or ISO 50001 (Energy Management Systems). This means applying a “Plan-Do-Check-Act” structure to AI sustainability: Plan goals and responsibilities, Do the activities, Check progress using data, and Act to improve based on what you learn. This helps ensure AI systems stay energy-efficient and aligned with evolving ESG targets over time. For example, review model efficiency reports every quarter and update design guidelines if needed.

Ultimately, governance ensures that sustainability goals do not remain aspirational—but become enforceable, operational, and measurable commitments within the AI portfolio.

Step 3: Establish Roles and Responsibilities

Role	Responsibility	Contact (Name & Contact Details)
AI Governance Team	Oversee AI lifecycle sustainability governance processes and ensure compliance with policies.	
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	Promote AI sustainability best practices and spearhead education throughout the organization, assess environmental, safety, and social impacts, and develop mitigation plans.	
Compliance Officer	Ensure AI applications comply with sustainability-related regulations.	
Business Unit Leads	Align AI application goals with corporate sustainability strategies.	
Green IT Program Lead	Evaluate AI's impact on energy use and related factors, aligning with IT Sustainability targets and roadmaps.	
Facilities Management Lead	Manage energy sources and power usage effectiveness (PUE) in data centers.	
IT and Cloud Infrastructure Teams	Assess and explore low-carbon computing options.	
Safety Manager	Responsible for the safety of the developed systems. In cases where AI is integrated, the safety manager should oversee compliance with AI safety standards, like ISO 8800 (for Automotive).	



Step 4: Acquire Tools and Resources

To establish AI sustainability governance, the team will need access to knowledge, resources and tools to help at various points along the AI system lifecycle. Among these are the following fundamental prerequisites:

- ✓ Documented corporate framework outlining sustainability goals and commitments
- ✓ Data center and energy consumption and emissions baselines (on-premises, cloud-hosted, and SaaS)
- ✓ Vendor and internal data center Power Usage Effectiveness (PUE) and Water Usage Effectiveness (WUE) scores
- ✓ Guidance for assessing AI models and architectures, including language models, which can vary widely in energy demand

The IT team should confer with the enterprise's sustainability and finance officers to what governance tools are already implemented and seek out experienced users to understand their benefits and limits. The following are resources and tools applicable to AI sustainability across the ESG spectrum. (Company and product names referenced in this list are for informational purposes and are not an endorsement by SustainableIT.org or its members and partners.) Also see the Toolkit's Reference Guide to Responsible AI Governance Frameworks, Principles, and Standards.

Environmental Tools and Resources

- » **ESG Automated Data Collection and Reporting Platforms** – Platforms from ServiceNow (**ESG Management**), IBM (**Envizi**), and Salesforce (**Net Zero Cloud**) aligned with reporting standards help companies measure, manage, and report ESG performance. Within AI governance, they facilitate transparency around sustainability metrics, ethical use, and alignment with regulations.
- » **ESG Performance and Risk Management Software** – Tools like **Datamaran** or **SpheraCloud** help identify, assess, and monitor environmental, social, and governance risks. In AI governance, they support decision-making by mapping sustainability risks related to AI development, such as energy consumption, labor impacts, and regulatory exposure.
- » **Software Carbon Assessment Methodologies** – Certified open source tools such as the Cloud **Carbon Footprint Methodology** or the ISO-certified **SCI specification** will help consistently quantify emissions associated with software, including AI systems, and progress against reduction targets.
- » **AI Carbon Impact Calculators** – Tools such as **CodeCarbon**, **EcoLogits**, and **ML CO2** Impact Tool estimate the greenhouse gas emissions from training and running GenAI models.
- » **AI Data Minimization Tools** – NIST offers an AI Minimization **Toolkit** to help machine learning model developers determine the minimal level of detail required for newly collected data to make accurate predictions with an AI model. Other frameworks and tools include IAPP **Supporting Data Minimization in AI** and **AI Now Institute's Data Minimization as a Tool for AI Accountability**.

Part 2 – Responsible Design and Development

Responsible design and development are the first and most consequential opportunity to embed sustainability into the AI lifecycle. However, before starting with this first step, a preliminary evaluation of whether AI is actually necessary for the use case at hand should be conducted. Given current trends, it is tempting to settle on AI quickly. Nonetheless, many use cases are much easier and more sustainably solved by conventional automation such as short scripts or low-code platforms. As a rule of thumb, the more structured the data is, the more likely that AI is neither the most functionally adequate nor the most sustainable solution.

While goals and governance policy establish the organizational intent, it is at the design stage that sustainability commitments are translated into real-world decisions—about which models to acquire and use, what infrastructure to run on, how much data to collect, and how to ensure equitable and efficient outcomes. In order to conduct this phase within the best interests of all stakeholders, it is important to foster a common understanding by offering workshops or training available to all stakeholders. (See the Toolkit's Persona-based Literacy Guide for Sustainable AI.)

This phase should proactively define sustainability-oriented design rules, architectural preferences, and model development parameters that align with enterprise ESG goals. These might include favoring energy-efficient model types, restricting the use of large general-purpose models where not necessary, or mandating inclusive design protocols and fairness guardrails from the outset.

If an AI solution is selected, then rather than assessing sustainability as an afterthought, responsible AI development requires that environmental, social, and governance considerations be treated as core design constraints—alongside accuracy, latency, scalability, and cost. This means prioritizing AI architectures that reduce emissions and energy intensity, incorporating reuse and circularity principles into infrastructure planning, and anticipating ethical and social risks that may emerge post-deployment.

This section outlines how to operationalize responsible design and development by (1) determining the most fit-for-purpose and efficient AI system for a given use case, and (2) evaluating its likely environmental and social sustainability impacts. These steps lay the foundation for more effective mitigation, optimization, and lifecycle governance later on.

Step 1: Conduct Fit-for-Purpose Evaluation

Actions

✓ Determine if the proposed use case merits traditional, generative or other forms of AI.

✓ Consider whether a Large Language Model is necessary for the AI application. More sustainable alternatives use considerably less energy and permit AI applications to run on lower-powered devices, potentially reducing hardware churn and e-waste impact:

- Small Language Models and narrow, domain-specific
- Model Pruning, quantization, and distillation techniques that shrink LLMs without significantly losing performance
- Retrieval-Augmented Generation, which combines a smaller base model with an external database or search system
- On-Device and Edge AI Models, which run locally on devices (phones, laptops, IoT) instead of in massive cloud servers

Utilizing these LLM alternatives can have significant impact on energy demand and carbon footprint with only slight performance compromises.

Option	Typical Carbon Footprint	Energy Demand	Performance Tradeoff
LLMs	Very high (training: hundreds of tons CO_2e)	Extremely high during inference	Very broad and versatile
Small Language Models	Very low to moderate	Low	Narrower capabilities
RAG with Small Models	Moderate	Low to moderate	Strong performance for info retrieval
Edge/On-Device AI	Very low	Very low	Limited by device capacity

An example of this evaluation is the implementation of a customer support chatbot on a company's website. Such a chatbot will generally conduct very short and simple conversations that do not require a broad scope and advanced reasoning. They do, however, need to access company policies such as the return policy. That being the case, a small retrieval-augmented model is most suitable.

Step 2: Embed Sustainable AI in Architecture

Sustainability must be embedded into the design phase of AI systems—not as an afterthought, but as a foundational set of parameters that guide technical decisions from the outset. This requires moving beyond reactive mitigation and toward intentional design choices that anticipate environmental and social risks while maximizing long-term value.

By defining sustainability objectives early in the development process—such as minimizing compute waste, extending hardware lifecycles, enabling equitable access, and ensuring model reuse—organizations can reduce unnecessary tradeoffs later and avoid lock-in to high-impact architectures. These actions not only align AI development with ESG commitments but also contribute to more resilient, cost-efficient, and future-ready AI systems.

The following practices provide a blueprint for embedding sustainability directly into the architecture, workflows, and operational design of AI applications.

✓ **Apply Design-for-Sustainability (DfS) Principles.**

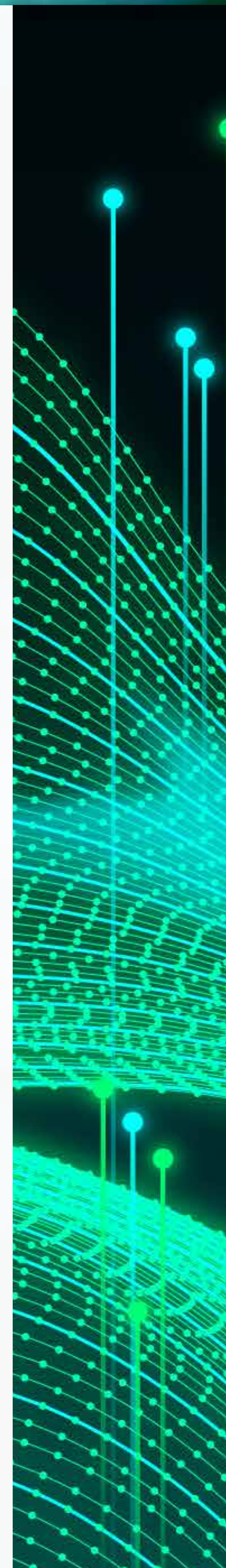
- Choose models that meet the “minimum viable intelligence” needed for the task.
- Avoid overtraining or overfitting that leads to unnecessary compute use.
- Design for graceful degradation—systems should remain effective with smaller retrained models or reduced input.

✓ **Prioritize Low-Impact Compute Architectures.**

- Deploy workloads on energy-efficient chipsets (e.g., SLM-optimized CPUs, TPUs).
- Use cloud-native autoscaling to reduce idle compute and emissions.
- Architect inference pipelines to minimize storage and retrieval redundancy.

✓ **Plan for Model and Infrastructure Longevity.**

- Build modular AI applications to extend reuse of models across use cases.
- Align model development timelines with expected hardware refresh cycles to reduce premature obsolescence.
- Use hardware lifecycle planning tools to flag AI-induced upgrade pressures early.



✓ **Define Sustainability Acceptance Criteria.**

- Add sustainability gates to technical review checklists (e.g., "Will this increase cloud emissions?").
- Require justification for GenAI model use and energy consumption estimates from prior model runs or from benchmarks in model cards (e.g., OpenAI, Hugging Face), and model profiling tools.
- Establish baseline model efficiency thresholds (e.g., gCO_2e/inference caps) by combining performance benchmarking and carbon estimation tools which correlate runtime and hardware usage to CO_2e.

Step 3: Conduct Sustainability Impact Assessment

To operationalize Sustainable AI, organizations must assess not only whether an AI system can perform a task, but whether it should—given its environmental, social, and governance (ESG) consequences. While AI systems can generate positive ESG value, they can also introduce significant sustainability risks through their energy use, carbon emissions, hardware demand, bias, or lack of explainability. Understanding these impacts early allows teams to make smarter, more responsible design and deployment decisions.

Sustainability Impact Assessment is a structured evaluation of a proposed AI system's potential environmental, social, and governance impacts across its lifecycle—before and during development, and after deployment. In this case, the goal is to determine whether the system contributes to or detracts from environmental sustainability goals, of the organization using a mix of qualitative judgment and quantitative indicators (e.g., emissions, energy and water use, compliance readiness). This assessment supports informed decision-making and enables sustainable design tradeoffs aligned with enterprise values and regulations.

Sustainability impact assessments do not need to be complex or time-consuming. When integrated into routine review workflows—such as during model selection, architecture design, or pilot evaluation—they can become a lightweight but powerful gatekeeper. To avoid analysis paralysis, assessments should be scoped to focus on the most material ESG dimensions relevant to the organization's stated sustainability priorities. Not all metrics are equally relevant for every use case.

Where possible, these assessments should be fast-tracked through:

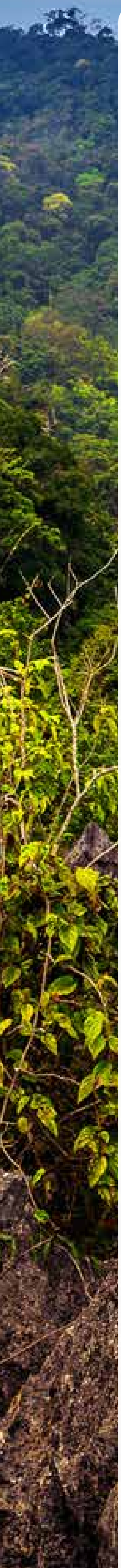
» Checklists and pre-approved model classes (e.g., low-impact SLMs)

» Pilot-stage measurement tools for estimating energy, emissions, or fairness

» Use of standardized questionnaires and dashboards to simplify documentation

» Reusing known metrics from similar applications already deployed

This section outlines how to perform a targeted sustainability impact assessment using common metrics.



Environmental Impacts and Metrics

There are four basic environmental metrics for IT, AI included—energy, emissions, water, and e-waste. These are often referred to as “hidden” impacts because they are difficult to isolate from other non-AI technology systems running in data centers or cloud. However, organizations with solid data center and cloud baselines can estimate and track changes associated with AI deployment. The fundamental impacts are:

» **Energy Consumption Rate** – Energy (kWh) used for AI model training and inference

  How to Collect: Use cloud provider usage dashboards (e.g., AWS CloudWatch, Azure Monitor) or on-premises power metering tools. For storage and data transfer, use cloud billing data to quantify GB transferred and stored

» **AI Carbon Footprint** – GHG emissions (kgCO_2e) associated with compute, storage, and data transmission

  How to Collect: Obtain energy consumption data as described above, then multiply energy usage by grid-specific carbon intensity factors or emissions coefficients provided by cloud providers or regional electricity authorities

» **Water Consumption Rate** – Water used for data center cooling (liters or m^3)

  How to Collect: Request WUE (Water Usage Effectiveness) from cloud providers or use estimated WUE x energy usage

» **E-waste Generation Rate** – Mass (kg/year) of electronic waste tied to early device retirement for AI capacity

  How to Collect: Track hardware replacement cycles and categorize AI-driven refreshes. Estimate device weight using vendor specs

The following are applicable metrics that may be selected for use depending on the priorities of the organization in terms of materiality, regulatory reporting requirements, risk appetite, sustainability commitments, and other factors.

» **Carbon Impact** – For an accurate perspective, totals should include emissions caused by the AI application’s compute, storage, and data movement. Metrics may include:

  - **Carbon Intensity** – Emissions (gCO_2e/kWh) per unit of energy used (gCO_2e/kWh), dependent on the grid mix or cloud provider’s sourcing

    How to Collect: Obtain regional grid data (e.g., EPA eGRID, IEA, Ember), or cloud provider’s regional mix

  - **Total Inference carbon footprint** – Estimate carbon emissions for a single inference—including compute and retrieval—based on the number of input and output tokens processed (gCO_2e/inference)

    How to collect: Estimate with CodeCarbon or similar, which may require AI token count × compute profile × energy/carbon intensity

  - **Lifecycle Carbon Footprint** – Full estimate including training, deployment, maintenance, and updates over the AI model's intended lifetime ($kgCO_2e$)

    How to collect: Calculate with LCA tools with inputs across training, deployment, and updates

  - **Scope 2 Emissions Attribution** – Allocation share of the organization's indirect emissions (e.g., cloud energy use) to AI applications

    How to collect: Obtain cloud billing + carbon reporting APIs, allocating based on usage per project

  - **Software Carbon Intensity Score** – Emissions per unit of functionality delivered by the AI model (e.g., gCO_2e per 1,000 inferences)

    How to collect: Use SCI specification from Green Software Foundation (emissions per 1,000 inferences)

- **Data Processing Efficiency Impact** – The amount of compute and energy used to produce AI system output. Metrics can include:

  - **Data-to-Computer Ratio** – Amount of input data (in GB) processed per TFLOP (Tera Floating Point Operation). Higher ratio = more efficient

    How to collect: Measure input GB ÷ TFLOPs consumed (from training/inference logs)

  - **Processing Latency** – Time taken to process a typical data input (ms or seconds)—lower is better

    How to collect: Run inference benchmarks; use profiling tools like NVIDIA Nsight or PyTorch Profiler

  - **Energy per Inference** – How much energy (Joules) it takes to produce one output

    How to collect: Energy consumption ÷ number of inferences (tools like CodeCarbon can automate this)

  - **Throughput (Inferences/Sec)** – How much energy (Joules) it takes to produce one output

    How to collect: Count inferences over time using model serving logs

- **Data Generation and Storage Efficiency Impact** – How efficiently the AI application generates new data (output) and stores models, logs, and artifacts. Metrics may include:

  - **Embodied Carbon of Storage** – Emissions ($kgCO_2e$) embedded in physical storage hardware use and lifecycle

    How to collect: Use vendor LCA data + storage capacity deployed

  - **Model Size** – Disk space (MB, GB) required for the deployed AI model (smaller sizes reduce storage burden)

    How to collect: Obtain binary file size of the trained model

- **Compression Ratio** – Percentage reduction achieved through data compression techniques on output or models

  How to collect: Obtain dataset sizes before and after compression

- **Dark Data Ratio** – Percentage of data that is generated but remains dormant and unused in a certain time period

  How to collect: Analyze access logs and storage activity to compare total data generated against inactive data over a defined period

- **Storage Energy Intensity** – Energy required to store model weights, output, or logs per GB annually (can be derived from cloud provider carbon data or data center PUE benchmarks

  How to collect: Calculate energy per GB × storage time, using figures from cloud provider or average data center PUE

- **Cold/hot Storage Proportion** – Percentage of data kept in “hot” (high-energy) versus “cold” (low-energy) storage tiers. More cold = more efficient

  How to collect: Calculate percentage split of storage classes (e.g., AWS S3 Standard vs Glacier)

» **Infrastructure Lifecycle Impact** – The demand for new, more powerful hardware necessitated by AI applications, as well as overprovisioning for the AI portfolio. Potential metrics include:

- **Hardware Refresh Rate** – Percentage of AI-related devices replaced annually (separate AI-driven refreshes from standard IT refresh)

  How to collect: Obtain annual procurement logs and compare against AI-specific needs

- **Compute Density Increase** – Percentage of upgrades to higher-compute-density devices (e.g., CPU to GPU, GPU to TPU/ASIC) necessitated by AI systems

  How to collect: Calculate and compare changing levels in deployment from CPUs to GPUs/TPUs

- **Embodied Carbon per Device and Total** – Carbon footprint (kgCO_2e) from manufacturing of new AI infrastructure hardware (for Scope 3 emissions reporting)

  How to collect: Use LCA figures from hardware manufacturers

- **Average Utilization Rate** – CPU/GPU/TPU utilization percentage post-AI deployment; Low utilization (<40%) suggests over-provisioning, worsening environmental impact.

  How to collect: Monitor usage logs across time with tools like NVIDIA-SMI or Prometheus

- **Hardware Overprovisioning Ratio** – Deployed capacity versus average needed capacity (a high ratios mean unsustainable hardware sizing)

  How to collect: Calculate total provisioned capacity ÷ average peak use

It is vital to understand that these metrics do not exist independently of each other. For instance, consider a data center using an open-loop cooling system. On the one hand the cooling effect by itself might lead to an increase in efficiency of the GPUs and thus to lower energy consumption and lower GHG emissions. On the other hand, an open-loop cooling system wastes significant amounts of water, which can exacerbate water scarcity in areas with already stressed water systems.

Once sustainability positive and negative AI impacts have been assessed and estimated, organizations can qualitatively score the AI system's "Sustainability Justification"—i.e., will the environmental cost of deploying the AI system justify the benefit it creates? The tradeoff between the AI system's environmental sustainability impact and its performance (accuracy, efficiency, speed) coupled with the organization's risk tolerance will all factor in the decision to deploy, scale, redesign or refuse the application. And since responsible governance carries on after deployment, the opportunity to further mitigate negative impacts and accentuate positives should be part of the AI lifecycle strategy.

Documentation

» **Sustainability Questionnaire Results** – Determines an AI application's potential sustainability impact, starting with general tech sustainability questions for suppliers and AI-specific questions such as estimated CO_2 per model call, energy usage and data center sourcing, and emissions across model training/inference lifecycle (see Appendix for Template 1)

» **AI Sustainability Risk Register** – Lists environmental sustainability risks, meant to be continuously updated (see Appendix for Template 2)

» **Vendor Sustainability & Compliance Assessment Report** – Evaluates third-party AI vendors for sustainable AI commitments and environmental footprint

» **AI Sustainability Compliance Checklist** – Assesses whether the AI portfolio complies with relevant sustainability frameworks, such as ISO 14001 and GHG Protocol

Part 3 – Advancing ESG Outcomes Through Impact Mitigation

Responsible AI design means going beyond minimizing harm—it involves actively shaping AI systems to support positive environmental, social, and governance outcomes. While risks may emerge throughout the AI lifecycle, organizations can embed resilient and regenerative practices early to enhance sustainability performance.

Yet, many companies have not yet adopted key practices to reduce AI's environmental impact. Capgemini's research shows that none of the main sustainability levers are in use by more than a third of organizations (see table below). This includes easy wins like green cloud hosting, smaller model architectures, or efficient training.

The problem is not a lack of tools. The problem is that most teams do not treat sustainability as a basic design requirement. These practices should not be seen as advanced or optional. They should be part of every AI development process.

The goal is simple—use what works. Make sustainability a routine step across the lifecycle. Add clear checks into design, vendor selection, deployment, and decommissioning. Track what is in place and close the gaps.

With this mindset, sustainability becomes a habit. AI systems will use less energy, run on cleaner infrastructure, and last longer with fewer upgrades. Over time, this builds stronger systems that are also better for the planet.

Area/ Lifecycle Focus	Exemplary Mitigation Levers	% Companies Implementing Measures
Hardware	- Energy efficient hardware usage	10%
Algorithm	- Recyclability	31%
Model Architecture	- E-waste management	6.60%
Infrastructure	- Energy efficient coding	11%
Data	- Energy efficient training algorithms	11%
Data Center	- Fine-tuning models	34%
Usage	- Using smaller models	19%

Percentage of Companies Mitigating Environmental Impacts Across AI Systems Lifecycle
Source: Capgemini Research Institute (2024) – *Developing Sustainable GenAI*

Step 1: Mitigate Environmental Impact

Actions

✓ **Prioritize AI Workload Placement by Carbon Intensity** – Run AI compute in cloud regions or on-premises data centers powered by renewable energy. Use carbon-aware workload orchestration to route model training and inference jobs to the lowest-carbon infrastructure available.

✓ **Optimize and Right-size AI Models** – Apply techniques such as model pruning and distillation to reduce the computational and energy demands of large-scale models without sacrificing accuracy.

✓ **Implement Hardware Efficiency Standards** – Require AI workloads to run on energy-efficient hardware (e.g., GPUs or TPUs with lower energy-per-operation ratings).

✓ **Implement Hardware Lifecycle Management** – Avoid premature end-of-life caused by model size escalation by extending the lifespan of AI-specific hardware through reuse, resale, or repurposing practices.

✓ **Apply Sustainability Criteria in Vendor Evaluations** – When selecting cloud, colocation or SaaS providers for AI workloads, evaluate environmental credentials including renewable energy sourcing, sustainability certifications, and climate-related transparency reporting. (See Appendix for Template 3.)

✓ **AI Lifecycle Data Exit Strategy** – Establish decommissioning protocols and data retirement criteria to prevent potential leakage and increase stakeholder trust.

✓ **AI Lifecycle Hardware Exit Strategy** – Establish decommissioning protocols that prioritize circular economy principles and consider future uses for hardware.

✓ **Automatic Dark Data Processing** – Establish automatic rules and procedures for deleting or achieving dormant and unused data after a certain period of time.

Documentation

» **ESG Impact Mitigation Plan** – Structured plan detailing mitigation measures, responsible parties, and implementation timelines

» **AI Sustainability Assessment Report** – Formalizes documentation of sustainability impacts and recommendations

» **AI Sustainability Decision Log** – Records key decisions on sustainability tradeoffs, accepted risks, and mitigation actions with executive signoff to enable accountability and audit readiness

» **AI Infrastructure Sustainability Baseline Report** – Summarizes carbon, energy, and hardware lifecycle benchmarks for current infrastructure supporting the AI portfolio (updated after major buildouts)

» **AI Carbon-Aware Deployment Record** – Logs AI portfolio workloads routed to low-carbon regions or infrastructure, including emission savings

Part 4 – Monitoring and Improvement

Sustainable AI systems require more than responsible design and implementation—they demand active oversight and continuous enhancement. Once deployed, AI models and infrastructure must be monitored for ongoing alignment with environmental and governance expectations. AI systems may experience “drift” when the real-world data or conditions they encounter during operation diverge from the data or assumptions they were originally trained on. This can lead to reduced reliability over time and may affect factors that impact environmental sustainability goals. For example, the AI model may compensate by generating more complex or frequent inferences to “catch up,” increasing compute cycles and energy use per inference.

This section outlines how enterprises can implement monitoring tools, define corrective action protocols, and establish learning loops that translate sustainability performance insights into measurable improvements. By embedding these practices into the AI lifecycle, organizations not only reduce risk and environmental impact, but also strengthen accountability, resilience, and alignment with enterprise ESG commitments.

Step 1: Establish Monitoring Mechanisms

AI system sustainability must be monitored after implementation because there will be systemic changes as new data enters the system and AI models “learn.” Selective monitoring should be applied to impact areas corresponding to the enterprise’s highest priority sustainability goals. Automated data collection and dashboards are best practices from corporate sustainability programs that should be applied to AI systems sustainability. Below is a representative list of monitoring tools.

Energy and Emissions Monitoring

- **Cloud Provider APIs** – APIs from cloud providers, such as **Google Cloud’s Carbon Footprint API**, Microsoft’s Azure **Sustainability Calculator**, and Amazon’s **Customer Carbon Footprint Tool** can facilitate energy and emissions monitoring for AI workloads
- **Data Center Infrastructure Management (DCIM) Platforms (e.g., Vertiv Environet Alert, Schneider EcoStruxure)** – For on-premises AI infrastructure, provides continuous monitoring of data center resource sustainability via sensors and centralized dashboard for automated alerts and sustainability KPIs

- **NVIDIA Data Center GPU Manager (DCGM)** – For GPU-intensive AI systems, real-time telemetry for energy use, thermal stress, hardware degradation, and utilization efficiency
- **Kubernetes-Based Resource Monitors (e.g., Prometheus with GPU and Power Exporters)** – For containerized AI workloads, continuous monitoring of resource usage

True automation of monitoring typically requires integration of these tools into MLOps, ITOps, or Infrastructure monitoring pipelines—not just manual dashboards. Hybrid setups (on-prem AI + cloud AI) will need combined cloud-native and on-premises approaches for full visibility.

Step 2: Define and Implement Continuous Improvement Processes

Results of monitoring must be actionable, fueling ongoing improvement to AI models and their application. A key early step is to assign responsibility for monitoring of primary AI sustainability impacts, mitigation efforts, and commitments. Ensure that monitoring is coordinated with enterprise audits and integrated into a feedback loop for actions that will support continuous improvement of the AI lifecycle sustainability. Establish threshold-based alerting and escalation protocols defining sustainability thresholds (e.g., energy use per inference, carbon intensity, bias scores). Configure alerts to notify responsible teams when thresholds are exceeded, triggering reviews or model rollbacks.

Actions

- ✓ Incorporate monitoring findings into AI lifecycle reviews, feeding results directly into design and governance gates (as defined in Part 3) to close the feedback loop and drive design-phase improvements in future AI projects.
- ✓ Incorporate sustainability criteria specific to AI systems into vendor and partner evaluation criteria, e.g., model efficiency metrics (e.g., FLOPs, gCO_2e/inference), emissions data per AI workload or service, and use of renewable energy in data centers. (See Appendix for Template 3).
- ✓ Benchmark sustainability performance against internal and external standards.
- ✓ Regularly compare monitored sustainability performance (e.g., compute intensity, energy efficiency) against enterprise baselines, industry benchmarks (e.g., MLCommons), or best-practice targets to identify areas for advancement.
- ✓ Implement corrective action plans based on monitoring outcomes, with accountable owners, deadlines, and follow-up assessments. A hypothetical example follows, which addresses traceability, accountability, and alignment with governance principles while actively improving the environmental performance of AI systems. (See Appendix for Template 4.)

Step 3: Institutionalize Learnings and Raise Sustainability Maturity

Continuous improvement should evolve beyond corrective actions to support strategic capability building. By embedding sustainability insights into strategic planning and organizational processes, enterprises can elevate AI sustainability from reactive compliance to innovation leadership.

Actions

- ✓ **Incorporate Sustainability Insights into AI Strategy Planning** – Use trends from monitoring to guide investment in sustainable AI tools, model architectures (e.g., small/efficient models), or infrastructure choices for future projects.
- ✓ **Create an AI Sustainability Performance Scorecard** – Develop an executive-level dashboard aggregating environmental, fairness, and governance indicators for all AI projects, to inform leadership decisions and accountability. (See Appendix for Template 5.)
- ✓ **Include Sustainability Metrics in Individual and Team Performance Reviews** – Tie responsible AI performance to KPIs for teams managing model development, infrastructure, and governance.
- ✓ **Host Quarterly AI Sustainability Review Forums** – Gather cross-functional stakeholders (e.g., ESG, IT, data science, procurement) to review performance, share lessons learned, and set new improvement targets.
- ✓ **Submit Annual AI Sustainability Impact Statements** – Contribute AI-specific disclosures to broader ESG reporting processes or publish standalone summaries to increase stakeholder transparency and credibility.

Documentation

- **Sustainability Metrics Record** – Maintains tracked indicators (e.g., energy per inference, bias levels, emissions) across AI systems, organized by project, team, or use case
- **Sustainability Compliance Update Record** – Tracks evolving ESG and AI-specific regulatory developments and how these influence AI design and monitoring processes
- **AI ESG Training and Awareness Documentation Update Record** – Captures completed training sessions, curriculum updates, participant metrics, and new topics integrated into Responsible AI or ESG curricula
- **AI Sustainability Audit Reports** – Reports from internal or third-party sustainability assessments of AI projects, including non-compliance findings, improvement recommendations, and audit trails
- **AI Sustainability Performance Scorecard** – A regularly updated, executive-level dashboard aggregating environmental (e.g., energy, emissions), fairness (e.g., bias scores), and governance metrics. Links project performance to corporate KPIs. (See Appendix for Template.)
- **Annual AI Sustainability Impact Statement** – Public- or stakeholder-facing summary of environmental and social impacts, corrective actions taken, and strategic changes made (may feed into ESG disclosures)

Conclusion

Sustainable AI is not achieved through a single policy or tool, but through the disciplined integration of ESG priorities across the entire AI lifecycle—from goal setting and design to deployment, monitoring, and continuous improvement. This Runbook provides a practical foundation to help enterprises turn sustainability commitments into operational practices that scale with their AI portfolios.

By embedding sustainability into governance, architecture, risk management, and performance oversight, organizations can ensure their AI investments are not only effective, but also ethical, efficient, and aligned with long-term value creation. As sustainability expectations and regulatory pressures evolve, this framework enables teams to adapt, demonstrate leadership, and deliver AI systems that benefit business, society, and the environment.

We welcome feedback and suggestions at info@sustainableIT.org.

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 the organization’s situations and priorities.

Template 1: AI Sustainability Impact Assessment Questionnaire

Use this template during the planning or design phase of any AI project to assess potential environmental impacts. This structured assessment can be integrated into model approval processes or design reviews.

Question	Response	Notes / Justification
What is the estimated carbon footprint per inference?	~0.08 gCO_2e/inference	Estimated using CodeCarbon with emissions factor from Electricity Map and model runtime profiling
Is the model deployed on low-carbon infrastructure (cloud/ data center)?	Yes – deployed in Google Cloud's europe-west4 region (carbon-free energy >90%)	Region selected based on GCP Carbon-Free Energy % dashboard and availability of sustainability SLAs
Does the model use energy-efficient architecture (e.g., SLMs, pruned LLMs)?	Yes – model uses a distilled transformer architecture with quantization	Chosen to reduce compute requirements and improve energy efficiency without sacrificing performance
Has model size been minimized or optimized (compression, distillation)?	Yes – original 3.4B parameter model compressed to 800M via distillation	Compression achieved a 75% reduction in storage and inference cost
What is the estimated energy consumption for training and inference?	Training: ~120 kWh; Inference: ~0.02 kWh per 1,000 inferences	Based on GPU logs and cloud usage data from Azure Sustainability Calculator
Are water usage or cooling impacts relevant and measured?	Yes – estimated 0.5 L/kWh based on region's WUE, totaling ~60 liters during training	Water estimate derived using regional water usage data from data center provider's ESG report

Template 2: AI Sustainability Risk Register

This risk register helps track, assess, and respond to sustainability-related risks specific to AI applications and infrastructure. Extend your enterprise's existing IT or AI risk register with these columns.

Risk Description	Impact (Low/ Med/ High)	Likelihood	Risk Response (Accept, Mitigate, Transfer, Avoid)	Owner	Mitigation Strategy	Status
Excessive compute energy usage exceeds sustainability thresholds	High	Medium	Mitigate	Green IT Lead	Model pruning, retraining with efficient architecture	Mitigated
Water consumption risk in selected data center location	Medium	Low	Transfer	Facilities Manager	Switch to region with WUE reporting	Open
Model overprovisioning leads to premature hardware refresh	High	High	Mitigate	Infra Ops Lead	Align model deployment with refresh cycle	In Progress

Template 3: Vendor Sustainability & Compliance Evaluation

Use this to evaluate AI vendors or cloud providers based on environmental sustainability criteria. Can be used as part of RFP processes or ongoing vendor assessments.

Risk Description	Impact (Low/ Med/ High)	Likelihood	Risk Response (Accept, Mitigate, Transfer, Avoid)	Owner	Mitigation Strategy	Status
Excessive compute energy usage exceeds sustainability thresholds	High	Medium	Mitigate	Green IT Lead	Model pruning, retraining with efficient architecture	Mitigated
Water consumption risk in selected data center location	Medium	Low	Transfer	Facilities Manager	Switch to region with WUE reporting	Open
Model overprovisioning leads to premature hardware refresh	High	High	Mitigate	Infra Ops Lead	Align model deployment with refresh cycle	In Progress

Template 4: AI Sustainability Incident Response Log

Use this log to document, analyze, and remediate incidents where AI system sustainability metrics deviate from established thresholds. This template supports continuous improvement and governance review.

Incident Summary

Issue Identified	**Energy use per inference for AI recommendation engine exceeded internal threshold of 0.05 kWh/inference for two consecutive weeks, reaching 0.08 kWh/inference (60% increase)**
Trigger Date	2025-04-30
Detection Method	Sustainability Monitoring System
Risk Priority	Medium
Escalation	Responsible AI Steering Committee notified
Integration into Continuous Improvement	Monitoring outcomes will be reviewed in the quarterly AI Lifecycle Governance Review. Lessons learned will inform updated sustainability design guidance and approval templates

Root Cause Analysis

Increased Model Complexity	**Model retraining added layers without optimization**
Lack of Post-Training Efficiency	No compression or distillation applied
Process Gap	Model deployed without sustainability review due to DevOps misalignment

Corrective Actions

Action	Owner	Deadline	Follow-up
Retrain and optimize model using quantization and pruning	AI Engineering Lead	May 15	Model audit report due May 20
Reintroduce sustainability checkpoint in MLOps pipeline	DevOps Manager	May 30	Compliance confirmation by internal audit
Conduct ML team refresher on energy-efficient design	Sustainability Program Lead	June 10	Training attendance and knowledge check
Re-baseline post-optimization energy/inference	Data Science Analyst	June 25	Dashboard updated, governance review
Update sustainability thresholds aligned to MLCommons	AI Governance Team	July 5	New thresholds integrated into alert system

Template 5: AI Sustainability Performance Scorecard

Use this scorecard to track key sustainability metrics across AI projects. This helps organizations maintain accountability and continuously improve.

Project / Model	Metric	Value	Target	Last Updated
Product Recommender	Energy per inference (kWh)	0.03	< 0.05	2025-05-01
Product Recommender	Model size (GB)	2.5	< 3.0	2025-05-01
NLP Chatbot	gCO_2e per inference	17	< 25	2025-05-01
Vision Model	GPU Utilization %	78%	> 60%	2025-05-01

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.