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

GENERATIVE AI TECHNOLOGY ECOSYSTEM

Generative AI Value Chain	**Technology**	**Data-center hosting**	**Open/closed source trained LL models**		**Use cases**		**AI-accelerated end-user services/devices**	
Examples	Google TPU v5e	Google Cloud	Gemini 1.5 Pro		Clinical trial design		Web portal, medical dashboards	
	NVIDIA H100	Azure	OpenAI GPT-4		Financial Services Chatbot		Embedded in Microsoft 365 Copilot	
	Amazon Trainium	AWS Europe	Meta LLaMA 3		Predictive maintenance & part design		Factory tablet dashboard, digital twins	
AI System Lifecycle	**Infrastructure** • GPU • TPU • CPU	**Hosting** • GPU • TPU • CPU	**Training** • Development • Production	**(Post) Training** • Fine tuning • Retraining	**Evaluation** • Quality • Safety • Impact	**Inference** • Use cases • Value add	**End-user function** • Prompt handling • Function handling • AI @ devi ce	**End-user devices** • Laptops • Mobile • Etc.
Sustainability factors and options	• Efficiency • Energy savings techniques • Support/software • GPU generations.	• Efficiency • Energy savings techniques • On-demand provisioning • Continuous rightsizing • Hardware lifecycle • Transparency	• Model architecture • Quantization • Hyperparameter optimization • Variations • Model lifecycle • Pre-trained models	• Efficiency/ reasoning • Pruning/compression • Modular training • Federated learning • Fixed precision training • Rendering	• Accuracy • Testing • Explainability • Responsible AI ◦ Ethical guardrails ◦ Environmental guardrails	• Business cases & trade-offs • Agents orchestration/ modularity • Maintenance • Performance monitoring • Impact monitoring	• Prompt handling • Conversations/agents • Embedded AI function ◦ SW & HW	• Energy saving techniques • AI chips in end-user devices • HW lifecycle/lifespan • Support/software
AI system inputs & outputs	**Minerals** • Bill of material • Sources • Virgin/circular • Waste	**Capacity** • Data Center • Operations	**Data - Generic** • Collection • Processing • Curation • Augmentation • Creation • IP /ownership	**Data - Specialized** • Collection • Labeling (conditions) • Proprietary	**Data - Test** • Principles • Values • Constraints • Regulations/policies (AI act) • Carbon limits	**Outcomes** • Productivity • Efficiency • Carbon savings • Cost • Revenue • Waste	**UX/UI** • Speed • Ease • Quality	**Minerals** • Bill of material • Sources • Virgin/circular • Waste
Infrastructure environmental impacts	E: Manufacturing	E: (DC) Operations	E: Training + Infra.	E: Training + Infra.	E: Eval./Test + Infrastruct.	E: Use phase + Infrastruct.	E: Dig. workplace + In fr a.	E: Manufacturing
	C: Embodied -3	C: Embodied -3	C: Embodied or attributed	C: Embodied or attributed.	C: Embodied or attributed	C: Embodied or attributed	C: Company attrib . -2	C: Embodied -3
	W: Manu factur ing	W: Cooling	W: Cooling	W: Cooling	W: Cooling	W: Cooling	W: N/A	W: Manufacturing
	Logistics & Packaging E: C: W:	E: Infra AI Supplier C: Embodied – 3 W: Supplier	**Infrastructure** Compute, Storage, Networking Orchestration, Monitoring, etc.)				E: Infra co. usi ng AI C:Company attrib. – 2 W: Operating company	Logistics & Packaging E: C: W:
	Excavation and/or circular materials	DC -Technology supplier (Scope 3) or DC operating company (Scope 2) allocated					Office (Scope2) Home (Scope3)	Excavation and/or circular materials

E = Energy
C = Carbon
W = Water

1 = Scope 1
2 = Scope 2
3 = Scope 3

Ecosystem Purpose and Use

Purpose:

The model emphasizes that AI's environmental impact extends far beyond the moment of use (inference), encompassing manufacturing, training, infrastructure, and end-of-life. The model can be used to better understand the stages of inputs and outputs of AI system capabilities and functionalities, and their relevant environmental sustainability impact areas.

Use this model to:

- Conduct lifecycle sustainability assessments for AI projects
- Align AI infrastructure procurement and data strategy with ESG goals
- Improve transparency and reporting for Scope 3 emissions
- Educate stakeholders on where sustainability and ethical risks are embedded in the AI pipeline
- Adjust the ecosystem elements for the specific AI models and the AI accelerated end-user devices in use and attribute the (additional) footprint stemming from AI adoption/integration.

What the model depicts:

This model describes different elements of and steps within the enterprise AI technology ecosystem, putting them in a logical order and indicating impact elements at each stage.

- **Gen AI value chain** -This top row shows the broad functional flow of Gen AI systems from infrastructure to user-facing applications. Three hypothetical examples are included in these rows.
- **AI System Lifecycle** -This row tracks major lifecycle stages for Gen AI systems, from hardware to model use.
- **Sustainability factors and options** -This row highlights technical decisions and optimization levers that influence energy/carbon/water impact
- **AI system inputs and outputs** -This row outlines the core resources and data flows required for building and running Gen AI systems, as well as the value-driving outcomes of the application/tool
- **Infrastructure environmental impacts** -The model's base depicts the foundational IT infrastructure and its relevant Energy (E), Carbon (C), and Water (W) footprints

Ecosystem model originally developed by Wiebren van der Zee, ABN AMRO Bank

Ecosystem Stages and Sustainability Factors

Model elements and their sustainability impacts:

Raw material and hardware manufacturing (not shown in ecosystem model)

Elements:

- Minerals: Feedstock sourcing (virgin vs. circular), bill of materials
- Manufacturing: Embodied carbon, energy, water

Sustainability factors: Scope 3 emissions from suppliers; energy and water use during production

Compute infrastructure and hosting

Elements:

- Data centers: Storage, compute, networking, orchestrations
- Suppliers vs. operators: Scope 2 vs. Scope 3 allocation
- Sustainability factors: Location-specific energy mix for data center operations; carbon (embodied or company attributed – Scope 2); water for cooling systems; capacity planning

Data for training and testing

Elements:

- Generic Data: Collection, processing, augmentation, IP
- Specialized Data: Labeling, proprietary datasets
- Test data: Validation against constraints, regulations (e.g., AI Act), ethical/principle-based filters

Sustainability considerations: Energy and infrastructure tied to data processing and hosting

Ecosystem Stages and Sustainability Factors - Continued

Model elements and their sustainability impacts:

Model training and inference

Elements:

- Open/closed source LLMs: GPT, Claude, LLaMA, etc.
- Functionality: Orchestration, modularity, performance optimization
- Prompt handling: Embedded AI functions in software/hardware

Sustainability factors: Energy (use phase + infrastructure); Carbon (embodied or attributed)

Use Cases

Elements:

- Business cases: Productivity, revenue, efficiency, carbon reduction, and waste
- Applications: Co-pilots, enterprise AI tools, embedded functions, agents

Sustainability factors: Trade-offs between performance and environmental efficiency

End-User Devices

Elements:

- Hardware: Laptops, mobile, AI PCs, cameras, vehicles (e.g., drones)
- AI performance features: On-device AI, modular inference
- UX/UI: Speed, ease, quality of AI-enabled experiences

Sustainability factors: Manufacturing Scope 3 emissions and water use; device lifespan and lifecycle management for circularity