Powering the Green Revolution with Energy-efficient AI Computing

Atos

Table of contents

Introduction	03
Deconstructing Energy Efficient AI Computing	04
An Overview of the Evolution	05
Game-changers: Industry Challenges and Opportunities	07
Key Innovations: Driving the Future of Green AI	08
Key Industry Statistics and Trends	08
Key Challenges in Adopting Energy-Efficient AI Computing	09
Growth Drivers for Faster Adoption of Energy-efficient	10
Key Components of Energy-efficient AI Computing	11
Solution Approach for Enterprises and GSIs	13
In summary	15
Way Ahead	16
Appendix	17

Introduction

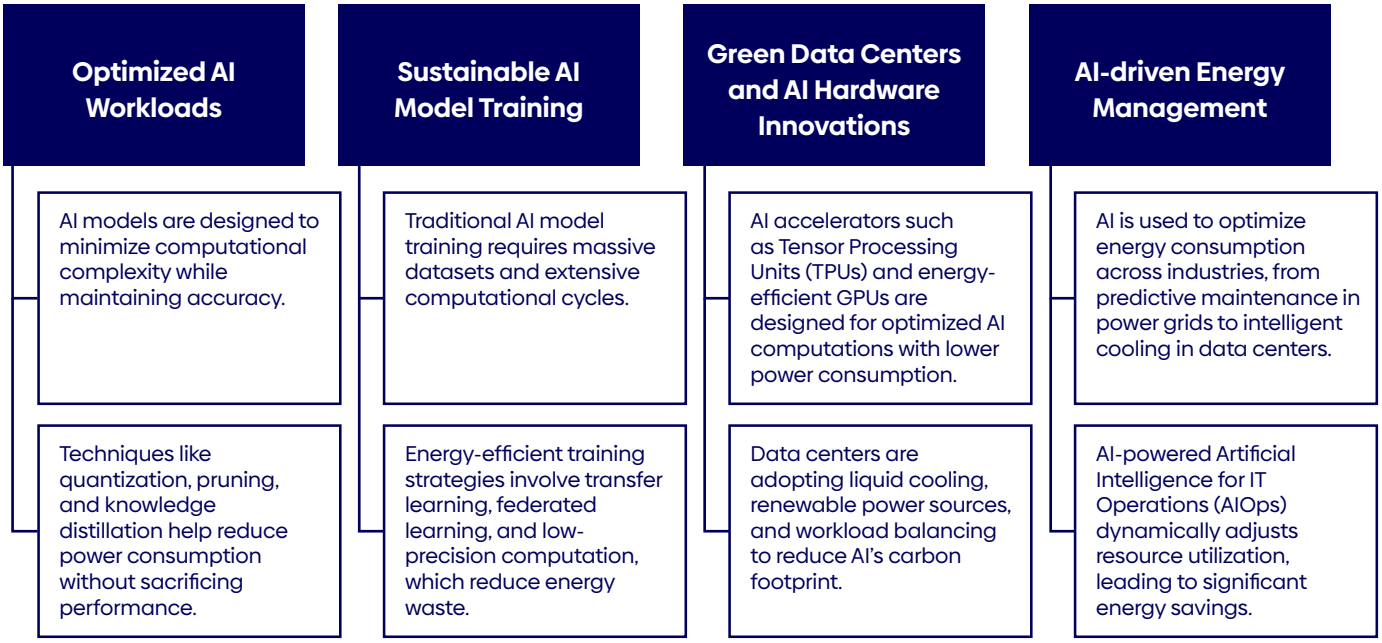
Artificial Intelligence (AI) is enabling businesses to make faster decisions, automate complex processes, and unlock new levels of efficiency. These transformations are occurring at an unprecedented pace, but with one key drawback — **energy consumption**. Training AI models, running large-scale inference systems, and maintaining AI-driven applications require enormous computational power, often leading to high electricity usage and increased carbon emissions.

Energy-efficient AI computing is an emerging discipline focused on reducing AI's energy footprint while maintaining or even improving performance. This approach combines innovative hardware, optimized algorithms, sustainable data center practices, and renewable energy integration to ensure the growth of AI aligns with the environmental responsibility shared by enterprises (industry customers), hyperscalers and CSPs, GSIs like Atos, and governments.

So, the responsibility doesn't lie with AI as a concept, but with the actors enabling and deploying AI. This refers to a collective obligation aligning to the theme of corporate social responsibility, regulatory compliance, and sustainability driven innovation.

With AI being a core enabler of digital transformation, energy-efficient AI is not just a technological necessity. It is a strategic imperative for businesses aiming to balance innovation with sustainability.

Deconstructing Energy Efficient AI Computing



The Green Revolution and AI

The concept of a **green revolution** in AI computing focuses on integrating energy-efficient AI technologies with mainstream IT infrastructure, cloud services, and industrial applications. Some of the key contributions of energy-efficient AI computing to sustainability include the following:

- **A reduced carbon footprint:** AI-powered optimization reduces unnecessary computational overhead, cutting down greenhouse gas emissions.
- **Integration with renewable energy:** AI can optimize solar, wind, and hydroelectric energy grids, ensuring maximum utilization of clean energy.
- **Improved resource management:** AI models are helping industries adopt circular economy principles, optimizing material usage and waste reduction.
- **Regulatory and ethical AI adoption:** Governments and enterprises are mandating energy-efficient AI practices, ensuring compliance with sustainability goals.

From healthcare and finance to smart cities and autonomous systems, industries have been embracing AI adoption, and as a result, **energy efficiency is no longer an afterthought but a necessity**. The next sections will explore how businesses can implement energy-efficient AI computing, the latest trends in Green AI, and a strategic roadmap for the future.

An Overview of the Evolution

The journey towards energy-efficient AI computing is an ongoing transformation, evolving from today's high-energy-demand AI models to a future where AI operates with minimal environmental impact.

This evolution can be broken down into three key phases:

1. Current state (today)
2. Interim advancements (near-term future)
3. Future vision (long-term sustainable AI)

1. Today: High-energy AI and Emerging Efficiency Measures

AI has become a core enabler of digital transformation, but its rapid expansion comes with significant energy demands. In this section, we survey the current landscape.

Challenges

Massive computational power requirements

Training state-of-the-art AI models like GPT-4 and Gemini requires thousands of high-performance GPUs, consuming enormous amounts of energy.

Data center energy consumption

AI workloads are processed in hyperscale data centers that contribute significantly to global electricity usage.

Carbon emissions

A single large-scale AI model can have a carbon footprint equivalent to multiple flights across continents.

Inefficient AI training and inference

Many AI models operate without energy optimization, leading to redundant computations.

Growing e-waste from AI hardware

Short lifecycles of AI chips and GPUs contribute to increasing electronic waste.

Current Efforts Toward Energy Efficiency

Early adoption of Green AI practices

Researchers focus on optimizing AI algorithms to minimize computational waste.

Specialized AI hardware (TPUs, FPGAs, neuromorphic chips)

Designed for energy efficiency, these are still not mainstream.

Renewable energy-powered data centers

Hyperscalers like AWS, Google, Microsoft and CSP like VMWare and DELL are investing in solar and wind-powered AI infrastructure.

Model compression and quantization

Reducing AI model size helps decrease energy consumption while maintaining accuracy.

Liquid cooling and heat recycling

Some data centers repurpose excess heat for secondary applications.

2. Interim Phase: Transitioning toward Scalable AI Sustainability

Over the next 3-5 years, we will see AI computing shifting toward more structured energy-efficient frameworks, driven by technological innovations and regulatory interventions.

Key Developments in the Interim Phase

Mainstream adoption of sustainable AI hardware
- Widespread use of low-power AI accelerators and photonics-based computing
- Advancements in neuromorphic computing that mimic brain-like efficiency

AI workload optimization becomes standard
- Adaptive AI models that dynamically adjust resource usage based on workload
- Edge AI becoming more prevalent, reducing reliance on cloud data centers

AI regulation and compliance maturity
- Governments enforce AI energy efficiency standards and carbon-neutral AI policies
- Digital sovereignty frameworks balance sustainable AI deployment with compliance needs

AI-driven energy grid optimization
- Enabling smart grids that dynamically allocate renewable energy to data centers
- AI-powered climate forecasting models for improved energy planning and resource management

Circular economy for AI hardware
- Recycling and repurposing AI chips and computing devices to reduce e-waste
- AI model-sharing platforms to prevent redundant model training

Expected Impact in the Interim Phase

- **Reduced energy intensity per AI model:** Efficient training and inference to reduce per-model electricity consumption
- **Carbon-neutral AI data centers:** Widespread use of renewable energy in AI infrastructure
- **Regulatory-backed AI sustainability metrics:** Reporting and optimization of AI's energy consumption

3. Future Vision: A Fully Sustainable AI Ecosystem

In another 7-10+ years, AI computing will be fundamentally transformed into an **ultra-efficient, carbon-neutral, and autonomous** system, seamlessly integrated with sustainable practices.

Future Impact of AI Sustainability

- **Zero-carbon AI models:** Net-zero emissions, driving the Green Revolution
- **Autonomous AI energy management:** Management and distribution of global energy resources
- **AI-powered sustainable innovation:** Enabling industries like manufacturing, logistics and smart cities to become carbon-neutral

The evolution of energy-efficient AI computing will be defined by continuous technological advancements, strategic regulatory interventions, and an industry-wide commitment to sustainability. While today's AI landscape is energy-intensive, the shift toward optimized AI models, green hardware, and sustainable data centers is accelerating.

Fun facts

A ChatGPT query is said to consume 10 times more electricity than a Google search.

Game-changers: Industry Challenges and Opportunities

The integration of AI across various industries presents both challenges and opportunities, particularly concerning energy consumption and sustainability. As AI adoption accelerates, understanding industry-specific issues and leveraging opportunities for energy-efficient AI computing become crucial.

At the forefront of tech: IT and Data Centers

Challenges

- **Escalating energy consumption** - AI workloads are projected to account for up to 20% of data center power consumption by 2028, increasing from 8% in 2024 (source: Statista). The significant energy demands of AI can lead to increased carbon emissions if it is not managed sustainably.

Opportunities

- **Adoption of energy-efficient hardware** - Specialized AI processors and optimizing hardware to reduce energy usage
- **Renewable energy integration** - Powering data centers with renewable sources to mitigate environmental impacts

Making it real: Manufacturing

Challenges

- **High energy usage** - AI-driven automation and predictive maintenance can lead to increased energy consumption (source: worldmetrics.org).

Opportunities

- **Energy waste reduction** – Up to 30% decrease in energy waste in manufacturing plants
- **Process optimization** – 15% boost in industrial processes' energy efficiency through AI apps

Powering the change: Energy Sector

Challenges

- **Grid management complexity** - Integrating AI increases the complexity of managing energy distribution and consumption (source: iea.org)

Opportunities

- **Operational cost savings** – Up to 25% savings on these costs
- **Enhanced grid reliability** – Up to 20% better grid reliability through AI-driven predictive maintenance and demand forecasting

Towards a better tomorrow: Transportation and Logistics

Challenges

- **Increased energy demand** - AI-powered systems for route optimization and autonomous vehicles may lead to higher energy consumption (source: sciencedirect.com)

Opportunities

- **Reduced fuel consumption** - Optimized logistics to reduce fuel usage and emissions
- **Predictive maintenance** - AI analytics reducing downtime and energy waste in transportation fleets

Caring for what matters: Healthcare

Challenges

- **Data processing demands** - AI applications in diagnostics and patient management require substantial computational power (source: sciencedirect.com)

Opportunities

- **Operational efficiency** - Streamlining hospital operations and reducing energy consumption through AI
- **Resource optimization** - Enhancing efficiency of medical equipment usage and energy management.

Key Innovations: Driving the Future of Green AI

Next-gen AI processors and Quantum AI computing	AI workloads shift to **quantum computing**, which drastically reduces energy consumption. **Biological and DNA-based computing** could emerge as low-energy alternatives.
AI operating on 100% renewable energy	AI systems become completely powered by **solar, wind, and kinetic energy harvesting**. AI data centers **self-optimize** based on real-time environmental conditions.
Self-regulating AI models	AI algorithms **auto-optimize** and reduce computational overhead dynamically. AI models trained on the **"Least Energy Path"** principle, ensuring the most efficient resource utilization.
AI's role in a sustainable world	AI actively contributes to climate action, optimizing **global carbon capture** and **enhancing biodiversity protection**. AI improves **precision agriculture, water conservation, and renewable energy storage.**
Policy-driven AI sustainability mandates	Global sustainability laws mandate that AI systems operate within strict energy efficiency thresholds. AI innovation is **governed by ethical AI energy frameworks** to ensure fairness, security, and sustainability.

Key Industry Statistics and Trends

According to statista.com, AI represented a global AI power use of 4.5 GW in 2024. That is approximately 8% of total DC power consumption. This is slated to grow to between 14-18.7 GW by 2028, i.e. ~20% of DC power usage.

Other key trends on our radar are as follows:

Energy efficiency improvements
AI can reduce energy use by ~20% in commercial buildings, cut energy waste by ~30% in manufacturing and lower greenhouse emissions by 4%.

Regulatory developments
Tech companies are increasingly asked to report data center energy use as AI booms, highlighting the need for transparency and accountability in energy consumption (source: guardian.com).

Key Challenges in Adopting Energy-Efficient AI Computing

The adoption of energy-efficient AI computing is influenced by several factors, spanning technological, economic, regulatory, and cultural dimensions. As AI innovation accelerates, addressing the energy demands associated with AI workloads is essential to align with sustainability goals and ensure long-term efficiency.

Balancing innovation with strategic carbon footprint reduction

- The rapid pace of AI advancements often prioritizes performance and scalability over energy efficiency.
- Organizations must strike a balance between delivering high-performance AI solutions and minimizing carbon emissions.
- The complexity of an AI model, such as LLMs like GPT-4, increases energy demand, making it difficult to maintain sustainability targets.

Trimming high AI, ML, and Gen AI resource consumption

- Training large AI models requires exponential computing power, with energy consumption doubling approximately every 3-4 months.
- A single AI training cycle (Ex: GPT-3) consumes an estimated 1,287 MWh, emitting 552 metric tons of CO_2, equivalent to 125 gasoline-powered cars driven for a year (Source: MIT Technology Review).
- AI inferencing also contributes significantly, accounting for 40-50% of AI-related energy consumption.

Developing efficient AI processing techniques

- Most AI models are trained and deployed on general-purpose GPUs, which are not optimized for energy efficiency specific to industry sectors.
- The need for hardware acceleration, such as TPUs, FPGAs, and neuromorphic computing, is growing but adoption has been sluggish due to inflated costs and limited software support.
- Lack of standardization in green AI computing architectures slows widespread adoption.

Containing costs and improving ROI

- Energy-efficient AI hardware and optimization techniques often come at a high upfront cost, delaying ROI realization.
- Companies struggle to justify sustainability-focused investments when short-term business KPIs prioritize speed and performance over efficiency.
- Lack of monetization models for energy-efficient AI creates low financial incentives for businesses to prioritize sustainability.

Raising awareness and promoting a culture of AI sustainability

- Many enterprises have not yet prioritized AI sustainability, resulting in low awareness and strategic alignment.
- AI developers and data scientists often lack training in green AI principles, leading to inefficient model development.
- Sustainability KPIs are rarely integrated into AI performance metrics, making energy efficiency an afterthought.

Embracing technology maturity and integration

- Most AI-driven energy optimization solutions are still in their infancy and lack widespread adoption.
- Integrating AI energy efficiency measures into existing cloud and data center infrastructure is complex as well as resource intensive.
- AI compute workloads still depend on legacy systems heavily, which were not designed for power efficiency.

Complying with regulatory requirements and digital sovereignty

- Emerging AI energy regulations such as the EU AI Act and California's AI sustainability mandates require enterprises to report energy usage, adding compliance complexity.
- Digital sovereignty concerns may restrict data and AI workloads to specific geographies, impacting energy-efficient deployment options like using offshore green data centers.
- Governments are increasing scrutiny on AI-driven power usage, with some regions even considering taxation on excessive AI energy consumption.

Broader industry and ecosystem barriers

- AI supply chains are heavily reliant on energy-intensive semiconductor manufacturing, which negates some efficiency gains.
- Many cloud providers do not yet offer transparent energy usage tracking for AI workloads, making it difficult for businesses to measure and optimize energy consumption.
- The lack of cross-industry collaboration on AI energy efficiency standards results in fragmented solutions and limited scalability.

Growth Drivers for Faster Adoption of Energy-efficient AI Computing

In this section, we take a look at some of the initiatives and incentives that are driving organizations to prioritize and adopt energy-efficient AI computing for tangible and transformational benefits.

1. Economic incentives and cost reductions

- AI energy costs are projected to rise 3x by 2030, making energy efficiency a financial imperative for enterprises.
- Adoption of AI-specific accelerators like TPUs, FPGAs, and custom AI chips can reduce power consumption by up to 90% in some workloads.
- Power-efficient AI deployment strategies such as Edge AI, federated learning, and model pruning can significantly lower costs.

2. Regulatory and compliance pressures

- Governments and regulatory bodies push for AI energy consumption disclosures.
- Carbon taxation on high-energy AI workloads may accelerate the shift towards energy-efficient computing.
- Companies aligning with net-zero commitments like Microsoft's AI sustainability pledge are driving industry-wide change.

3. Advances in AI model optimization

- Smaller, more efficient AI models are being developed to maintain accuracy while reducing compute demands.
- Quantum computing and neuromorphic chips show promise in reducing AI energy requirements.
- Techniques like knowledge distillation, low-rank adaptation (LoRA), and sparsity optimization improve AI efficiency.

4. Shift towards renewable energy in AI computing

- Hyperscalers like AWS, Google, and Microsoft and CSPs like VMWare are investing in 100% renewable-powered data centers to offset AI's energy impact.
- The transition to solar, wind, and hydro-powered AI computing can lower emissions by up to 75%.
- AI workload scheduling based on real-time energy availability can maximize the use of green power sources.

5. AI-driven energy management solutions

- AI can be leveraged to optimize data center energy consumption, leading to 20-30% reductions in power usage.
- AI-powered dynamic workload balancing and cooling optimization significantly improve energy efficiency in hyperscale data centers.

6. Industry collaborations and standardizations

- Initiatives like Green Software Foundation (GSF), Climate Neutral Data Center Pact, and the Energy Star for AI initiative are pushing for sustainable AI development.
- AI energy efficiency benchmarks like MLPerf Green and AI Sustainability Index are introduced to track and compare energy-efficient AI models.

7. Growing consumer and investor pressure on sustainability

- According to IBM, 80% of global consumers prefer brands that prioritize sustainability.
- Environmental, social and governance-driven investment strategies are pushing enterprises to adopt greener AI solutions to maintain investor confidence.
- Large enterprises are implementing sustainability-linked AI procurement policies, requiring vendors to comply with energy efficiency standards.

With these global changes and the continuous push for sustainability, we are confident that more and more organizations will embrace this for a future-ready business strategy. However, many seek guidance to help them on this journey. What are the different types of solutions available for swifter energy efficient AI computing? How soon can they expect results and larger profit margins?

Let's take a look.

Key Components of Energy-efficient AI Computing

Energy-efficient AI computing is built upon multiple interconnected components, spanning hardware, software, data management, and regulatory frameworks.

These components collectively enable AI to operate with minimal energy consumption while supporting sustainability initiatives and business objectives. We have identified six types here.

1. Energy-efficient AI model design

Algorithm Optimization

- **Model pruning and quantization** – Reducing model complexity by eliminating redundant parameters and using lower-precision arithmetic, like from FP32 to INT8, to reduce energy consumption
- **Sparse computation** – Leveraging sparsity in neural networks to minimize unnecessary computations
- **Federated learning** – Decentralized AI model training that reduces energy spent on data transfers and centralized processing
- **Knowledge distillation** – Training smaller, energy-efficient AI models by transferring knowledge from larger models

Low-power Inference and Training

- Efficient use of **edge AI** and **tinyML** to reduce reliance on power-hungry cloud infrastructure
- Utilizing **adaptive AI models** that dynamically adjust computational intensity based on workload requirements

2. Energy-efficient AI hardware and infrastructure

Next-gen AI Chips

- **Custom-built AI processors like Tensor Processing Units (TPUs), neuromorphic chips, and low-power GPUs** - Optimizes energy efficiency for AI workloads
- **Emerging photonic computing and analog AI systems** - Slated to reduce power needs

Green Data Centers

- **Liquid cooling and immersion cooling** – Reduces reliance on traditional air cooling systems, cutting power consumption significantly
- **Renewable energy integration** – Powering AI workloads with solar, wind, or geothermal energy
- **Workload orchestration** – Intelligent scheduling of AI tasks based on power availability and efficiency metrics

Edge AI and Distributed Computing

- **Shifting AI workloads to edge devices** - Reduces reliance on large cloud data centers, saving energy
- **AI models deployed in 5G and IoT ecosystems** - Minimizes data transmission energy costs

3. Responsible data and AI practices

Data Lifecycle Management

- **Data compression, deduplication, and intelligent storage strategies** - Minimizes energy spent on redundant data
- **Cold storage strategies** - Ensures long-term data retention to reduce energy-intensive active storage
- **Green data labeling and processing** – AI-driven automation for efficient annotation and reduced computational load

Ethical AI and Fairness

- Ensuring **AI model development aligns with sustainability goals** without amplifying biases or environmental risks
- Prioritizing **explainability and interpretability** in AI models to prevent unnecessary re-computation

4. Regulatory compliance and digital sovereignty

Green AI Policies and Standards

- Adhering to **ISO 50001 Energy Management** standards in AI-driven data centers
- Compliance with **EU Green Digital Transformation** mandates and US **AI energy efficiency regulations**
- Industry-driven initiatives such as the **Green Software Foundation** focusing on low-energy AI workloads

Digital Sovereignty and Data Localization

- Balancing energy-efficient AI with **regional data storage laws** and **AI TRISM** policies
- AI model governance frameworks ensuring sustainability does not compromise security and sovereignty

5. AI for environmental sustainability

AI-powered Energy Optimization

- **AI-driven grid management** - Optimizes renewable energy distribution
- **Predictive analytics** for energy-efficient urban planning and smart cities

AI in Climate Tech and Resource Management

- AI-assisted **carbon capture and emission tracking** for enterprises
- Smart AI-based **water conservation and waste reduction** techniques in industries

6. Social responsibility and industry innovation

Sustainable AI Innovation

- Encouraging **open-source AI models** to prevent energy waste from repeated training of proprietary models
- Investments in **circular economy models** for AI hardware recycling

Collaborations and Public-Private Partnerships

- Tech companies, academia, and governments working together to establish **Green AI benchmarks**
- AI-driven sustainability initiatives in **manufacturing, logistics, and agriculture**

Energy-efficient AI computing is a convergence of technological advancements, responsible AI practices, regulatory mandates, and sustainability-driven innovations.

By focusing on optimized AI workloads, next-gen energy-efficient hardware, responsible data management, and adherence to green regulations, the tech industry can significantly reduce AI's carbon footprint.

The future of AI is not just about being more intelligent. It is about being more sustainable.

Solution Approach for Enterprises and GSIs

Dos:

- Optimize AI models using pruning, quantization, and low-precision computing.
- Use AI-specialized hardware like TPUs, FPGAs, and neuromorphic computing for efficiency gains.
- Deploy AI in renewable-powered data centers or leverage carbon offsetting strategies.
- Implement automated workload scheduling to minimize peak-hour energy usage.
- Ensure compliance with sustainability regulations in AI computing.
- Foster a culture of AI energy efficiency awareness and governance.

Don'ts:

- Don't deploy AI models without considering their energy and carbon impact.
- Reconsider training large AI models from scratch when transfer learning is viable.
- Don't overuse cloud AI resources without monitoring energy efficiency.
- Never ignore AI energy consumption tracking and sustainability KPIs.
- Don't delay investments in energy-efficient AI hardware and software optimizations.
- Don't assume AI sustainability is only an IT or cloud provider responsibility. It requires cross-functional enterprise commitment.

[I] AI model, algorithm optimization, hardware, and infrastructure efficiency

1

Adopt Efficient Model Architectures

- Use smaller, efficient AI models like DistilBERT, LLaMA, and LoRA fine-tuning vs. computationally heavy models
- Leverage knowledge distillation, model pruning, quantization to reduce computational needs
- Use transfer learning instead of retraining large models from scratch.

2

Optimize Model Training and Inferencing

- Use structured sparsity to remove redundant computations
- Prefer low-precision computing like FP16 and INT8 over standard floating-point ops
- Deploy edge AI and federated learning to reduce cloud data transfers and improve efficiency.

3

Adopt Specialized AI Hardware

- Use AI-optimized processors such as TPUs, FPGAs, neuromorphic with higher compute efficiency per watt
- Deploy liquid cooling, dynamic power scaling in data centers to reduce energy waste
- Prioritize hyperscale green data centers powered by renewable energy sources.

4

Leverage AI Workload Scheduling for Energy Efficiency

- Use dynamic AI workload orchestration to run intensive AI tasks when renewable energy is most available
- Implement serverless AI inference to optimize resource allocation and avoid over-provisioning.

[II] AI governance, compliance, digital sovereignty and sustainable data lifecycle management

1

Reduce Unnecessary Data Processing

- Apply intelligent data retention policies to store only essential data
- Use data deduplication and compression techniques to minimize storage needs
- Shift towards cold storage for inactive AI datasets.

2

Improve Data Transfer Efficiency

- Adopt edge AI to reduce network traffic and processing at centralized data centers
- Use energy-aware networking protocols to optimize data movement.

3

Adopt Responsible AI and Sustainability Metrics

- Integrate AI energy efficiency KPIs in AI development and operations
- Implement ESG-aligned AI governance frameworks that track energy consumption and carbon emissions.

4

Ensure Regulatory Compliance

- Align AI deployments with regional AI sustainability regulations like the EU AI Act and California Energy Standards
- Partner with Green AI organizations like Green Software Foundation and Climate Neutral Data Centers for compliance validation.

[III] Financial and economic considerations with organizational and cultural transformation

1

Emphasize Digital Sovereignty and Ethical AI

- Ensure AI workloads comply with local energy and data sovereignty laws
- Avoid excessive AI-driven surveillance or invasive data practices that conflict with sustainability ethics.

2

Develop Sustainable AI Cost Models

- Encourage Pay-as-You-Go (PAYG) AI computing models to align costs with energy usage
- Introduce carbon credit-based incentives for AI models with lower energy footprints.

3

AI Carbon Offsetting and Sustainability Awareness

- Collaborate with carbon-neutral AI infra providers
- AI-driven sustainability reports for ESG goals
- Train AI developers and data scientists in green AI best practices
- Promote sustainability focused AI R&D.

4

Encourage Industry Collaboration and Standardization

- Work with cloud providers, chip manufacturers, and regulatory bodies to create industry-wide AI efficiency standards
- Support open-source energy-efficient AI libraries and toolkits.

In summary

AI's energy footprint is a growing concern	• AI workloads are energy-intensive, especially GenAI models like GPT, BERT, and DALL·E. • Without intervention, AI energy consumption could surpass entire nations' electricity usage.
Efficiency must be embedded at every layer	• Optimizing AI models (pruning, quantization, distillation) reduces computational power needs. • Energy-efficient AI chips (TPUs, FPGAs, neuromorphic computing) lower power consumption. • Green cloud computing using renewable energy is key to sustainable AI operations.
Responsible Data and AI Practices are critical	• Data lifecycle optimization minimizes redundant processing and storage costs. • Federated learning and edge AI reduce unnecessary cloud computing. • Ethical AI frameworks ensure compliance with ESG goals and regulations.
Regulatory compliance and Digital Sovereignty must be addressed	• AI sustainability is increasingly governed by global and regional regulations like EU AI Act and SEC ESG disclosures. • Organizations must align AI deployments with energy efficiency and carbon neutrality standards.
Sustainable AI is a business advantage	• Energy-efficient AI computing reduces operational costs and enhances long-term AI scalability. • Enterprises prioritizing AI sustainability will gain a competitive edge in compliance, innovation and brand reputation.
The road ahead	• Future AI innovations must balance performance, cost and environmental responsibility. • Cross-industry collaboration between hyperscalers, enterprises, regulators and GSIs will drive scalable green AI solutions. • AI sustainability metrics should be a core component of AI governance strategies.

Way Ahead

The convergence of AI, energy demands, and the green revolution presents a critical opportunity for enterprises and technology providers to drive sustainable AI innovation. As AI adoption grows across industries, the energy consumption of AI workloads, especially large-scale ML and GenAI models, has become a significant concern. Addressing this challenge requires a comprehensive approach that integrates AI model efficiency, energy-optimized hardware, sustainable cloud infrastructure, responsible AI governance, and regulatory compliance.

By adopting energy-efficient AI computing strategies, organizations can minimize environmental impact while maximizing AI-driven business value. Investing in hardware acceleration, intelligent workload scheduling, data lifecycle management, and carbon-conscious AI practices will not only reduce costs but also enhance operational resilience and digital responsibility.

GSIs and enterprises must recognize that energy-efficient AI is not just a technological imperative but a business and ethical necessity. Companies that proactively embed sustainability into AI development and deployment will gain a competitive edge, comply with emerging regulations, and contribute meaningfully to the global net-zero agenda.

Atos is a global leader in digital transformation and European number one in cybersecurity, cloud and high-performance computing. A pioneer in decarbonization services and products, it focuses specially on sustainability, energy, and AI. Collaborating with a GSI like Atos will help organizations craft tailored cloud strategies that embed sustainability into their core operations.

Together, we can shape a greener, more efficient, and more responsible future.

The future of AI must be intelligent, responsible, and energy-aware ensuring continued innovation without compromising planetary sustainability.

>> Learn more about how Atos can help you get started on your journey to energy-efficient AI computing. Contact our team of AI and sustainability experts today.

>> Find out how Atos solutions are making a difference in the world around us with cutting-edge solutions across Smart Platform, Gen AI, Cybersecurity, Data and AI Platforms, Industry Cloud and IoT.

Appendix

Comparative Analysis: AI model architectures, hardware, and software from an energy-efficiency perspective

Category	AI type / Component	Nature of Purpose / Task	Resource Consumption (Compute, Power, Data Center Energy)	Sustainability Factors (CO2 Footprint, Emissions, E-Waste)	Training Options	Data Processing Techniques
AI Model Architecture	**Large Language Models (LLMs)** GPT, BERT, LLaMA)	NLP, text generation, summarization	**High compute & power demand** (~10,000 GPUs for training), Large-scale cloud clusters	**High carbon footprint**, ~500 tons CO2 GPT-3 training	**Centralized cloud-based training**, fine-tuning via APIs	Transformer-based deep learning
	Convolutional Neural Networks (CNNs)	Image recognition, object detection	Moderate compute (depends on complexity)	**Lower emissions vs LLMs**, but still significant	Pre-trained models + transfer learning	Feature extraction deep learning
	Recurrent Neural Networks (RNNs) & LSTMs	Time-series forecasting, speech recognition	Moderate compute, memory-intensive	**Moderate CO2 footprint**, can be reduced via pruning	Pre-training + online learning	Sequence-based memory processing
	Graph Neural Networks (GNNs)	Fraud detection, social networks, recommendation systems	High compute demand for large graphs	High emissions but optimized via sparse processing	Specialized graph databases for training	Graph-based parallel processing
AI Hardware	**GPUs (Nvidia A100, H100, AMD Instinct MI250)**	AI/ML model training & inference	High power draw (~300 W per chip), massive parallelism	High emissions unless run on green data centers	Tensor-based acceleration	SIMD (Single Instruction Multiple Data)
	TPUs (Google tensor Processing Units)	AI inference, optimized for deep learning	More energy-efficient vs GPUs (~40% less power)	Lower CO2 footprint, designed for efficiency	Pre-trained models deployed via TPU cloud	Matrix multiplications
	Neuromorphic Chips (Intel Loihi, IBM TrueNorth)	Brain-like AI processing, ultra-low power	Extremely low energy (~10W), event-driven computing	**Minimal emissions**, sustainable AI chip	On-device self-learning	Spike-based computing
	FPGAs (Field Programmable Gate Arrays)	AI model acceleration, real-time inference	Power-efficient, but less flexible than GPUs	Low CO2 footprint, reprogrammable	Custom AI model optimizations	Parallel computing
AI Software & Frameworks	**TensorFlow PyTorch, JAX**	AI/ML model development	Runs on CPU, GPU, or Tpu - depends on deployment	**Sustainability depends on hardware & optimization**	Cloud-based or local training	Stochastic gradient descent
	ML Ops Platforms (AWS SageMaker, Azure ML, Google Vertex AI)	ML lifecycle management, monitoring & governance	Cloud-based, energy usage varies by instance type	**Can be optimized using renewable cloud instances**	Cloud-hosted training pipelines	AutoML, Hyperparameter tuning
	AI Inference Engines (ONNX, TensorRT, OpenVINO)	AI model deployment & inference	Edge AI-friendly, optimized for low-power inference	**Low emissions if used on optimized hardware**	Optimized for inference	Model compression, quantization
AI Model Architecture	**Large Language Models (LLMs) (GPT, BERT, LLaMa)**	Medium (depends on energy source)	Moderate (Cloud-hosted, often US/EU servers)	$$$	Finance, Healthcare, Retail, Customer Support	OpenAI, Google DeepMind, Meta, Microsoft
	Convolutional Neural Networks (CNNs)	High (if using efficient chips)	High (on-prem & edge possible)	$$	Manufacturing, Healthcare, Security	Nvidia, Intel, Google, AWS
	Recurrent Neural Networks (RNNs) & LSTMs	High (if optimized with pruning)	Moderate	$$	Healthcare, IoT, Telecom	Google, Meta, IBM
	Graph Neural Networks (GNNs)	Medium (enegy-intensive training)	Moderate	$$	Cybersecurity, Social Media, Pharma	Microsoft, AWS, DeepMind

Category	AI type / Component	Nature of Purpose / Task	Resource Consumption (Compute, Power, Data Center Energy)	Sustainability Factors (CO2 Footprint, Emissions, E-Waste)	Training Options	Data Processing Techniques
AI Hardware	**GPUs (Nvidia A100, H100, AMD Instinct MI250)**	Medium (green AI data centers help)	Low (Mostly US/ EU-made)	$$$	AI Research, HPC, Cloud AI	Nvidia, AMD
	TPUs (Google Tensor Processing Units)	High (optimized for cloud AI)	Low (Google-dependent)	$$	Cloud AI, Search, Vision AI	Google
	Neuromorphic Chips (Intel Loihi, IBM TrueNorth)	Very High (lower energy AI)	High (local & sovereign AI possible)	$$	Robotics, Edge AI, AI powered Iot	Intel, IBM
	FPGAs (Field Programmable Gate Arrays)	High (lower than GPUs)	High (Edge AI possible)	$$	Autonomous Vehicles, Financial AI	Xilinx, Altera (Intel)
AI Software & Frameworks	**TensorFlow, PyTorch, JAX**	Medium	Moderate (Cloud-heavy)	$$	AI R&D, Data Science, Cloud AI	Google (TF), Meta (PyTorch), DeepMind (JAX)
	ML Ops Platforms (AWS SageMaker, Azure ML, Google Vertex AI)	High (depends on cloud sustainability)	Moderate	$$	Enterprise AI, FinTech, Manufacturing	AWS, Microsoft, Google
	AI Inference Engines (ONNX, TensorRT, OpenVINO)	High (optimized for performance)	High (Edge AI & Sovereign AI possible)	$	Healthcare AI, IoT, Smart Cities	Intel, Nvidia, OpenAI

Insights and analysis

AI model architecture trends

- LLMs like GPT models are highly energy-intensive, requiring large-scale data center resources, whereas CNNs and RNNs offer better efficiency for specific applications.
- Emerging models such as Graph Neural Networks (GNNs) can optimize energy-efficient AI processing for large, complex datasets.

AI hardware optimization

- Neuromorphic chips and FPGAs provide extremely low-power AI computation, making them ideal for Edge AI and sustainable AI.
- GPUs remain the dominant choice but are power-hungry, requiring sustainable energy solutions in AI data centers.
- TPUs are 40% more energy-efficient than GPUs, making them a more sustainable option for cloud-based AI inference.

AI software and ML operations

- ML platforms like AWS SageMaker and Azure ML are integrating energy-aware AI deployment for enterprises.
- ONNX and OpenVINO allow for low-power AI inference, promoting AI efficiency for industrial and IoT applications.

Sustainability and ESG compliance

- AI hardware sustainability is improving with energy-efficient silicon. Consider Nvidia's H100 with 30% lower emissions.
- AI models must adopt Green AI principles optimizing training, data processing, and inference efficiency.
- Renewable energy-powered AI data centers will be a major trend for hyperscalers, CSPs and enterprises adopting GenAI.

Digital sovereignty and regional considerations

- European AI regulations (EU AI Act) and digital sovereignty concerns drive on-prem AI adoption instead of cloud dependence.
- AI deployment must consider cross-border data laws to ensure compliance and energy efficiency.

About Atos

Atos is a global leader in digital transformation with c. 78,000 employees and annual revenue of c. € 10 billion. European number one in cybersecurity, cloud and high-performance computing, the Group provides tailored end-to-end solutions for all industries in 68 countries. A pioneer in decarbonization services and products, Atos is committed to a secure and decarbonized digital for its clients. Atos is a SE (Societas Europaea) and listed on Euronext Paris.

The purpose of Atos is to help design the future of the information space. Its expertise and services support the development of knowledge, education and research in a multicultural approach and contribute to the development of scientific and technological excellence. Across the world, the Group enables its customers and employees, and members of societies at large to live, work and develop sustainably, in a safe and secure information space.

Find out more about us

atos.net
atos.net/career

Let's start a discussion together

Atos is a registered trademark of Atos SE.
May 2025. © Copyright 2025, Atos SE.
Confidential Information owned by Atos group, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval of Atos.

106228 - EM - ST