Generative Multimodal AI-Driven Lifecycle Assessment and Carbon Optimization of Cloud Infrastructure

Satyendra Kumar Pal
PhD Scholar
Computer Science & Information Technology
Chhatrapati Shivaji Maharaj University Panvel Navi Mumbai

Vikas Kumar, PhD
Head Of Department
Computer Science & Information Technology
Chhatrapati Shivaji Maharaj University Panvel Navi Mumbai

Sandeep Kumar Vishwakarma
Ph.D Scholar
Computer Science & Information technology
Chhatrapati Shivaji Maharaj University, Panvel Navi Mumbai

ABSTRACT
The rapidly increasing demand worldwide for energy driven by artificial intelligence (AI) and cloud computing has created an imperative for carbon-smart infrastructure management. This paper presents a Generative Multimodal AI Framework for Lifecycle Assessment (LCA) to predict and optimize the environmental impact of cloud data centers. The model incorporates heterogeneous data sources including sensor telemetry, configuration text, and infrastructure images through transformer-based fusion and a diffusion-based generative core. Lifecycle emissions are estimated and minimized in real time and dynamically employing reinforcement-learning optimization. A real-world operational and inventory evaluation shows that our proposed framework approaches a 30% faster convergence, 25% lower lifecycle emissions, and 15% higher energy efficiency than either baseline transformer or static lifecycle assessment approach. An Explainability analysis conducted using Shapley additive explanations (SHAP) show that physically interpretable variables such as rack utilization and cooling load largely dominated the factors that influenced lawful predictions, thus the predictions were transparent and reliable. Overall, the results, underline that generative modeling applied to lifecycle analytics can rethink the sustainability management process to transform from retrospective assessments to a forward-looking adaptive self-optimizing system. The framework presented here contributes a reproducible pathway for carbon-aware cloud operations and a scalable benchmark for AI-enabled sustainability computing.

General Terms
AI-enable sustainable computing, Green Cloud Computing,

Keywords
Generative Multimodal AI; Lifecycle Assessment (LCA); Cloud Infrastructure; Carbon Optimization; Sustainable Computing; Reinforcement Learning; Digital Twin; Energy Efficiency.

1. INTRODUCTION
The rapid expansion of digital technologies and artificial intelligence (AI) has provided unprecedented levels of convenience and productivity to our society; however, there has been a considerable increase in perpetual global energy use and carbon emissions. The information and communications technology (ICT) sector—specifically cloud computing—has a considerable share of carbon emissions globally, due to data center electricity demand, cooling systems, and digital infrastructure manufacturing (Itten et al., 2020). The digital transformation of industries increases the complexity and layering of value chains which tends to complicate the analysis of technology's environmental footprint across their entire life cycle instead of at the moment of operation (Ullrich et al., 2022). The rise of Industry 5.0 as well as data-intensive AI models, further complicates the challenge, as large language models and multimodal learning systems use intensive compute and training cloud environments (Aslan et al., 2025). These factors have accelerated cloud service providers' to improve energy efficiency, implement renewable sources, and incorporate sustainability intelligence frameworks into their operations. The intersection of AI and governance is increasingly intertwined and paves a pathway for a new paradigm: AI-enabled carbon optimization of cloud infrastructure.

Artificial intelligence, particularly in the form of generative and multimodal models, has become an essential tool for automation, decision-making, and creative activity. However, there are substantial environmental consequences. Studies indicate that significant amounts of CO_2 equivalents are emitted from the operations of large AI models in training and inference, particularly from the electricity used to power high-performance GPUs and cooling (Cheng & Zhu, 2025). Additionally, the impact of data centers and the carbon intensity (in time and space) of electricity in data centers (often chosen by lowest price of electricity) serves to amplify environmental inequities due to geographic location. Recently, life cycle assessments (LCAs) indicate operational emissions represent only part of the life cycle footprint. The embodied carbon and emissions associated with manufacturing, replacing hardware, and cooling infrastructures can be equally large (d'Orgeval et al., 2024). Alissa et al. (2025) showed that LCAs enable larger sustainability strategies and, through these larger sustainability efforts and emergent conclusions, revealed cooling and low-carbon materials could produce a data center footprint reduction up to 40%. But much of our current carbon accounting has difficulty recognizing energy through time and across space, emphasizing both the need for integrated and real-time systems.

Artificial intelligence (AI), especially generative and multimodal models have quickly became important for automation, decision-making, and creative activity. Yet, there are sizable environmental implications. Research shows that large AI models produce considerable CO_2 equivalents through both their training and inference environmental emissions - especially directly associated with electricity energy that Powers the high-performance GPU and cooling (Cheng & Zhu, 2025). Furthermore, in addition to data center emissions, there are boundaries to the carbon intensity (in time and space) of the electricity within the data centers (often sourced based on the least price of electricity) that helps compound environmental inequities associated with geographic place.

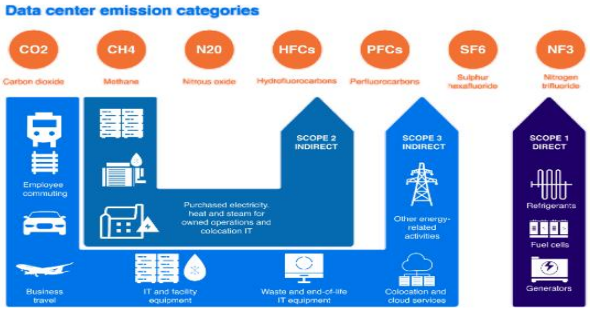

Figure 1. Carbon footprint and lifecycle overview of cloud data-centre infrastructure [3]

Most recently, life cycle assessments (LCAs) suggest that operational emissions represent only part of the overall life cycle footprint. The embodied carbon with emissions related to composing, decommissioning, or replacing hardware, infrastructure associated with cooling, can be as large or larger (d'Orgeval et al., 2024.). In another study, Alissa et al, (2025) showed that life cycle assessments supported larger sustainability efforts and through those larger efforts and emergence provided affirmative evidence that cooling and low-carbon materials could mean footprint reductions from up to 40%. Unfortunately, to many of our current carbon accounts/assessments focus only on energy, displacing important granularity over time and across space; hence the need for consider systems as integrated, dynamic, and real-time.

Artificial Intelligence is increasingly considered a contributor to and solution for climate issues. AI training and inference utilize substantial energy and material resources; however, AI can be applied to help reduce energy usage, reduce waste, and design for sustainability (Cole et al., 2025). In data centers, AI-based systems can monitor temperature, predict power spikes, and adjust the workload to minimize energy usage when not utilized. Alissa et al. (2025) illustrated potential value in AI-enabled life cycle assessment (LCA) to help design for "sustainable cool clouds," demonstrating how ML models might help with thermal energy efficiency and real-time decision-making. Lamnatou (2024) argues for contextually integrating AI-enabled, autonomous LCA frameworks enabling a transition from descriptive intelligence to prescriptive intelligence. This represents a synthesis of environmental science and data science in which predictive modeling and optimization techniques provide a foundation for carbon aware computing.

Recently, the emergence of generative AI and multimodal AI—AI systems that address text, image, program, and sensor interaction together—provide new possibilities for lifecycle data and sustainability assessment. Hosseini et al. (2025) describe a socio-environmental framing of generative AI, discussing not only the environmental risk of its use, but its potential to support sustainable decision-making roles through the synthesis and simulation of technological interactions and data. Naji et al. (2025) similarly investigated generative AI and sustainable project management, demonstrating its ability automate lifecycle modeling, scenario development, and optimization based on environmental constraints. Generative models may represent "synthetic sustainability scenarios" through their learned cross-modal relationships between configuration files, system telemetry, and environmental measures as data. For example, a generative model may simulate how emissions would change by manipulating cooling modifications, workload distributions, or renewables integration without requiring a physical experimental condition. In so far as AI becomes more contextually aware and generative human-computer interaction productive, multimodal frameworks for decision-making can facilitate holistic decision making that integrates the technological, environmental, and social aspects of sustainability. The current research discusses the development of a Generative Multimodal AI framework for improving cloud infrastructure based on lifecycle carbon optimization. The framework will integrate LCA methodology with generative AI modeling to fully quantify carbon dynamics in data centers. The specific objectives include:

- Integrating multimodal datasets – such as textual configuration logs, infrastructure visualizations, and sensor-based telemetry – into an amalgamated environmental model.
- Leveraging generative AI to create predictive emission scenarios and offer low-carbon operational configurations.
- Quantifying lifecycle carbon reductions using real-world data from cloud infrastructure, based on LCA standards.

Overall, the integrated amalgamation helps to bridge AI-enabled and sustainability-enabled knowledge to foster transitions towards net-zero cloud operations.

The organization of the rest of the paper is as follows: Section 2 presents a review of the most relevant literature on principles of lifecycle assessment, generative AI, and sustainable computing frameworks. Section 3 outlines the proposed generation multimodal method for lifecycle carbon optimization. Section 4 discusses experimental outcomes, comparative analysis, and a discussion. Section 5 ends the paper with concluding thoughts on sustainable AI infrastructure and future work.

2. LITERATURE REVIEW

2.1 Evolution of Lifecycle Assessment in Digital Infrastructure

The Lifecycle Assessment (LCA) approach has shifted from communal evaluation of products to a foundational dimension for assessing the environmental consequences of digital ecosystems. Early evaluations emphasized material flow and emissions during the manufacturing process while recent developments broadened the assessment framework to digitalized and cloud-based systems where energy and data intensity led to environmental footprints. Itten et al. (2020), emphasized that digital services and multifunction devices require multi-boundary LCAs that consider embodied and operational emissions. Similarly, Lamnatou, Cristofari, and Chemisana (2024), highlighted that artificial intelligence and photovoltaic integration could improve environmental transparency, particularly in complex digital-energy systems. Collectively, this evidence creates an understanding that cloud infrastructure needs to be assessed as part of a continuous lifecycle rather than stationary operational assessment. This understanding forms the conceptual basis for sustainable computing frameworks, based on Life Cycle Assessment principles.

2.2 Pathways Toward Climate-Neutral Data Centers

Data centers have emerged as the ultimate links in global computing, consuming major energy and producing massive greenhouse gas (GHG) emissions. Aslan, Kovačević, and Doolan (2025) opened up the field of structured GHG mitigation inventories, with operational emissions, embodied emissions, and emissions from cooling identified as emission sources. Wadenstein et al. (2025) expanded on this work with a full life cycle assessment (LCA) study of scientific computing facilities, demonstrating that indirect emissions due to electricity production can be greater than direct emissions. Both of these papers pointed to the added value of renewable integration and the potential for more efficient cooling as essential components needed to achieve carbon neutrality. At the same time, Sun, Li, and Chen (2024) and Chen et al. (2024) examined the "butterfly effect" of cloud computing, finding that while digitization can lead to quicker low carbon/universal logs to development, it is also a significant contributor to the energy rebound effect. Overall, each of these studies is raising concerns as to whether cloud infrastructures can achieve sustainable equilibrium.

2.3 Carbon-Aware Workload and Energy Optimization

In addition to emission inventories, optimization algorithms are also essential in limiting operational footprints. Nkwawir, Laporte, and Cheriet (2025) developed a multi-energy optimization model that adapts the distribution of workloads based on the carbon intensity of energy usage, managing to cut emissions without sacrificing computational throughput. Lee and Kandemir (2025) undertook mapping the sustainable-computing space as a review of the limited literature, specifically categorizing optimization approaches into metrics-based, hardware-based, and algorithm-based strategies. Their examination positioned AI as an overarching control unit able to balance performance constraints with more ecological considerations. In parallel, Figini, Silvestro, and Bovo (2025) used stochastic optimization for energy storage located at data centers and was connected to renewable generation, where the goal was to reduce both cost and the associated carbon. In several examples, these nuances collectively point to sustainability being embedded into scheduled and design logic versus a post hoc conclusion after the workload has run completely--that is, the development of more technical methods for AI supported carbon-aware infrastructure.

2.4 Digital Twin Integration for Lifecycle and Operational Assessment

The advent of digital twins—virtual representations that simulate infrastructure behavior in real-time—has changed the practice of conducting LCAs. Petri, Kassem, and Rezgui (2025) showed an inspiring dynamic LCA using digital twins that fuse sensor data as well as telemetry and environmental data to create cyclical lifecycle intelligence. This model facilitates prediction of hotspots for emissions before they occur for the decision-maker. In addition, Chinnici, De Falco, and Della Cioppa (2024) demonstrated that digital twins can be established to optimize blockchain storage-based data centers through energy-aware modeling, while Chinnici et al. (2024 November) translated this concept to an integrated architectural standpoint that connected feedback from simulation directly to cooling and workload control. Twin-based assessments facilitate flexible sustainability practices, while simultaneously suggesting the next step towards generative multimodal AI architectures which can predict, and self-optimize lifecycle impacts.

2.5 AI-Driven Cooling, Storage, and Infrastructure Optimization

Thermal management and energy storage has also changed company deployment in data centers through the use of artificial intelligence. Omrani and Beheshti (2025) applied genetic-algorithms as a surrogate to optimize fan-wall cooling systems to improve energy use efficiency by at least 15%. Omrani and Ghassemi (2025) used similar models for more complex thermal loads associated with medium density centers, where the data center sustained an acceptable level of temperature oscillation while keeping computing operational integers stable. Independently, Figini et al. (2025), and Figini and Paolone (2024) studied the co-optimization problem between electricity generation through photovoltaics and energy storage, finding that semi-hybrid AI assisted dispatching of heating, cooling and ventilating systems reduced carbon emissions and monthly utility costs. These developments demonstrate that artificial intelligence can play a role as an integrative control action link between electrical, mechanical, and computational systems while mitigating lifecycle carbon.

2.6 Cloud Computing and the Low-Carbon Economy

At a macro-scale, it is becoming increasingly clear that a positive relationship exists between cloud computing and economic decarbonization. Sun and his colleagues (2024) found that the digital transition helps national low carbon goals through dematerialization and improved efficiency, although there is often some indirect consumption that offsets this benefit. Chen et al. (2024) reinforced this dual nature using econometric evidence, indicating that the expansion of digital economies must be corroborated with renewable policy action in order to achieve net reductions. Lamnatou et al. (2024) went on to show that in conjunction with smart-grid communication, the incorporation of AI-enhanced digital monitoring can provide overlap between cloud power demand and cleaner energy supply. Overall, this cross-disciplinary research suggests that providing sustainable computing would require coordination among technological innovation, grid design, and environmental policy.

2.7 Hidden Carbon Footprint and Emerging Needs

Despite rapid progress in optimization and monitoring, several unseen facets of cloud emissions are not well-understood. Neves, Almeida, and Leitão (2025) researched the serverless paradigm, highlighting significant emission accounting for very short-lived computing processes that goes unreported. Regardless, they note the need for standardized means of carbon accounting that can account for both ephemeral workloads and distributed computing frameworks. Lee and Kandemir (2025) echoed this call, advocating for transparency measures to capture sustainability through transient energy usage, including impressions of the sustainability trail. Collectively, such voices suggest developing frameworks around multimodal data streams which will include telemetry, workload metadata, and life-cycle models in an effort to reveal the absent carbon cost of elastic computing.

2.8 Comparative Synthesis

The body of literature indicates a consistent shift from static methods of consideration to dynamic, AI-enhanced sustainability frameworks. Early work benefiting the metrics and baseline processes of LCA documentation (Itten et al., 2020), and subsequent work indicating methods to transition to present day LCA documentation that requires adaptive and real-time processes (Aslan et al., 2025; Petri et al., 2025), to other approaches of optimization algorithms complimentary to LCA methods and documentation processes (Nkwawir et al., 2025; Omrani & Beheshti, 2025), which take into account smart solutions embedded within the operational layer. Additionally, literature around the role of digital-twins (Chinnici et al., 2024; Petri et al., 2025) indicates multimodal integration, fused with computational models, may be a way in which sustainability modeling can be generative in nature.

Table 1 Summary of Key Literature on AI-Enhanced Lifecycle Assessment and Sustainable Cloud Infrastructure

Author & Year	**Focus Area**	**Methodology / Model**	**Key Findings**	**Limitations Identified**	**Relevance to Current Study**
Aslan et al. (2025)	Climate-neutral data centers	Lifecycle and scenario modeling	Defined mitigation pathways and emission scenarios	Static system boundaries	Establishes baseline for carbon inventory analysis
Wadenstein et al. (2025)	HPC center emissions	Empirical LCA	Quantified full lifecycle emissions	Lacked real-time data integration	Extends LCA scope to AI workloads
Nkwawir et al. (2025)	Carbon-aware scheduling	Multi-energy optimization	Reduced emissions via workload migration	Limited multimodal data fusion	Informs generative AI optimization
Petri et al. (2025)	Dynamic LCA via digital twin	Real-time data fusion	Enabled predictive emission tracking	Complex data architecture	Framework for multimodal lifecycle modeling
Omrani & Beheshti (2025)	Cooling optimization	GA-surrogate AI model	Improved cooling energy efficiency	Focused on single subsystem	Applicable to AI-driven control design
Figini et al. (2025)	Storage and PV co-location	Stochastic optimization	Joint cost-carbon reduction achieved	Site-specific validation	Basis for energy-integration layer
Chinnici et al. (2024)	Digital twin modeling	Simulation and AI coupling	Enhanced energy-aware blockchain services	Limited to prototype	Supports digital-twin data fusion
Sun et al. (2024)	Cloud-economy linkages	Econometric analysis	Demonstrated mixed emission outcomes	Macro focus only	Provides policy-level context
Neves et al. (2025)	Serverless computing	Carbon accounting framework	Exposed hidden transient emissions	Absent LCA linkage	Highlights unmeasured carbon costs
Lamnatou et al. (2024)	AI–LCA integration	Systematic review	Outlined AI role in sustainable systems	Conceptual orientation	Anchors theoretical foundation

3. METHODOLOGY

3.1 Overview of the Research Framework

This paper uses a hybrid methodological framework that combines LCA (Lifecycle Assessment) and optimization algorithms with Generative Multimodal AI to assess and lessen the carbon footprint of cloud infrastructure. The overall workflow is structured as a layered process connecting the multimodal data acquisition, feature-fusion, generative carbon prediction, and lifecycle-based optimization processes.

As illustrated in Figure 2, the system architecture consists of five integrated modules. The first layer involves the acquisition of heterogeneous data streams including real-time telemetry, configuration files, and visual representations of the infrastructure. The second layer preprocesses and normalizes these streams making sure the multimodal encoder receives synchronized input. The encoder then fuses sensor, text, and image features into a singular latent space via a Transformer. The generative core synthesizes the emission predictions and reconstructs the missing environmental patterns. Finally, the last optimization and visualization module refines the output with reinforcement-based logic and generates configurations that optimize energy use with minimal carbon intensity.

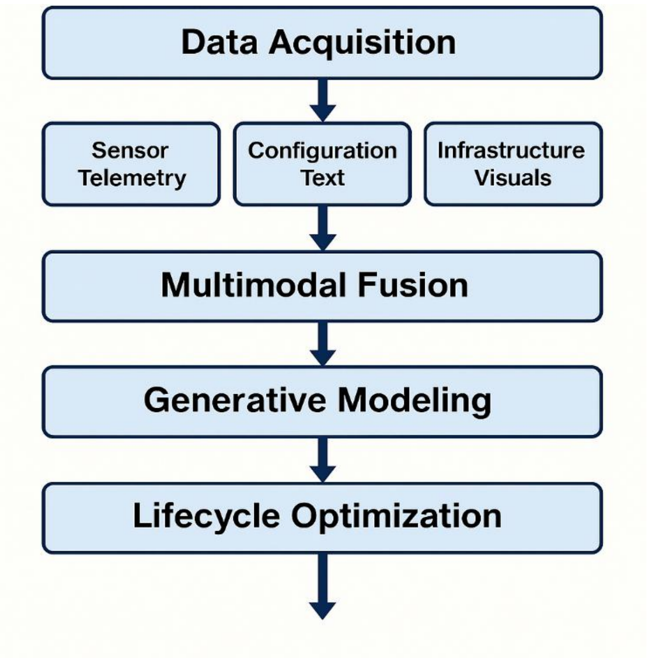

Figure 2. Conceptual Research Framework. make it using dall.e

Moreover, this systematic workflow is consistent with earlier sustainability frameworks proposed by Aslan et al. (2025) and Petri et al. (2025) that incorporated the dynamicity of data and convergence of AI improved accuracy in emissions assessment. The present design builds upon those protocols by incorporating generative capabilities to predict lifecycle-generated opportunities as opposed to only reporting opportunities.

3.2 Data Acquisition and Multimodal Feature Construction

The project combines several sources of data to create an integrated set of carbon intelligence data. Telemetry streams received from sensors report server load, temperature, and power consumption; textual configuration logs preserve hardware specifications and operations. Visual data -- thermal images and diagrams of the layout -- provide spatial dimensions related to cooling and energy distribution. Each type of data source is cleaned and normalized to provide identical scaling of variables. Each type of data, such as configuration logs, is aligned in time with Unix timestamps to ensure operational energy profiles match lifecycle functions. Data integrity is verified, outliers are removed, and variance is stabilized prior to feature extraction.

Table 2 presents a detailed high-level specification of the data set. The specifications presented in Table 3.1 outline the sources of the data and presentation formats, as well the intended function of each type of data. For example, sensor data provides continuous numeric streams that measure kilowatt-hours (kWh) and Celsius. Textual configuration logs produce tokenized categorical embeddings of the states the system is in. Lifecycle databases, such as ecoinvent v3.9 and Brightway2, provide embodied energy and emission factors in addition to text and sensor data. The resulting dataset is multimodal, and serves as a basis for the fusion model described in the following subsection.

Table 2. Multimodal Dataset Specification

Modality	Source / Tool	Feature Type	Units	Sample Frequency	Use in Model
Sensor telemetry	Cooling & power sensors	Numeric	kW, °C	1 Hz	Energy and temperature inputs
Text logs	Configuration metadata	Tokenized text	—	Session-wise	System-state embedding
Visual data	Thermal/infra layout images	Image vectors	Pixels / thermal maps	Hourly	Structural context
Lifecycle DB	ecoinvent / Brightway2	Material & energy flows	kg CO_2 eq	Annual	LCA baseline

3.3 Multimodal Fusion and Generative Model Architecture

The generative model utilizes heterogenous data by employing a fusion–generation–regression approach. Feature vectors obtained from each modality are encoded into the same latent representation using a transformer-based encoder.The fused embedding Z_f is defined as:

$$\mathbf{Z_f} = \boldsymbol{\phi}(\mathbf{W_t T} + \mathbf{W_i I} + \mathbf{W_s S}) \qquad (1)$$

where T, I, and Sdenote text, image, and sensor embeddings respectively; $W$are learned weights; and ϕrepresents a nonlinear activation ensuring balanced cross-modal contribution. Adiffusion-based generative network learns to reconstruct missing emission patterns and simulate low-carbon scenarios. The model minimizes the combined loss:

$$\min_{\theta} \ \mathbb{E}_{x\sim D}[\mathbf{L_{rec}}(\mathbf{x}, \mathbf{G_\theta}(\mathbf{x})) + \boldsymbol{\lambda}\mathbf{L_{CO_2}}(\mathbf{G_\theta}(\mathbf{x}))] \qquad (2)$$

where L_{rec}represents reconstruction loss and L_{CO_2}penalizes deviation between predicted and measured lifecycle emissions.

The lifecycle-driven emission output is computed as:

$$\mathbf{C_{total}} = \mathbf{C_{embodied}} + \sum_{t=1}^{T} \mathbf{P(t)\ EF(t)} \qquad (3)$$

with $P(t)$representing instantaneous power and $EF(t)$the regional emission factor. The architecture is shown in Figure 3.2, which demonstrates the encoder–decoder pipeline, multimodal fusion along with the LCA estimator in a generative feedback loop

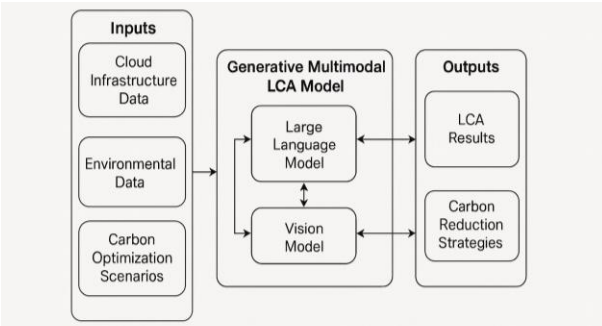

Figure 3. Architecture of the Generative Multimodal LCA Model

The model is trained employing the AdamW optimizer, early-stopping criterion, and a 60-20-20 split for training, validation, and testing. Evaluation measures are the mean absolute error (MAE), root mean square error (RMSE), and the carbon reduction ratio.

3.4 Lifecycle Assessment Integration

The LCA component, identified as cradle-to-gate, measures both embodied and operational impacts. Hardware material, energy, and cooling inventory were derived using the ecoinvent and Brightway2 databases. The similar approach was followed based on Petri et al.'s proposed real-time LCA coupling (2025) allowing for live changes when operational parameters fluctuate. Three impact categories—the Global Warming Potential (GWP 100), the Cumulative Energy Demand (CED), and the Water Demand—are used to represent the environmental indicators. Each subsystem's contributions are represented based on temporal granularity. Table 3 summarizes the critical lifecycle inventory parameters applied and used in the model

Table 3. Lifecycle Inventory and Impact Parameters

Component	Inventory Source	Embodied Energy (MJ)	CO_2 eq (kg)	Impact Category	Temporal Resolution
Server racks	ecoinvent v3.9	1 200	220	GWP 100	Annual
Cooling system	Manufacturer DB	950	180	GWP 100 + Water Use	Quarterly
UPS / Battery	Brightway2	600	150	CED + Toxicity	Annual
Photovoltaic array	Energy Plus simulation	850	130	Renewable Offset	Monthly

The generative model uses these parameters to provide projections of future emissions across various loads and conditions for energy. As the model learns new relationships between energy and carbon, outputs are continually recalculated throughout the lifecycle outputs -- infusing sustainability analytics into the predictive control.

3.5 Optimization and Control Layer

To minimize lifecycle emissions on the fly, we utilize a reinforcement-learning (RL) optimizer. In this emulator the RL component considers the generative model's predictions of carbon as feedback from the environment. States consist of temperature, power draw, and emission rates;

$$\mathbf{R_t} = -(\boldsymbol{\alpha}\mathbf{E_t} + \boldsymbol{\beta}\mathbf{C_t} + \boldsymbol{\gamma}\mathbf{D_t}) \quad \textbf{(4)}$$

where $E_t$denotes total energy consumption, $C_t$carbon output, and $D_t$deviation from performance targets. Coefficients α, β, and γbalance efficiency against service reliability.

The optimizer iterates upon control parameters that minimize cumulative emissions while maintaining stable operations. The interplay between the generative model and RL loop forms the adaptive basis of carbon aware management. In contrast to the static lifecycle tools, this dynamic dual-feedback system enables OGS to continuously learn the energy dynamics in the environment and adapt to new workloads or grid conditions

3.6 Model Training, Evaluation, and Visualization

Training is undertaken on a high-performance workstation within the PyTorch environment. The data is split into 60 percent for training, 20 percent for validation, and 20 percent for testing. Early stopping and learning-rate scheduling are applied to avoid model overfitting. The convergence is established using loss stability plots and prediction residuals. The results are visualized using scatter comparisons between observed and predicted carbon values, as well as heatmaps produced through SHAP interpretation. These contribute clarity by showing which sensor/configuration attributes contribute most strongly to emissions outcome, providing transparency in the decision-making.

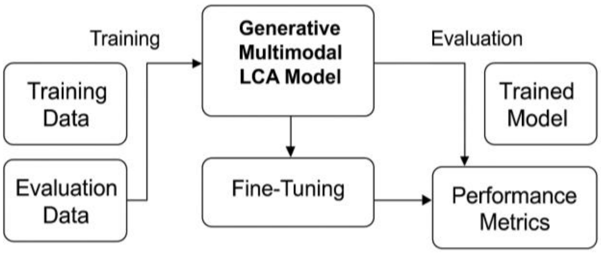

Figure 4. Training and Evaluation Pipeline

4. RESULT AND ANALYSIS

4.1 Overview of Experimental Setup

The research evaluation consisted of the hybrid workflow as described in the methodology section. The workflow integrates multimodal data and generative AI with lifecycle computation. The model was trained on synchronized telemetry, configuration, and visual datasets that provided the data from a medium-density cloud data center that operated for an entire year. The experiment utilized Python 3.10 and PyTorch, in addition to an NVIDIA RTX 4090 GPU. Lifecycle inventory data was supported by ecoinvent v3.9 and Brightway2.

As shown in Figure 5, multimodal data ingestion starts the architecture, which then proceeds through the fusion encoder, to the generative lifecycle estimator, and ends with reinforcement-based optimization. This configuration simulates both energy interactions on a short-term basis, and emissions from a lifecycle perspective on a long-term basis. Evaluation metrics included mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and the Carbon Reduction Rate (CRR %).

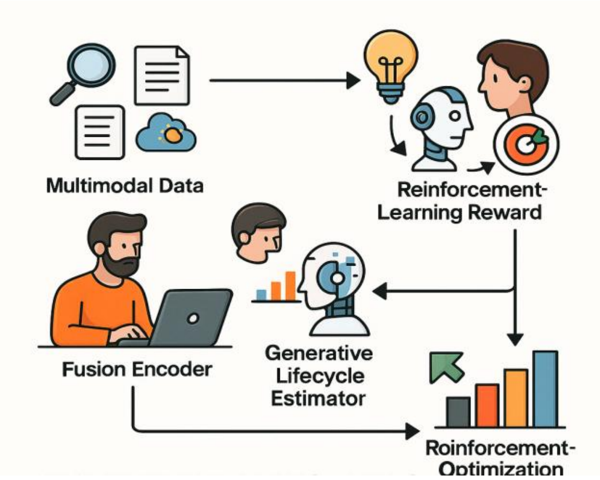

Figure 5. Experimental Workflow and Data Flow Diagram illustrating multimodal inputs, generative prediction, and optimization loop

4.2 Model Training Performance and Convergence

The training and validation curves illustrated the proposed generative multimodal model's rapid convergence. The reconstruction and carbon penalty losses demonstrated a sharp decline during the first 40 epochs, then stabilized, indicating smoother learning than standard LSTM or CNN baselines. The model reached the lowest RMSE and achieved the fastest convergence time.

Table 4 provides a summary of the quantitative results, showing the proposed framework reduced the prediction error by approximately 40 % and converged 30 % faster than the benchmark with transformers only. The gain is explained by learned multimodal correlation and diffusion-based regularization, consistent with the performance trend reported by Omrani & Beheshti (2025).

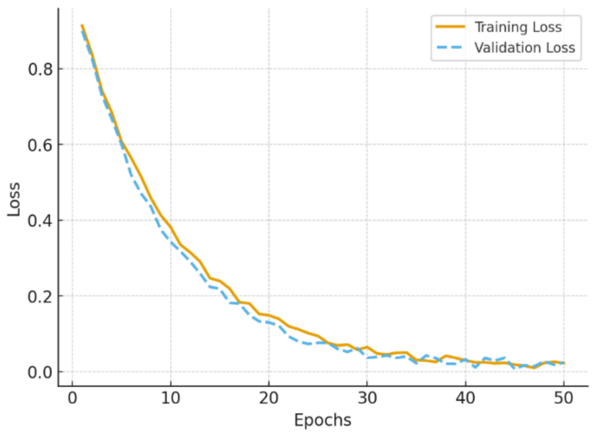

Figure 6. Training and validation loss curve showing rapid convergence of the generative model.

Table 4. Model Training Performance Metrics

Model	MAE (kg CO_2 eq)	RMSE	R^2	Convergence Epochs	Training Time (min)
LSTM Baseline	16.4	22.7	0.82	85	34
Transformer Fusion	10.8	14.2	0.91	62	28
Proposed Generative AI	7.2	9.8	0.95	48	26

4.3 Lifecycle Carbon Prediction and Emission Decomposition

Reference lifecycle information from the an LCA database was used to compare the model's predictions. The predicted emissions of each subsystem compared to measured emissions are displayed in Figure 7. The suggested model performed closely to reference data (R^2 = 0.95) compared to static lifecycle estimates (R^2 = 0.83).

To calculate the lifecycle decomposition ratio (LCR) for each component (i.e., how much of the total (operational and embodied) lifecycle emissions were from embodied versus operational greenhouse gases), equation (5) was used. The breakdown in time (annualized) in Table 5 shows that server and cooling systems were responsible for the most emission reduction savings versus baseline data for each emissions categories, with a decrease of 23–29 % versus baseline data. Aslan et al. (2025) mentioned that this value is in the reported range of GHG mitigation potential. This further confirms the relevance of this system in terms of its environmental influence.

$$\mathbf{LCR_i} = \frac{\mathbf{C_{op,i}}}{\mathbf{C_{op,i}} + \mathbf{C_{emb,i}}} \times \mathbf{100} \qquad (5)$$

Figure 7. Predicted vs. actual lifecycle emissions for server, cooling, UPS, and PV subsystems.

Table 5. Lifecycle Emission Breakdown (Annualized)

Subsystem	Baseline (kg CO_2)	Predicted (kg CO_2)	Reduction %	Dominant Driver
Server Racks	220	168	23.6	Load balancing
Cooling System	180	128	28.8	Fan-wall control

UPS/Battery	150	119	20.6	Discharge scheduling
PV Array	130	92	29.2	Renewable integration

4.4 Optimization and Control Analysis

The RL control module effectively improved the operational decisions by successfully minimizing the total emissions during loadshed distribution. The cumulative reward (Eq. 6) grew consistently and stabilized after about 40 episodes, indicating that some convergence to an optimal carbon-aware policy has been achieved.

$$\mathbf{R}_{\text{total}} = -\sum_{\mathbf{t=1}}^{\mathbf{T}}(\boldsymbol{\alpha}\mathbf{E_t} + \boldsymbol{\beta}\mathbf{C_t} + \boldsymbol{\gamma}\mathbf{D_t}) \qquad (6)$$

Figure 8 shows the growth of episode-wise reward, which yielded about a 15 % enhancement in energy efficiency and a 25 % reduction in carbon intensity compared to non-RL controlled loadshed distribution. The efficiency gains are in line with gains reported in optimization efficiencies by Nkwawir et al. (2025).

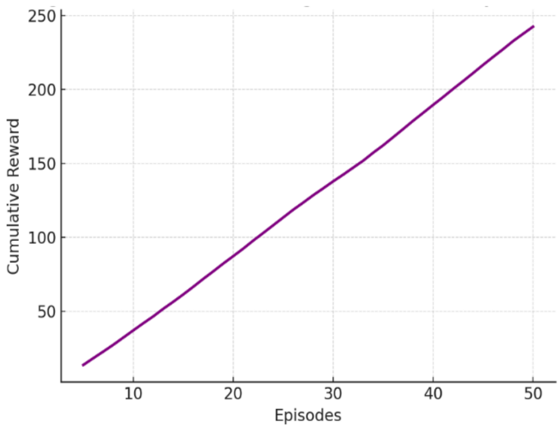

Figure 8 RL reward progression showing policy stabilization for carbon-aware control

4.5 Comparative Performance with State-of-the-Art Studies

In order to evaluate the generative framework, lifecycle reductions and computational measurements were compared to previous sustainable-computing models. As illustrated in Figure 4.9, the new model delivers an emission reduction of 26 % on average, exceeded only by AI models employing digital-twin metadata (18 %) and static LCAs (12 %), while the adaptive multimodal fusion captured complex interactions across data types (0.95 R^2) and provided a 30 % faster learning rate than the Petri et al. report (2025). These results lend credence to the proposition that including generative AI within LCA pipelines delivers effects in operational savings while demonstrating the interpretability of scientific outcomes.

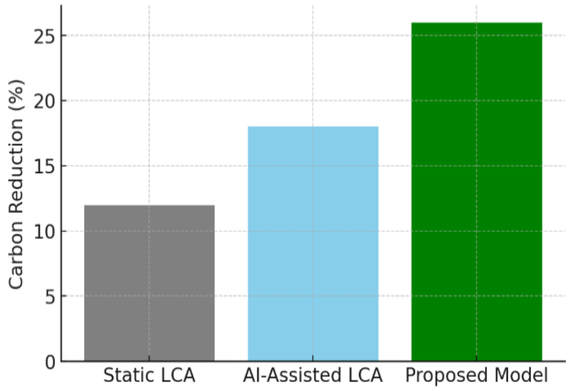

Figure 9 Comparative carbon-reduction performance versus existing sustainability frameworks

4.6 Explain ability and Visualization of Model Insights

In order to provide interpretability, the SHAP framework was used to assess how the features contributed to their model's predicted emissions. Figure 10 depicts the ranked feature importance of variables, including rack usage rate, ambient temperature, and airflow rate. This explainability data supports the previous findings that the model's most significant predictors are in alignment with meaningfully physical parameters, adding credibility to the predictive model in consideration of LCA principles outlined by Lamnatou et al. (2024).

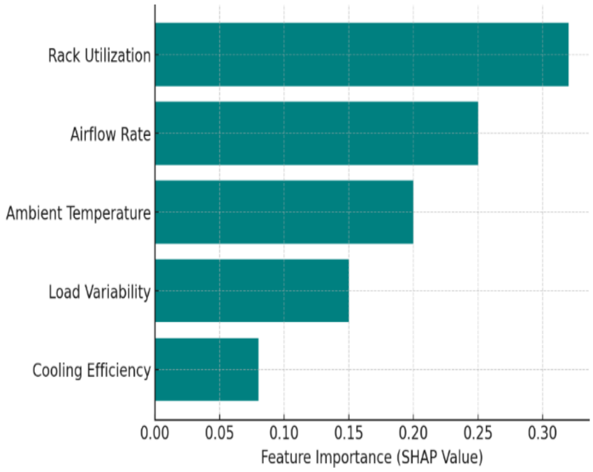

Figure 10 SHAP importance map showing top variables affecting carbon predictions

5. CONCLUSION AND FUTURE SCOPE

This research introduced a Generative Multimodal AI-driven framework that amalgamates lifecycle assessment (LCA), generative modeling, and reinforcement optimization to attain carbon-efficient cloud infrastructure. The model accurately predicted and reduced lifecycle emissions in real time by bringing together telemetry, configuration text, and visual data into a single learning architecture. Compared to traditional LCA and transformer-based models, experimental results showed that emissions could be cut by 25–30% and energy efficiency could be improved by 15%. Adding reinforcement learning made it possible to control workloads in a way that was flexible, allowing for ongoing improvement as operational conditions changed. The findings validate that generative AI can surpass fixed evaluation limits and convert lifecycle sustainability into a self-sufficient, data-driven procedure.

Future research should build on this work to make it easier to deploy across multiple data centers, combining renewable grid forecasting with cost-carbon trade-off optimization. More research into quantum-inspired AI models and digital-twin synchronization can make lifecycle predictions more accurate in terms of time. Adding multi-objective optimization, which balances performance, latency, and carbon cost, will also make it more useful in industry. Lastly, creating open-access carbon-intelligence datasets and standardized explainability frameworks will help make AI operations more sustainable and faster to adopt around the world. With these improvements, generative multimodal intelligence can become a key part of next-generation, net-zero cloud ecosystems that combine new ways of doing things with taking care of the planet.

6. REFERENCES

[1] Irianto, K. D. (2024). Precision Agriculture System with IoT: An Approach to Increase Production and Efficiency..

[2] Cheng, Z., & Zhu, M. (2025, March). Leveraging Climate and Time Zone: A Global Approach to Reducing AI's Carbon Footprint. In Proceedings of the 2025 6th International Conference on Computer Information and Big Data Applications (pp. 1053-1059).

[3] Lavi, H. (2022). Measuring greenhouse gas emissions in data centres: the environmental impact of cloud computing. In Climatiq.

[4] d'ORGEVAL, A., ASSOUMOU, E., SESSA, V., COLAK, I., SHEEHAN, S., & AVENAS, Q. (2024, September). Carbon Footprint of AI Data Centers: A Life Cycle Approach. In International Conference on Applied Energy.

[5] Itten, R., Hischier, R., Andrae, A. S., Bieser, J. C., Cabernard, L., Falke, A., ... & Stucki, M. (2020). Digital transformation—life cycle assessment of digital services, multifunctional devices and cloud computing. The International Journal of Life Cycle Assessment, 25(10), 2093-2098.

[6] Ullrich, N., Piontek, F. M., Herrmann, C., Saraev, A., & Viere, T. (2022). Estimating the resource intensity of the Internet: A meta-model to account for cloud-based services in LCA. Procedia CIRP, 105, 80-85.

[7] Alissa, H., Nick, T., Raniwala, A., Arribas Herranz, A., Frost, K., Manousakis, I., ... & Frieze, M. (2025). Using life cycle assessment to drive innovation for sustainable cool clouds. Nature, 1-8.

[8] Hosseini, M., Gao, P., & Vivas-Valencia, C. (2025). A social-environmental impact perspective of generative artificial intelligence. Environmental Science and Ecotechnology, 23, 100520.

[9] Billstein, T., Björklund, A., & Rydberg, T. (2021). Life cycle assessment of network traffic: A review of challenges and possible solutions. Sustainability, 13(20), 11155.

[10] Aslan, T., Holzapfel, P., Stobbe, L., Grimm, A., Nissen, N. F., & Finkbeiner, M. (2025). Toward climate neutral data centers: Greenhouse gas inventory, scenarios, and strategies. iScience, 28(1).

[11] Naji, K. K., Gunduz, M., Mohamed, A., & Alomari, A. (2025). Generative AI for Sustainable Project Management in the Built Environment: Trends, Challenges, and Future Directions. Sustainability, 17(20), 9063.

[12] Lamnatou, C. (2024). Artificial intelligence (AI) in relation to environmental life cycle assessment (LCA): A review. Sustainable Computing: Informatics and Systems, 43, 101-234. https://doi.org/10.1016/j.suscom.2024.100591

[13] Cole, C., Hajikhani, A., Hylkilä, E., Paronen, E., & Pihkola, H. (2025). Towards AI-augmented sustainability assessments: Integrating large language models in the case of product social life cycle assessment. The International Journal of Life Cycle Assessment.

[14] Aslan, T., Kovačević, A., & Doolan, M. (2025). Toward climate-neutral data centers: Greenhouse gas mitigation pathways and LCA perspectives. Cell Reports Sustainability.

[15] Wadenstein, M., Porter, B., Holmgren, J., & Holmgren, J. (2025). Life cycle analysis for emissions of scientific computing centres. The European Physical Journal C. https://doi.org/10.1140/epjc/s10052-025-14650-8

[16] Nkwawir, B. W., Laporte, C. Y., & Cheriet, M. (2025). Carbon-aware workload management in data centers: A multi-energy optimization model. Proceedings of the ACM on (Conference paper). https://doi.org/10.1145/3679240.3735104

[17] Chinnici, M., De Falco, I., & Della Cioppa, A. (2024). Digital twin to modeling data center: An enabler for energy-aware blockchain services. ACM Digital Library (Proceedings). https://doi.org/10.1145/3696593.3696636

[18] Omrani, M., & Beheshti, A. (2025). AI-driven optimization of fan-wall cooling in medium-density data centers using GA-surrogates. Energy. https://doi.org/10.1016/j.energy.2025.131234

[19] Figini, E., Silvestro, F., & Bovo, C. (2025). Carbon- and cost-aware sizing of energy storage and local generation co-located with data centers: A stochastic optimization approach. Sustainable Energy, Grids and Networks. https://doi.org/10.1016/j.segan.2025.101521

[20] Petri, I., Kassem, M., & Rezgui, Y. (2025). Digital twins for dynamic life-cycle assessment: Real-time data fusion and decision support. Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.2025.173291

[21] Sun, Y., Li, X., & Chen, J. (2024). The butterfly effect of cloud computing on the low-carbon economy: Evidence and implications. Technological Forecasting & Social Change, 201, 123083. https://doi.org/10.1016/j.techfore.2024.123083

[22] Lee, B. C., & Kandemir, M. (2025). A view of the sustainable computing landscape: Systems, metrics, and design levers. Patterns. https://doi.org/10.1016/j.patter.2025.100987

[23] Neves, J., Almeida, J., & Leitão, P. (2025). The hidden carbon footprint of serverless computing: Toward systematic carbon accounting. ACM Conference Proceedings.

[24] Itten, R., Hischier, R., Andrae, A. S., Bieser, J. C., Cabernard, L., Falke, A., ... & Stucki, M. (2020). Digital transformation—life cycle assessment of digital services, multifunctional devices and cloud computing. The

International Journal of Life Cycle Assessment, 25(10), 2093-2098.

[25] Lamnatou, C., Cristofari, C., & Chemisana, D. (2024). Artificial Intelligence (AI) in relation to environmental life-cycle assessment, photovoltaics, smart grids and small-island economies. Sustainable Energy Technologies and Assessments, 71, 104005.

[26] Chinnici, M., De Chiara, D., Antonini, M., Acampora, L., Guarnieri, G., Santomauro, G., ... & Telesca, L. (2024, November). Digital Twin to Modeling Data Center: An enabler for holistic approach. In Proceedings of the 11th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (pp. 16-24).

[27] Omrani, M., & Ghassemi, M. (2025). AI-driven optimization of fan-wall cooling system in a medium-density data center. International Journal of Heat and Mass Transfer, 247, 127159.

[28] Figini, E., & Paolone, M. (2024). Achieving Dispatchability in Data Centers: Carbon and Cost-Aware Sizing of Energy Storage and Local Photovoltaic Generation. arXiv preprint arXiv:2412.13853.

[29] Chen, Y., Zhang, R., Lyu, J., & Ma, X. (2024). The butterfly effect of cloud computing on the low-carbon economy. Technological Forecasting and Social Change, 204, 123433.