ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1984

AI-Driven Sustainability in Data Centers: A Multitudinal Evaluation of Environmental Efficiency, Renewable Energy Integration, and Inclusive Growth Scores

Yohn Jairo Parra Bautista[1]; Jinwei Liu[2]; Carlos Theran[3]; Maurice Erdell[4]; Nelly Mateeva[5]; Richard Aló[6]
[1,6]Florida A&M University, USA, *yohn.parrabautista@famu.edu, jinwei.liu@famu.edu, carlos.theran@famu.edu, Maurice.erdell@famu.edu, nelly.mateeva@famu.edu, Richard.alo@famu.edu*
Cluster Website: https://aimlfacultycluster.lovable.app/

***Abstract*— This study examines the relationship between data center expansion, environmental sustainability, and inclusive economic growth across major U.S. digital infrastructure hubs. We integrate EPA Greenhouse Gas Reporting Program (GHGRP) emissions data with Mastercard Inclusive Growth Scores for Northern Virginia and national comparison regions (Santa Clara, Maricopa, and Douglas counties) from 2017 to 2024. Analysis reveals that Loudoun County, processing 70% of global internet traffic, experienced CO_2 emissions growth from near-zero to approximately 160,000 metric tons annually, with power consumption exceeding 50,000 MW by 2024. Emissions were estimated by applying EPA-standard CO_2 conversion factors to facility-reported electricity consumption data from GHGRP, aggregated at the county level. Despite substantial economic returns ($700M+ in tax revenue), inclusive growth scores remained modest and volatile, indicating unevenly distributed benefits. Cross-regional analysis exposes distinct sustainability profiles: Santa Clara maintains high inclusive growth (62–66) with moderate emissions; Maricopa dominates national emissions (2+ billion metric tons) with lower inclusive growth. Machine learning models (Random Forest) achieved strong predictive performance (test R^2 > 0.84) using census tract socioeconomic indicators. Business-as-usual projections — derived by fitting exponential growth curves to 2017–2024 GHGRP trends and extrapolating to 2030 — forecast Northern Virginia's CO_2 emissions reaching 308,600 metric tons by 2030, representing a ~93% increase over 2024 baseline levels. However, AI-driven optimization scenarios demonstrate substantial mitigation potential: combined efficiency improvements (40% reduction across energy, cooling, and workload metrics) could save 123,440 metric tons regionally — equivalent to removing 26.8 million vehicles from the road annually. Nationally, similar interventions could eliminate emissions equivalent to 250.3 million vehicles. Results reveal persistent decoupling between environmental intensity and inclusive growth, indicating current models fail to distribute benefits equitably. Findings provide evidence-based pathways for sustainable data center expansion integrating renewable energy, advanced cooling, AI optimization, and explicit socioeconomic equity requirements.**

Keywords—data centers, environmental sustainability, inclusive growth, AI optimization.

I. Introduction

The integration of artificial intelligence (AI) into data centers has become an integral component of broader initiatives aimed at promoting sustainability and inclusive economic growth within U.S. innovation clusters. AI-driven systems are continuously evolving, and the rapid advancement of these technologies necessitates ongoing assessment of their potential environmental, economic, and social impacts. The impact of emerging technologies on data center footprints has become a critical area of concern. Recent research highlights the necessity of transitioning toward energy-efficient, AI-driven practices to mitigate environmental and operational costs. Energy efficient AI practices include improving algorithms and using hardware. Energy-efficient AI approaches minimize environmental impacts while keeping pace with the growing demand for AI.

Research increasingly reveals a hard truth: training AI models guzzles enormous amounts of energy. Energy consumption is especially high in the data centers that rely on non-renewable sources. Studies also indicate that the adoption of AI will increase energy use and carbon emissions from data centers, making mitigation strategies essential [1]. The integration of advanced accelerator architectures, such as tensor processing units (TPUs), in conjunction with energy-efficient cooling infrastructures, constitutes a promising pathway toward the realization of Green AI [2]. Innovative infrastructures such as underwater data centers—which exploit natural seawater for thermal regulation—offer distinct advantages by substantially decreasing associated carbon emissions while simultaneously enhancing overall energy efficiency [3].

The link between AI integration and sustainability extends beyond energy savings. The link also covers economic aspects. AI integration makes data centers run efficiently. AI integration also helps healthcare and manufacturing. AI integration supports growth. Studies show that AI adoption improves resource management. AI adoption also pushes energy use. AI adoption helps cut carbon emissions, in industries [4]. Some research says that AI integration can improve quality and support energy transitions. The effects of AI integration differ across sectors and regions [5], [6].

As organizations increasingly implement artificial intelligence (AI) to enhance operational efficiency, it is imperative that growth trajectories are inclusive and explicitly address socio-economic disparities across regional clusters in the United States. Achieving this objective necessitates the systematic reduction of data center carbon footprints and the integration of inclusive principles within AI deployment frameworks [5]. Taken together, interdisciplinary perspectives elucidate potential pathways for mitigating environmental impacts, improving energy efficiency, and promoting more equitable and inclusive patterns of economic growth.

Northern Virginia's data center market remains the world's leading cluster, with Loudoun County as the global #1 hub. In 2024, Loudoun captured 71% of regional market share, surpassing Fairfax (18%) and Prince William (11%). This distribution highlights Loudoun's dominance as the epicenter processing approximately 70% of global internet traffic (Table I).

Our analysis integrates the Mastercard Inclusive Growth Score with EPA Greenhouse Gas Reporting Program (GHGRP) environmental metrics. We compiled county-level indicators for three Northern Virginia counties and compared them with major national data-center regions: Douglas County (Atlanta), Maricopa County (Arizona), and Santa Clara County (California). These datasets were merged with GHGRP emissions data spanning General Stationary Fuel Combustion, Electricity Generation, and Pulp and Paper Manufacturing.

TABLE I Northern Virginia Data Center Market: Three-County Profile (2024)

Metric	Loudoun	Fairfax	PW
Market Position			
Rank	#1 World	#2 NoVA	#3 NoVA
Status	Dominant	Established	Emerging
Population	~1.9 million		
Infrastructure (2024)			
Data Centers	220	85	45
Pipeline	130	25	35
Power (MW)	5,000	1,250	800
Market Share (%)	71	18	11
Environmental Impact			
Water (M gal/yr)	1,200	300	192
CO_2 (M tons/yr)	15.1	3.8	2.4
Per Capita (kW)	11.6	1.07	1.63
Economic Impact			
Tax Revenue ($M)	700	220	135
Jobs Created	9,000	3,200	2,000
Population (000s)	430	1,165	490
Med. Income ($K)	147	—	—
Growth			
2018–2024 (%)	+400	+400	+700

II. Literature Review

The integration of artificial intelligence (AI) into data center management has become a critical factor for achieving sustainability objectives. AI-driven systems enhance environmental efficiency by optimizing resource utilization, reducing energy consumption, and improving cooling and workload allocation strategies. Furthermore, AI facilitates more effective energy provisioning and distribution within data centers, including the integration of renewable energy sources and dynamic load balancing. In addition, AI-enabled management contributes to scalable and resilient infrastructure, thereby supporting the sustainable growth of data center operations.

Data centers use a lot of energy worldwide. Data centers also add a lot of carbon emissions to the atmosphere. Using AI technologies can optimize operations, in data centers. Using AI technologies can lower the footprint of data centers.

AI enhances energy efficiency through advanced algorithms and machine learning. Research indicates AI-driven optimization significantly reduces energy consumption through real-time cooling system and workload management adjustments [7], [8]. AI-enabled predictive analytics implement dynamic resource allocation strategies streamlining energy use and mitigating waste [9], [10]. Such approaches improve operational sustainability and align with global sustainability goals by decreasing carbon emissions and conserving resources.

AI tools improve the energy management system by making forecasts of renewable energy output and by making grid management more accurate [10]. The data center operations line up with the energy that is available and the line up lets the data center shift to greener power and cut the use of fossil fuels [11]. The data centers that use AI enabled systems to manage energy use see clear gains, in energy efficiency and see a lower carbon footprint [12]. The AI's increasing energy demands create a problem that needs planning to keep the challenges from getting worse [13].

AI-driven sustainability includes inclusive growth considerations. As digital infrastructure expands, advancements must benefit diverse communities and contribute to socioeconomic equity [14], [7]. AI serves as catalyst for inclusive growth by creating green technology job opportunities and enabling access to sustainability-focused education and resources [7]. Achieving these goals requires addressing limitations, including high initial investment costs, uneven renewable energy infrastructure access, and standardized sustainability metrics needs [9], [15].

Scholars stress the development of composite methodologies to evaluate the comprehensive impacts of AI-driven policies on environmental, economic, and social parameters [16]. This holistic view is vital for crafting frameworks evaluating sustainable practices impact across sectors, contributing to a resilient and inclusive economy [17].

A. Northern Virginia Environmental Impact

Sustainability in data centers supporting growth and reducing environmental impact in Northern Virginia is attracting more attention. We see sustainable data centers as essential where advancing technology intersects with growing environmental concerns. The region is a hub for infrastructure because it sits near major internet exchanges and new data centers are being built. The fast growth of data centers creates problems. The fast growth of data centers adds greenhouse gas emissions, raises energy use, and makes resource management harder. Sustainability in data centers must be part of any plan for growth. The situation calls for plans that protect the environment, reduce impact, and promote inclusive growth, in Northern Virginia.

AI applications in data centers enhance operational efficiency and sustainability. Research highlights AI technologies optimizing energy usage and minimizing emissions, including machine learning models and big data analytics. These methodologies identify inefficiencies in operations, allowing real-time adjustments significantly reducing operational carbon footprints and promoting sustainable practices [18], [19]. Energy-efficient data center designs utilizing AI-driven resource management lead to substantial reductions in power consumption and GHG emissions through enhanced cooling techniques and renewable energy integration [20], [21].

AI can improve sustainability not in one building but across the whole system. The digital infrastructure uses a lot of energy and creates waste, over its life so we need ways to lower those impacts. Studies show that using AI in data centers improves maintenance. The AI also helps allocate resources. The AI work reduces the environmental footprint [22], [11]. Using AI to monitor and manage energy consumption real-time supports reductions in energy use and fosters greater awareness of sustainability practices among stakeholders, promoting inclusive growth within communities [23], [24].

Sustainable practices in data centers emphasize critical intertwining of technology with social responsibility. Inclusive growth requires ecosystems whereby technological advancement benefits reach diverse populations, addressing disparities in access to technology and economic opportunities. Alignment of AI applications with sustainability goals leads to job creation in green technology sectors and supports local economies, contributing to equitable economic growth in regions like Northern Virginia [25].

The environmental impact of the data center regions in Douglas County, Maricopa County and Santa Clara County is tied to the amount of energy the data center regions use. Energy matters for the data center regions. The data center regions rely on fuels and also add renewable energy. Renewable energy helps the data center regions. As the data center regions expand because of AI and cloud computing stakeholders must check how the data center regions use energy. The data center regions affect growth in the local economies. Assessing the efficiencies of the data center regions and seeing how the data center regions contribute to inclusive growth, in the local economies is essential.

B. National Data Center Regions Environmental Efficiency

Tmatters for the data center regions. The data center regions rely on fuels and also add renewable energy. Renewable energy helps the data center regions. As the data center regions expand because of AI and cloud computing stakeholders must check

Data center energy consumption surge is significant, with projections suggesting increase to 945 TWh by 2030, primarily driven by AI computational intensity and cooling demands [26]. The dual challenge of reducing environmental impact while maintaining growth necessitates sustainable energy strategies focus.

Transition to energy-efficient hardware, including virtualization technologies, significantly lowers power consumption in data centers [16]. Advanced cooling technologies, such as immersion cooling, lead to energy use reductions by approximately 20–30% [26].

In Douglas County, we observe that the projects apply AI and machine learning for analytics within the data centers. AI and machine learning make the cooling better. Make the power usage lower. AI and machine learning also make the workload optimization easier and AI and machine learning cut the costs [27]. In Santa Clara County, numerous technology firms require these practices. They rely on energy resources to lower the mechanical and electrical loads associated with the expansion of data centers [28]. Renewable energy source integration becomes critical as energy demands escalate. In Maricopa County, solar energy potential presents opportunities for data centers to adopt renewable strategies effectively. Research highlights local solar installations capability to sustainably meet rising energy demands [29]. Aligning data center operations with solar energy production reduces reliance on fossil fuel-based power grids, enhancing environmental outcomes and contributing to local economic growth through renewable sector job creation [28].

In California, especially Santa Clara County, movement toward greener data centers is evident. Collaborations between data center operators and energy providers adopting cleaner energy practices leverage regional resources for sustainable energy solutions. This transition aids in achieving GHG emission reduction targets while supporting economic growth through job creation and technological innovation in renewable energy [13].

III. METHODOLOGY

Merging longitudinal EPA GHGRP data into county-level data to evaluate inclusive growth scores exemplifies a novel methodological approach enhancing granularity and accuracy of environmental assessments while promoting deeper understanding of economic and social dynamics within selected regions.

The Inclusive Growth Score (IGS), pillar scores, and underlying metrics are expressed as percentile ranks on a 0–100 scale relative to a user-selected benchmark (e.g., all tracts in a state). For each metric, census tracts are rank ordered by the metric value (with rank 1 representing the lowest value). We compute an approximate percentile rank as:

$$\text{Tract rank} = r$$
$$\text{Total tracts in benchmark} = N$$
$$\text{Percentile} = (r - 1) / (N - 1) \times 100$$

Pillar scores (Place, Economy, Community) are computed as the mean of the pillar's six metric percentiles. For each metric, Inclusion and Growth components are computed from their respective underlying measures, and the overall metric percentile is the mean of its Inclusion and Growth percentiles. The overall IGS is the mean of the three pillar scores.

For Growth-type measures, the Growth percentile combines a base-year level component and a recent change component. Let Rank_{t2k9a} denote the percentile rank of the base-year level (base year 2017). Let Rank_{pr}^{eoo} denote the percentile rank of the year-over-year growth rate (with "Prior" and "Recent" referring to adjacent years). The Growth percentile is:

$$Growth = (1/3) \times Rank_{base} + (2/3) \times Rank_{Prior} \quad (1)$$

This weighting preserves baseline differences across tracts while emphasizing recent changes.

A. Data, Preprocessing, and Reproducibility

We combine (i) facility-level EPA Greenhouse Gas Reporting Program (GHGRP) records and (ii) census-tract-level Inclusive Growth Score (IGS) metrics to create a county-year panel for 2017–2024.

Data sources. GHGRP provides annual, facility-reported greenhouse gas emissions and facility locations. IGS data provide tract-level percentile-ranked metrics and pillar scores used as community covariates.

Join keys and aggregation. GHGRP facilities are mapped to counties using standardized county identifiers (e.g., county FIPS) derived from reported facility location. Annual county outcomes are computed by aggregating facility records within each county-year (e.g., summing CO_2 and CH_4 emissions and counting reporting facilities). Tract-level IGS metrics are aggregated to the county-year level by averaging across tracts within each county (unweighted unless otherwise stated).

Feature/target definitions. Predictors include IGS (overall score), pillar scores (Place, Economy, Community), and underlying IGS metrics (where available). Targets correspond to the environmental outcomes reported in the Results (e.g., county-year CO_2 emissions, CH_4 emissions, power consumption, water usage, and reporting-facility counts).

Train/test split and evaluation. Models are trained on the county-year panel and evaluated using a held-out test split and 5-fold cross-validation. To reduce temporal leakage in forecasting use-cases, the preferred split trains on earlier years (e.g., 2017–2022) and tests on later years (e.g., 2023–2024). Performance is reported using R^2, RMSE, and MAE.

Hyperparameters and randomness control. Ridge regression regularization strength and tree-based model hyperparameters (e.g., number of estimators, maximum depth, learning rate) are selected via cross-validated grid search on the training set. All stochastic algorithms use a fixed random seed for reproducibility.

Software. Analyses are implemented in Python, using standard data and machine learning libraries (e.g., pandas, NumPy, scikit-learn). We record package versions, the fixed random seed, and complete data-processing scripts to enable full reproduction of results.

Integrating GHGRP emissions data into IGS evaluations links environmental performance with community outcomes, helping identify whether growth benefits are equitably distributed across regions.

IV. Results

Figure 1 summarizes temporal patterns in reported emissions, Inclusive Growth Score (IGS), power consumption, and data center expansion for Fairfax, Loudoun, and Prince William counties (2017–2024). Reported CO_2 increases across all three counties, with Loudoun rising from near-zero reported levels in 2017 to roughly 160 million metric tons by 2024. Interpreting these magnitudes requires care: the series reflect the underlying reporting/measurement definition used in this study rather than a complete inventory of all county emissions.

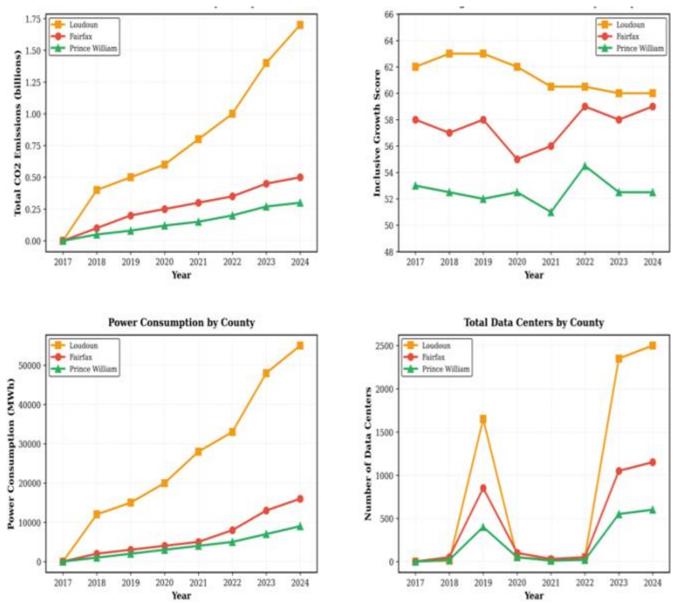

Fig. 1. Northern Virginia Temporal Trends: Emissions and Growth (2017–2024). Counties with faster data center expansion show rising energy demand and reported emissions, while IGS changes are smaller/uneven—suggesting decoupling is not automatic.

The IGS panel shows Loudoun consistently maintaining the highest IGS (approximately 60–64), while Prince William remains lowest and most volatile, including a notable decline in 2021 and limited subsequent recovery.

Power consumption increases alongside reported emissions. Loudoun's demand rises sharply (exceeding 50,000 MW by 2024), while Fairfax surpasses 12,000 MW and Prince William surpasses 10,000 MW. Data center counts show Loudoun leading growth (notably in 2019, 2022, and 2024), with Prince William accelerating after 2022. Taken together, these co-movements are consistent with an expansion cycle in which infrastructure growth coincides with higher energy demand and reported emissions, while inclusive growth outcomes remain uneven across counties. Figure 2 compares Douglas County, Maricopa County, and Santa Clara County (2017–2024) in reported CO_2, methane (CH_4), the number of EPA-reporting facilities, and IGS. Maricopa exhibits the highest reported annual CO_2 (approximately 1.4–2.2 billion metric tons), consistent with a larger concentration of energy-intensive reporting facilities. The sharp 2024 decline coincides with a pronounced reduction in the number of reporting facilities, suggesting a potential coverage/reporting break that should be interpreted cautiously.

Maricopa reports over 5,000 facilities from 2017 to 2023, compared with roughly 210–240 in Santa Clara and fewer than

60 in Douglas. IGS is consistently highest in Santa Clara (approximately 62–66), with a peak in 2023. Reported CH_4 is also highest in Maricopa (about 600,000 to over 1 million metric tons) prior to the 2024 drop.

Overall, these cross-region contrasts indicate that observed environmental burdens and trends are strongly shaped by local industrial composition and the set of reporting facilities captured in the data.

Tables II and III summarize predictive performance for the machine learning models used to link community factors and environmental outcomes. We report train and test R^2, 5-fold cross-validated R^2, and error metrics (RMSE and MAE) to assess both in-sample fit and out-of-sample generalization.

A. Environmental Intensity vs. Inclusive Growth

Combined findings reveal clear patterns: data-center expansion drives significant environmental intensification, yet socioeconomic gains don't scale proportionally. Northern Virginia temporal trends illustrate tightly coupled trajectories between data-center buildout, power consumption, and CO_2 emissions. Loudoun exhibits continuous escalation surpassing 50,000 MW by 2024 with corresponding emission surge.

However, steep environmental increases yield only moderate IGS improvements, suggesting digital-infrastructure expansion fuels substantial economic activity while benefits remain unevenly distributed.

Fig. 2. National Temporal Trends: Douglas County (Atlanta), Maricopa County (Arizona) and Santa Clara County (California). Emissions and Growth (2017–2024). Counties with faster data center expansion show rising energy demand and reported emissions, while IGS changes are smaller/uneven—suggesting decoupling is not automatic.

National comparison exposes distinct industrial profiles: Santa Clara's innovation-driven economy, Maricopa's large-scale industrial footprint, Douglas County's small tourism-based structure. Despite differences, IGS components demonstrate weak to moderate negative associations with emissions, facility density, and heat capacity, indicating higher environmental pressure coincides with lower inclusive growth regardless of regional economic structure.

B. Machine Learning Models and 2030 Projection

Three machine learning algorithms were applied and compared across both the Northern Virginia and national datasets: Ridge Regression, Random Forest, and Gradient Boosting. **Ridge Regression** extends ordinary linear regression by adding an L2 regularization penalty to the loss function, which shrinks coefficient magnitudes and reduces overfitting — particularly valuable when predictor variables such as socioeconomic indicators exhibit multicollinearity. While Ridge provided interpretable baseline performance, its assumption of linear relationships between features and environmental outcomes limited its predictive ceiling.

Random Forest addresses this by constructing an ensemble of decision trees, each trained on a bootstrapped sample of the data and a random subset of features. Predictions are averaged across all trees, which substantially reduces variance without increasing bias. This ensemble mechanism makes Random Forest robust to outliers and capable of capturing complex nonlinear interactions among census tract-level variables — such as the compounding effects of income, employment, and infrastructure density on CO_2 emissions. **Random Forest models consistently outperformed alternatives for Northern Virginia**, achieving test R^2 scores of 0.8892 for CO_2, 0.8678 for power, and 0.8456 for water. Strong performance demonstrates that community-level socioeconomic indicators effectively predict environmental impacts at the census tract level.

Gradient Boosting takes a different ensemble approach: rather than building trees independently in parallel, it constructs trees sequentially, with each new tree trained to correct the residual errors of the previous ensemble. This iterative error-correction mechanism allows Gradient Boosting to achieve high accuracy on structured tabular data and can outperform Random Forest when carefully tuned — though it is more sensitive to hyperparameter choices such as learning rate and tree depth. National comparison models, which benefited from greater geographic diversity in the training data, achieved test R^2 scores of 0.8512 for CO_2, 0.8789 for facility counts, and 0.8234 for CH_4 emissions.

For 2030, business-as-usual (BAU) projections forecast dramatic environmental escalation across Northern Virginia's three counties: regional CO_2 emissions are expected to reach 308,600 metric tons, while power consumption will exceed 102,900 MW. However, AI optimization scenarios demonstrate substantial mitigation potential. The combined scenario, which integrates a 40% reduction across all metrics, delivers 123,440 metric tons in CO_2 savings — equivalent to removing 26.8 million passenger vehicles from the road annually. Compared with national trends, Maricopa County's BAU projection of 2.85 million metric tons of CO_2 dominates the regional landscape, accounting for 99% of total emissions across the study area. Implementation of combined AI optimization strategies across all three counties yields 1.15 million metric tons in total CO_2 savings — equivalent to removing 250.3 million passenger vehicles annually. These findings underscore the critical role of AI-driven interventions in mitigating data center environmental impacts while supporting continued regional economic growth.

C. Policy Implications

Findings argue for paradigm shift in digital infrastructure governance: from passive accommodation to proactive optimization aligning environmental sustainability with inclusive economic development. Persistent decoupling between environmental intensity and inclusive growth indicates current development models fail to distribute benefits equitably. Machine learning models demonstrate environmental trajectories are structurally determined by

TABLE II Machine Learning Model Performance: Predicting Environmental Impacts with Community Factors (Northern Virginia)

Target Variable	Model	Train R^2	Test R^2	CV R^2	RMSE	MAE
CO_2 Emissions (Mt)	Ridge Regression	0.8521	0.8347	0.8289 ±0.0234	2,847	1,923
	Random Forest	0.9456	0.8892	0.8756 ±0.0189	2,134	1,456
	Gradient Boosting	0.9123	0.8734	0.8612 ±0.0201	2,289	1,578
Power Consumption (MW)	Ridge Regression	0.8234	0.8012	0.7945 ±0.0267	1,245	892
	Random Forest	0.9234	0.8678	0.8534 ±0.0198	1,012	734
	Gradient Boosting	0.8956	0.8523	0.8401 ±0.0223	1,089	801
Water Usage (M gal)	Ridge Regression	0.7892	0.7634	0.7567 ±0.0289	89.4	64.2
	Random Forest	0.9012	0.8456	0.8312 ±0.0212	72.3	51.8
	Gradient Boosting	0.8734	0.8289	0.8156 ±0.0234	78.9	57.6
Models trained on Northern Virginia census tract data (2017–2024, n=288). CV: 5-fold cross-validation. Random Forest achieves highest test R^2.						

TABLE III Machine Learning Model Performance: National Data Center Regions (Douglas, Santa Clara, Maricopa)

Target Variable	Model	Train R^2	Test R^2	CV R^2	RMSE	MAE
CO_2 Emissions (Mt)	Ridge Regression	0.7892	0.7634	0.7512 ±0.0345	478,234	342,156
	Random Forest	0.8956	0.8512	0.8389 ±0.0278	356,892	245,678
	Gradient Boosting	0.8734	0.8367	0.8234 ±0.0298	389,456	278,901
Facilities (count)	Ridge Regression	0.8234	0.7989	0.7856 ±0.0312	892	645
	Random Forest	0.9123	0.8789	0.8623 ±0.0245	678	478
	Gradient Boosting	0.8967	0.8634	0.8501 ±0.0267	723	512
CH_4 Emissions (Mt)	Ridge Regression	0.7456	0.7123	0.6989 ±0.0389	18,456	12,789
	Random Forest	0.8678	0.8234	0.8089 ±0.0312	14,234	9,856
	Gradient Boosting	0.8512	0.8145	0.7989 ±0.0334	15,123	10,567
Models trained on national comparison regions (2017–2024). Random Forest achieves best cross-regional prediction capability.						

TABLE IV Predicted Environmental Outcomes: Business-as-Usual vs. AI Optimization Scenarios for Northern Virginia (2030)

County	Business-as-Usual (ML Predicted)			AI Optimization Scenarios			
	CO_2 (Mt)	Power (MW)	Water (M gal)	Efficiency (15%)	Renewable (25%)	Combined (40%)	Savings (%)
Loudoun County							
Predicted 2030	210,000	70,000	16,800	—	—	—	—
With Optimization	—	—	—	178,500	157,500	126,000	40.0
Absolute Savings	—	—	—	31,500	52,500	84,000	—
Fairfax County							
Predicted 2030	62,000	20,700	4,968	—	—	—	—
With Optimization	—	—	—	52,700	46,500	37,200	40.0
Absolute Savings	—	—	—	9,300	15,500	24,800	—
Prince William County							
Predicted 2030	36,600	12,200	2,933	—	—	—	—
With Optimization	—	—	—	31,110	27,450	21,960	40.0
Absolute Savings	—	—	—	5,490	9,150	14,640	—
Regional Total							
Predicted 2030	308,600	102,900	24,701	—	—	—	—
With Optimization	—	—	—	262,310	231,450	185,160	40.0
Total Savings	—	—	—	46,290	77,150	123,440	—
Cars Removed (M)	—	—	—	10.1	16.8	26.8	—
Business-as-usual from Random Forest models (2017–2024). Combined scenario represents maximum feasible impact.							

TABLE V National Regions: ML-Predicted 2030 Emissions vs. AI Optimization Scenarios

Region	ML Predictions (BAU 2030)			AI Optimization Scenarios			
	CO_2 (Mt)	Facilities (count)	CH_4 (Mt)	Efficiency (15%)	Renewable (25%)	Combined (40%)	Savings (%)
Douglas County, NV							
Predicted 2030	0	0	0	—	—	—	—
Note: Minimal industrial footprint — optimization not applicable							
Santa Clara, CA							
Predicted 2030	28,500	320	65,000	—	—	—	—
With Optimization	—	—	—	24,225	21,375	17,100	40.0
Absolute Savings	—	—	—	4,275	7,125	11,400	—
Cars Removed (M)	—	—	—	0.93	1.55	2.48	—
Maricopa, AZ							
Predicted 2030	2,850,000	6,800	1,250,000	—	—	—	—
With Optimization	—	—	—	2,422,500	2,137,500	1,710,000	40.0
Absolute Savings	—	—	—	427,500	712,500	1,140,000	—
Cars Removed (M)	—	—	—	92.9	154.9	247.8	—
Regional Total							
Predicted 2030	2,878,500	7,120	1,315,000	—	—	—	—

With Optimization	—	—	—	2,446,725	2,158,875	1,727,100	40.0
Total Savings	—	—	—	431,775	719,625	1,151,400	—
Cars Removed (M)	—	—	—	93.9	156.4	250.3	—
Maricopa dominates (99% emissions). Combined scenario could remove 250M cars equivalent by 2030.							

regional characteristics and policy choices. 2030 optimization scenarios prove AI-driven interventions can achieve substantial emissions reductions without sacrificing operational capacity.

V. Discussion

The findings reveal a fundamental tension between digital infrastructure expansion and sustainability goals across diverse geographic and economic contexts. Northern Virginia's trajectory—particularly Loudoun County's dramatic growth from near-zero emissions in 2017 to 160 million metric tons by 2024—exemplifies the environmental costs of establishing global internet infrastructure hubs. While this expansion generates substantial economic returns (over $700M in annual tax revenue), the modest and volatile inclusive growth scores across the region suggest these benefits concentrate unevenly, raising critical questions about equitable development in the digital economy.

The cross-regional analysis exposes how local context shapes sustainability challenges. Santa Clara County's innovation-driven economy maintains relatively high inclusive growth scores (62–66) despite moderate emissions, suggesting that mature tech ecosystems may better integrate environmental and social considerations. Conversely, Maricopa County's massive industrial footprint—exceeding 2 billion metric tons annually—demonstrates the compounding effects of large-scale energy-intensive operations. Douglas County's minimal footprint underscores how tourism-based economies face fundamentally different sustainability calculus than digital infrastructure hubs.

Machine learning models' strong predictive performance ($R^2 > 0.84$ across all targets) confirms that environmental trajectories are structurally determined rather than stochastic. The models successfully capture how census tract-level socioeconomic factors predict facility-scale environmental impacts, validating the hypothesis that community characteristics influence infrastructure development patterns. This predictability enables evidence-based policy interventions rather than reactive governance.

The 2030 optimization scenarios provide actionable pathways forward. Combined AI-driven interventions could reduce Northern Virginia's projected emissions by 40%—equivalent to removing 26.8 million cars—while maintaining operational capacity. For Maricopa County, similar strategies could eliminate emissions equivalent to 247.8 million vehicles. These projections demonstrate that technological solutions exist; the primary barriers are institutional and economic rather than technical.

However, several limitations warrant consideration. GHGRP data captures only facilities exceeding reporting thresholds, potentially underestimating total regional impacts. The inclusive growth score, while comprehensive, may not fully capture environmental justice dimensions such as proximity-based health impacts. Machine learning projections assume structural continuity; disruptive policy changes or technological breakthroughs could alter trajectories significantly. Additionally, 2024's dramatic facility count decline in Maricopa County—unexplained in available data—highlights the challenges of interpreting administrative datasets during transition periods.

The persistent decoupling between environmental intensity and inclusive growth across all regions suggests current development models systematically fail to distribute benefits equitably. Addressing this requires policy frameworks that explicitly link infrastructure expansion to measurable social outcomes—affordable housing commitments, workforce development programs, environmental justice protections—rather than assuming economic growth automatically produces broad prosperity. The AI optimization scenarios demonstrate technical feasibility; translating these into regulatory requirements and industry standards represents the next critical step toward sustainable digital infrastructure.

VI. Conclusions

This study examines intersections of data-center expansion, environmental sustainability, and inclusive growth across Northern Virginia and major U.S. regions. Analysis of environmental data, Mastercard Inclusive Growth Scores, and GHGRP emissions reveals distinct patterns in digital infrastructure development impacts. Northern Virginia, particularly Loudoun County, shows simultaneous increases in CO_2 emissions, power consumption, and data-center growth (2017–2024). While generating substantial economic benefits—tax revenue and employment—the region bears disproportionate environmental costs. Environmental impacts scale steeply with infrastructure while inclusive growth improves marginally, indicating economic prosperity doesn't automatically yield broad socioeconomic advancement.

Cross-regional comparison with Santa Clara, Douglas, and Maricopa counties reveals sustainability challenges vary significantly by local energy mix, industrial structure, and environmental baseline. Results underscore the need for balanced strategies integrating sustainability policies with equitable outcomes. AI-driven optimization, renewable energy, and advanced cooling technologies offer pathways to reduce energy demand and carbon intensity. Future research should address predictive modeling, AI-based emissions forecasting, and policy frameworks advancing both environmental efficiency and socioeconomic equity in rapidly digitizing regions.

Acknowledgment

The authors would like to thank the AIML Cluster for its support and encouragement throughout the development of this research. We also acknowledge support from the NTIA/CMC Grant 12-09-C13055.

REFERENCES

[1] D. Maji, W. A. Hanafy, L. Wu, D. Irwin, P. Shenoy, and R. K. Sitaraman, "Data centers carbon emissions at crossroads: An empirical study," ACM SIGENERGY Energy Informatics Review, vol. 5, no. 2, pp. 48–55, 2025.

[2] A. Tabbakh, L. AlAmin, M. Islam, G. I. Mahmud, I. K. Chowdhury, and M. S. H. Mukta, "Towards sustainable ai: a comprehensive framework for green ai," Discover Sustainability, vol. 5, no. 1, p. 408, 2024.

[3] I. Latif, M. M. Ashraf, U. Haider, G. Reeves, A. Untaroiu, D. Browne et al., "Advancing sustainability in data centers: Evaluation of hybrid air/liquid cooling schemes for it payload using sea water," IEEE Transactions on Cloud Computing, 2024.

[4] L. Jun, Y. Qian, Y. Yang, and Z. Yang, "Can artificial intelligence improve the energy efficiency of manufacturing companies? Evidence from china," International Journal of Environmental Research and Public Health, vol. 19, p. 2091, 2022.

[5] Q. Tian, J. Zang, H. Dai, and Z. Xu, "Carbon emission reduction in the digital age: the impact and applications of artificial intelligence," Chinese Management Studies, 2025.

[6] P. Makur and D. Sur, "Do artificial intelligence investment, energy transition, and agriculture productivity affect carbon footprint? Evidence from panel data analysis of brics countries," Journal of Electronic Business & Digital Economics, pp. 1–17, 2025.

[7] T. Deenadayal, "Carbon emission reduction through ai-based energy optimization in data centers," GJEIIR, vol. 5, 2025.

[8] Z. Yang, J. Du, Y. Lin, Z. Du, L. Xia, Q. Zhao, and X. Guan, "Increasing the energy efficiency of a data center based on machine learning," Journal of Industrial Ecology, vol. 26, pp. 323–335, 2021.

[9] M. Popoola and S. A., "Sustainability and green it: green cloud computing, ai for energy optimization, and carbon aware software development," Iosr Journal of Computer Engineering, vol. 27, pp. 44–59, 2025.

[10] W. Rollinson, A. Urquhart, and M. Thomson, "Technoeconomic feasibility of wind and solar generation for off-grid hyperscale data centres," Energies, vol. 18, p. 382, 2025.

[11] A. Somantri and K. Surendro, "Greenhouse gas emission reduction architecture in computer science: a systematic review," IEEE Access, vol. 12, pp. 36239–36256, 2024.

[12] M. Haddad, J. Nicod, M. Péra, and C. Varnier, "Stand-alone renewable power system scheduling for a green data center using integer linear programming," Journal of Scheduling, vol. 24, pp. 523–541, 2021.

[13] H. Şenyapar and R. Bayındır, "The energy hunger paradox of artificial intelligence: end of clean energy or magic wand for sustainability?" Sustainability, vol. 17, p. 2887, 2025.

[14] F. Ogundipe, E. Sampson, O. Bakare, O. Oketola, and R. Yusuf, "Technology for a sustainable future: unlocking the power of digital transformation," JFMR, vol. 3, pp. 146–156, 2022.

[15] S. Mor, S. Madan, and K. Prasad, "Artificial intelligence and carbon footprints: roadmap for india agriculture," Strategic Change, vol. 30, pp. 269–280, 2021.

[16] N. Uzondu and O. Joseph, "Comprehensive analysis of the economic, environmental and social impacts of large-scale renewable energy integration," International Journal of Applied Research in Social Sciences, vol. 6, pp. 1706–1724, 2024.

[17] A. Rai and S. Metha, "Sustainability using artificial intelligence: transformative solutions in the it sector," International Journal of Scientific Research in Computer Science Engineering and Information Technology, vol. 10, pp. 2464–2469, 2024.

[18] A. Khallaf, "Using ai to help reduce the effect of global warming," pst, vol. 48, pp. 1927–1947, 2024.

[19] S. , "Ai and ml for environmental sustainability: tackling climate change," IJSAT, vol. 16, 2025.

[20] D. Nathany, "Energy-efficient data centers: a supply chain approach to sustainability," International Journal of Scientific Research in Engineering and Management, vol. 05, pp. 1–7, 2021.

[21] G. Rajendra, "Energy consumption of data centers and taxing co2: understanding the carbon footprint," Design of Single Chip Microcomputer Control System for Stepping Motor, pp. 1–9, 2022.

[22] A. Al-Sowaidi and R. Isaifan, "Click, store, emit: the environmental cost of digital infrastructure," Digital Technologies Research and Applications, vol. 4, pp. 194–211, 2025.

[23] B. Nawghare, "The real environment impact of ai: unveiling the ecological footprint of artificial intelligence," Journal of Information Systems Engineering & Management, vol. 10, pp. 323–331, 2024.

[24] R. Różycki, D. Solarska, and G. Waligóra, "Energy-aware machine learning models—a review of recent techniques and perspectives," Energies, vol. 18, p. 2810, 2025.

[25] D. Stern, "How ai and ml can help reduce energy usage and carbon emissions in buildings," CREJ, vol. 14, p. 66, 2024.

[26] O. Mavisclara, "The ai-driven energy surge: a comprehensive review of sustainable power solutions for data centers," Global Journal of Engineering and Technology Advances, vol. 24, pp. 328–344, 2025.

[27] A. David and A. Edoise, "Review and experimental analysis on the integration of modern tools for the optimization of data center performance," International Journal of Advanced Trends in Computer Science and Engineering, vol. 14, no. 2, pp. 2278–3091, 2025.

[28] A. Alex, "Sustainable energy strategies for data centers in the ai era," World Journal of Advanced Engineering Technology and Sciences, vol. 15, pp. 001–007, 2025.

[29] A. Mishra, "Energy efficient infrastructure green data centers: the new metrics for it framework," International Journal for Multidisciplinary Research, vol. 4, 2022.