

**CAPSTONE PROJECT**

**ON**

**­­­­­­­­**

**The Green Algorithms**

**SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE IN MASTER OF BUSINESS ADMINISTRATION (MBA)**

**IN**

Business Analytics

**UNDER THE SUPERVISION OF: SUBMITTED BY:**

Faculty Name - Dr. Anand Student Name- Prakash Kumar

Designation - Associate Professor MBA-BA-603

E-Code- E13756 UID - 23MBA20167

**SUBMITTED TO:**



**CHANDIGARH UNIVERSITY**

**March 2025**



**BONAFIDE CERTIFICATE**

This is to certify that **Prakash Kumar**, a student of Master of Business Administration- **Business Analytics** in the 4th semester at Apex Institute of Management- Chandigarh University, has completed a capstone project work on **“ The Green Algorithms ”** under the guidance of **Dr. Anand Sharma**. The work completed by the student was satisfactory.

We wish **Prakash Kumar** all the best in their future endeavors.

**SIGNATURE**

Mrs. Prerna Dhawan

**HEAD OF THE DEPARTMENT**

Apex Institute of Management

**SIGNATURE**

Dr. Anand Sharma

**SUPERVISOR**

Apex Institute of Management

Submitted for the project viva-voce examination held on \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

INTERNAL EXAMINER EXTERNAL EXAMINER

**ACKNOWLEDGEMENT**

It gives me great pleasure to express my gratitude to everyone who contributed to this effort. Without their efforts, it would not have been feasible. It would not have been possible without the kind support and help of many individuals and the organizations indulged along the way. I would like to extend my sincere thanks to all of them.

I am highly indebted to my supervisor Dr. Anand for her guidance, support, consultant supervision and for providing all the necessary information to me regarding the project.

I need to deliver my thank you additionally to Ms. Prerna Dhawan, Head of Operations, Management-Apex, Chandigarh University for her assistance and guidance. And finally, I would like to thank my family and friends for their continuous support and encouragement.

Sincerely

PRAKASH KUMAR

(UID: 23MBA20167)

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Particulars** | **Page No.** |
| 4.2.1 | Emission Efficiency by Model Type | 37 |
| 4.2.2 | Average Emissions by Cloud Provider | 38 |
| 4.2.3 | Average Emissions by Region | 39 |
| 4.2.4 | Average Emissions by Energy Source | 40 |
| 4.2.5 | Histogram: Total Emissions Distribution | 41 |
| 4.2.6 | Scatter Plot: Compute Hours vs Emissions | 42 |
| 4.2.7 | Correlation Heatmap – Operational Metrics | 43 |
| 4.2.8 | Bar Chart – Avg Power Draw by Model Type | 44 |
| 4.2.9 | Boxplot – Efficiency Score by Region | 45 |

|  |  |  |
| --- | --- | --- |
|  | **TABLE OF CONTENT** |  |

|  |  |
| --- | --- |
| Bonafide Certificate | 1 |
| Acknowledgements | 2 |
| List of figures | 3 |
| Abstract | 6 |
| **CHAPTER 1. INTRODUCTION** | 7 |
| * 1. Introduction | 7 |
| * 1. Problem Identification | 9 |
| **CHAPTER 2. LITERATURE REVIEW** | 10 |
| 2.1 Review of related literature | 10 |
| 2.2 Case Studies | 11 |
| 2.3 Bibliometrics Analysis | 18 |
| **CHAPTER 3. RESEARCH METHODLOGY** | 23 |
| 3.1 Need and Significance of the study | 23 |
| 3.2 Objectives of the study | 25 |
| 3.3 Research Design | 26 |
| 3.4 Implementation methodology | 30 |
| 3.5 Data Collection | 32 |
| 3.6 Tools and Techniques | 32 |
| **CHAPTER 4. ANALYSIS & INTERPRETATIONS** | 35 |
| 4.1 Dataset | 35 |
| 4.2 Exploratory Data Analysis | 37 |
| **CHAPTER 5. FINDINGS & SUGGESTIONS** | 46 |
| 5.1 Findings of the study | 46 |
| 5.2 Suggestions | 47 |
| **CHAPTER 6. CONCLUSION & FUTURE SCOPE** | 51 |
| 6.1 Conclusion | 51 |
| 6.2 Scope of future work | 52 |
| **REFERENCES** | 55 |

## ABSTRACT

## The exponential growth of artificial intelligence (AI) has led to substantial environmental implications, particularly in the form of increased energy consumption and carbon emissions. This study introduces *The Green Algorithm*, a holistic framework designed to assess and benchmark the environmental sustainability of AI systems. The research leverages a simulated dataset of 50 AI models deployed across various regions and data centers, capturing both operational and environmental metrics such as carbon intensity, power usage effectiveness (PUE), green energy adoption, and training efficiency. A novel composite metric—*Green Score Index*—was developed to quantify sustainability, enabling fair comparison between models irrespective of their type or location. Statistical analysis, data visualization, and predictive modeling (including support vector regression) revealed strong correlations between compute hours, energy sources, and emissions, while also highlighting the variance in sustainability even among models with similar accuracy. The findings underscore the importance of carbon-aware deployment, optimization techniques, and cloud region selection. This study not only fills a significant gap in AI sustainability assessment but also provides actionable insights for developers, organizations, and policymakers aiming to reduce the environmental footprint of AI. The Green Algorithm framework offers a practical path forward for aligning AI development with global climate goals.

# INTRODUCTION

## Introduction

In today's rapidly evolving technological landscape, artificial intelligence (AI) has emerged as a transformative force across industries, driving innovation and efficiency. However, this unprecedented growth in AI adoption has raised significant concerns about its environmental impact, particularly regarding energy consumption and carbon emissions. As organizations worldwide increasingly deploy AI systems, the collective environmental footprint of these technologies has become too substantial to ignore. This research examines the concept of "Green AI" - a paradigm focused on developing and deploying AI systems that minimize environmental harm while maintaining performance efficacy. The environmental implications of AI are multifaceted. According to recent studies, training complex AI models can generate carbon emissions equivalent to the lifetime emissions of five cars.

Data centers, which power AI operations, consume approximately 1% of global electricity and contribute to 0.3% of annual carbon emissions. As AI applications continue to proliferate, these figures are projected to increase significantly, potentially undermining global sustainability efforts. While AI presents environmental challenges, it simultaneously offers promising solutions for sustainability. As evidenced in the case of Google DeepMind, AI algorithms have successfully optimized energy usage in data centers, reducing cooling energy consumption by up to 30%. Similar applications across sectors like manufacturing, transportation, and urban planning demonstrate AI's potential to drive resource efficiency and minimize waste. This dual nature of AI—as both contributor to and potential solution for environmental challenges—underscores the complexity of the sustainability question in AI development.

The concept of Green AI encompasses various approaches to mitigate environmental impacts. These include designing energy-efficient algorithms, optimizing hardware utilization, considering the entire lifecycle of AI systems, and leveraging AI itself to solve sustainability challenges. While recent years have seen growing interest in Green AI, comprehensive frameworks for measuring and enhancing sustainability across AI systems remain limited.

This research seeks to address this gap by developing a holistic methodology for measuring sustainability in AI operations. By examining both direct impacts (such as energy consumption during training and inference) and indirect effects (such as hardware lifecycle considerations), this study aims to provide organizations with practical tools for assessing and improving the environmental performance of their AI systems.

The significance of this research extends beyond academic interest. As regulatory frameworks increasingly incorporate environmental considerations and stakeholders demand greater corporate responsibility, organizations must demonstrate sustainable practices across all operations, including AI development and deployment. By establishing clear metrics and methodologies for sustainable AI, this research contributes to both environmental stewardship and organizational resilience in an evolving business landscape.

## Identification of Problem

 **High Energy Consumption of AI Models**  
Training large-scale AI models (e.g., GPT, BERT, CV systems) requires massive computational resources, leading to significant energy usage and CO₂ emissions.

 **Environmental Impact Often Overlooked**  
Most organizations focus on improving model accuracy and speed, with minimal attention to the environmental costs associated with training and deployment.

 **Lack of a Standardized Sustainability Framework**  
There is no universally accepted metric or scoring system to evaluate and compare the environmental impact of AI models alongside their technical performance.

 **Neglect of Regional Emission Variability**  
Geographic differences in energy sources (e.g., coal vs. hydro) are not considered when selecting data center locations for AI deployment, leading to avoidable emissions.

 **Limited Transparency from Cloud Providers**  
Users often lack visibility into whether their cloud-based workloads are running on renewable or fossil fuel-based energy.

 **No Tools for Composite Sustainability Assessment**  
There are few decision-support tools that integrate **operational metrics** (like training hours, energy consumption) with **environmental indicators** (like carbon intensity, PUE, green energy usage).

 **Absence of Green AI Benchmarks**  
The AI research and development community lacks environmental benchmarking systems that reward low-emission, energy-efficient models.

 **No Unified Metric to Guide Model Selection**  
Currently, there is no composite index that enables comparison of AI models based on both performance and sustainability.

 **Low Awareness and Adoption of Green AI Practices**  
Many developers and organizations are unaware of their AI carbon footprint or how to reduce it, resulting in unsustainable AI growth.

# LITERATURE REVIEW/BACKGROUND STUDY

## Review of Related Literature

The increasing adoption of Artificial Intelligence (AI) across industries has brought not only technological transformation but also environmental challenges. As AI models grow larger and more complex, so does their energy consumption and carbon footprint. This concern has led to the emergence of the concept known as **"Green AI"**, which emphasizes the importance of developing AI systems that are not only accurate and powerful but also energy-efficient and environmentally sustainable.

The term **Green AI** was first formally introduced by Schwartz et al. (2019), who argued for a shift in AI research from performance-centric approaches to those that also prioritize computational efficiency and ecological responsibility. They highlighted how AI research has traditionally focused on accuracy and speed, with little consideration for the environmental costs of training and deploying large-scale models. Their work sparked widespread interest in understanding and reducing the environmental footprint of AI technologies.

Subsequent studies by researchers like **Strubell et al. (2019)** provided alarming data on the emissions caused by large AI models. Their analysis revealed that training a single NLP model such as BERT can emit over **300,000 kg of CO₂**, which is equivalent to the lifetime emissions of five passenger vehicles. These findings underscored the urgent need for sustainable AI development practices.

Responding to this need, **Patterson et al. (2021)** from Google introduced the concept of **carbon-aware model training**, which involves selecting geographic regions with cleaner energy sources for running AI training tasks. Their research showed that emissions could be reduced by over 70% simply by optimizing where AI workloads are run. Similarly, **Microsoft’s sustainability team (2022)** promoted deploying data centers in regions with high renewable energy availability, such as Canada and Northern Europe, to minimize environmental impact.

The role of **data centers** is crucial in this conversation. According to the International Energy Agency (IEA, 2021), data centers consume around **1% of the world’s electricity**, making their efficiency vital to AI’s sustainability. Concepts like **Power Usage Effectiveness (PUE)** have become standard measures for evaluating data center performance. Innovations like **immersion cooling** and **modular infrastructure** have been identified as effective ways to reduce power consumption and water usage.

Additionally, researchers have begun to develop **composite sustainability metrics** such as the **Green Score Index**, which incorporate various indicators like carbon intensity, green energy usage, water consumption, and energy efficiency. These metrics help assess not just the computational performance of AI models, but their overall environmental sustainability as well.

Despite growing research interest, several gaps remain. There is still no universally accepted framework for evaluating AI sustainability, and few tools exist that allow businesses to make real-time, data-driven decisions about the environmental impact of their AI systems. Most existing studies focus either on operational AI performance or environmental factors, but not both. This lack of integration limits the ability to make holistic sustainability assessments.

In conclusion, the existing literature strongly supports the case for Green AI and highlights the need for a unified approach that brings together operational performance and environmental data. This project addresses that gap by proposing **The Green Algorithm**—a decision-support model that evaluates the sustainability of AI models using both operational efficiency and environmental responsibility metrics.

## Case Studies

# Carbon Footprint Management in Global Supply Chains: A Data-Driven Approach Utilizing Artificial Intelligence Algorithms by [Rong Huang](https://ieeexplore.ieee.org/author/918192640561094)

This study presents a data-driven framework that uses artificial intelligence (AI) to manage and reduce carbon footprints in global supply chains. With growing emphasis on sustainability, the proposed approach leverages AI algorithms to collect, analyze, and monitor emissions data from various supply chain stages such as transportation, manufacturing, and sourcing. By applying machine learning and optimization techniques, the framework identifies key areas for emission reduction while maintaining efficiency. Real-time monitoring and predictive analytics support proactive decision-making, helping businesses adapt to environmental regulations and market changes. Overall, this AI-powered model enhances sustainability efforts and promotes responsible supply chain management.

1. **Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools by** [**Nesrine Bannour**](https://aclanthology.org/people/n/nesrine-bannour/)

This paper explores the environmental impact of modern Natural Language Processing (NLP), which heavily relies on deep learning techniques. Given the high energy consumption of these methods, the authors advocate for a cost-benefit analysis that includes both accuracy and carbon footprint. The paper reviews six tools—Carbon Tracker, Experiment Impact Tracker, Green Algorithms, ML CO2 Impact, Energy Usage, and Cumulator—used to measure energy use and CO2 emissions in NLP experiments. These tools are tested on named entity recognition tasks across different computational setups (local servers and computing facilities). The study concludes with practical recommendations for accurately assessing the environmental impact of NLP research.

1. **A systematic review of Green AI by** [Roberto Verdecchia](https://wires.onlinelibrary.wiley.com/authored-by/Verdecchia/Roberto)

As AI adoption grows rapidly, its environmental impact—particularly carbon emissions—has become a significant concern. This has led to the emergence of *Green AI*, a research field focused on making AI development more environmentally sustainable. This article provides a systematic review of 98 Green AI studies, highlighting trends and findings. Since 2020, interest in Green AI has surged, with most research focusing on monitoring energy usage, optimizing hyperparameters, and benchmarking models for sustainability. Studies are predominantly based on training neural networks with image data, using lab experiments, and often yield energy savings exceeding 50%. While industry involvement exists, most studies target academic audiences, and practical Green AI tools remain limited. The review concludes that the field has matured and is now ready to shift toward industrial applications and broader research strategies.

1. **A review of green artificial intelligence: Towards a more sustainable future by Verónica Bolón-Canedo**

This paper explores *Green AI* as a key strategy for improving the environmental sustainability and inclusivity of artificial intelligence. Unlike traditional AI, Green AI focuses on achieving high accuracy without increasing computational costs, enabling researchers with limited resources—such as just a laptop—to conduct quality research. The paper highlights three main aspects: (1) *Green-by AI*, which uses AI to support eco-friendly practices in other domains; (2) *Green-in AI*, which involves designing energy-efficient machine learning algorithms; and (3) tools for measuring and optimizing energy usage. It also discusses the role of regulations in encouraging sustainable AI development and outlines future directions for creating more eco-conscious AI systems.

# Green AI for IIoT: Energy Efficient Intelligent Edge Computing for Industrial Internet of Things by [Sha Zhu](https://ieeexplore.ieee.org/author/37089280589)

This article explores how *intelligent edge computing* can enable energy-efficient, or *Green AI*, solutions for the Industrial Internet of Things (IIoT) in the context of Industry 4.0. Traditional AI applications in IIoT often rely on high-end servers, leading to significant energy consumption. To address this, the authors propose an intelligent edge computing framework with a heterogeneous architecture that offloads AI tasks from centralized servers to edge devices. A novel scheduling algorithm is introduced to optimize energy use across various computing resources. Performance tests using a physical testbed and extensive simulations demonstrate that the proposed strategy significantly reduces energy consumption—using less than 80% of the energy required by static scheduling and 70% compared to FIFO scheduling. This approach highlights the potential of intelligent edge computing to support sustainable and efficient AI-driven IIoT systems.

# Green Artificial Intelligence: Towards an Efficient, Sustainable and Equitable Technology for Smart Cities and Futures by Tan Yigitcanlar

This perspective paper discusses the limitations of current AI applications in smart cities and advocates for the adoption of *Green AI* as a transformative approach. Many existing AI-driven urban initiatives have failed due to narrow, technocentric, and reductionist strategies that overlook the complexity of urban systems. The authors argue that successful smart city development requires AI solutions that prioritize not just efficiency, but also sustainability and equity. Green AI is presented as a way to shift from purely technical fixes to holistic approaches that better align with the goals of smart, inclusive urban futures. The paper evaluates current trends and practices in AI and smart city planning, urging policymakers and urban planners to embrace Green AI for more balanced and environmentally responsible city development.

# Improving Automated Machine-Learning Systems through Green AI by Dagoberto Castellanos-Nieves

This paper examines the environmental impact of *Automated Machine Learning (AutoML)* and explores ways to enhance its sustainability. While AutoML simplifies the design and optimization of machine learning models, it often involves high computational and energy costs, which have been largely overlooked. Focusing on sustainability, the study conducts a proof-of-concept using the Scikit-learn library to evaluate energy efficiency during hyperparameter tuning with Bayesian and random search methods. The results show that integrating energy efficiency metrics into AutoML processes can significantly reduce environmental impact. These findings support the principles of *Green AI*, promoting more ecologically responsible AI development. The proposed approach not only improves the sustainability of AutoML but also shows potential for broader application in similar computational tasks.

1. **UTILIZING ARTIFICIAL INTELLIGENCE IN ENERGY MANAGEMENT SYSTEMS TO IMPROVE CARBON EMISSION REDUCTION AND SUSTAINABILITY by Eda Tabaku**

# This article explores the transformative role of *artificial intelligence (AI)* in enhancing *Energy Management Systems (EMS)* to reduce carbon emissions and combat climate change. Through a comprehensive literature review, the study evaluates how AI-driven EMS outperforms traditional static and reactive methods by offering real-time optimization, predictive maintenance, and advanced data analytics. Case studies from industrial and public sectors demonstrate tangible benefits such as reduced operational costs, better integration of renewable energy, and improved sustainability practices. While challenges like high implementation costs, data privacy, and regulatory hurdles exist, the paper highlights the need for supportive policies and strategic investments. Ultimately, the study concludes that AI-powered EMS is crucial for achieving significant emissions reductions and building resilient, eco-efficient energy infrastructures aligned with global climate goals.

# The Optimization of Carbon Emission Prediction in Low Carbon Energy Economy Under Big Data by [Ji Luo](https://ieeexplore.ieee.org/author/37089870800)

This study addresses the urgent need for low-carbon energy solutions amid escalating climate change by introducing an advanced AI-based model for carbon emission prediction and low-carbon economic evaluation. The proposed model, called the *Multi-universe Quantum Harmony Search-Algorithm Dynamic Fuzzy System Ensemble (MUQHS-DMFSE)*, integrates the MUQHS optimization algorithm with a dynamic fuzzy prediction system. This composite model employs a sliding factor matrix to enhance prediction accuracy, achieving a Mean Absolute Percentage Error (MAPE) under 3.5% and minimal MAE and RMSE values, confirming its high accuracy.

For assessing low-carbon economic performance, the study applies *Data Envelopment Analysis (DEA)* using CCR and BCC models to evaluate technical efficiency, input redundancy, and output gaps across decision-making units. Using data from S Province, the analysis highlights the need for structural changes such as promoting clean energy, controlling emissions, and developing the tertiary sector.

The research underscores the practical and scientific value of combining *artificial intelligence* and *big data* for precise carbon emission forecasting and informed energy policy planning, contributing significantly to sustainable economic development strategies.

# How does artificial intelligence affect pollutant emissions by improving energy efficiency and developing green technology by Wei Zhou a b

# This study investigates the environmental impact of industrial robot adoption in China, focusing on how these AI-powered technologies influence pollution emissions across different regions. Using provincial panel data from 2010 to 2019, the research finds that industrial robots significantly reduce pollution emissions intensity. This conclusion remains robust across various tests. The mechanism analysis reveals that the reduction is achieved through improved energy efficiency and the advancement of pollution reduction technologies. Regional differences are also noted: in the eastern region, pollution reduction is mainly driven by technology upgrades, while in the western regions and areas outside the Yangtze River Economic Belt, both improved energy efficiency and technological enhancements contribute. These findings highlight the role of AI and industrial automation in promoting green production practices and offer valuable insights for regional pollution control strategies in China.

# AI-Driven Green Optimization in Well Construction: Carbon Emission Management Through Technical Limit Performance Benchmarking by [K. W. Amadi](javascript:;)

This paper explores the use of *machine learning (ML)* to reduce carbon emissions in drilling operations, which account for 10–15% of global energy-related emissions—approximately 5 billion tonnes of greenhouse gases. The study focuses on optimizing drilling performance to lower CO₂ emissions by applying ML models to estimate *Unconfined Compressive Strength (UCS)* using basic, low-cost drilling parameters such as weight on bit (WOB), rotary speed (RPM), penetration rate (ROP), and torque (TOR). Four supervised learning algorithms—Random Forest, Support Vector Regression, Artificial Neural Networks (ANN), and CatBoost—were tested, with ANN and CatBoost yielding the highest prediction accuracy (R² = 0.77 and 0.75, respectively).

The ML-driven UCS predictions allowed for benchmarking drilling performance within technical limits, enabling comparison between actual and optimal drill rates across various lithologies. A case study demonstrated that this benchmarking approach improved drilling efficiency by 30–60% and reduced CO₂ emissions by 50%, equivalent to 20 tonnes of CO₂ emissions saved. The research highlights the potential of ML in enhancing operational efficiency, reducing environmental impact, and supporting sustainable energy extraction practices.

1. **Leveraging Reinforcement Learning and Genetic Algorithms for Enhanced Optimization of Sustainability Practices in AI Systems by Aravind Kumar Kalusivalingam**

This research presents a novel hybrid approach combining Reinforcement Learning (RL) and Genetic Algorithms (GA) to enhance the sustainability of artificial intelligence (AI) systems. As AI becomes more widespread and complex, its environmental impact—particularly in terms of energy consumption and carbon emissions—has grown significantly. The proposed framework leverages RL to adapt AI system parameters in real-time based on environmental performance metrics, while GA facilitates the evolution of these parameters to discover the most energy-efficient configurations. Simulation results show that this combined approach significantly reduces energy use and emissions compared to conventional optimization methods. The framework also demonstrates strong potential for generalization across various AI applications, offering a scalable solution for sustainable AI development. This research marks an important step toward integrating environmental responsibility into AI design and sets a foundation for future efforts in building eco-friendly AI systems.

# Artificial Intelligence and Carbon Emissions in Manufacturing Firms: The Moderating Role of Green Innovation by Yixuan Chen

# This study investigates the role of artificial intelligence (AI) in reducing carbon emissions within China's manufacturing sector, a major contributor to global emissions. Analyzing data from Chinese A-share listed manufacturing companies from 2012 to 2021 using a fixed-effects regression model, the research confirms that the adoption of AI technologies significantly contributes to carbon emission reductions. Furthermore, the study explores the moderating role of green innovation—including green technological innovation, green management innovation, and green product innovation. These innovations amplify the effectiveness of AI in lowering emissions. The findings highlight the strategic importance of intelligent manufacturing and digital transformation in promoting sustainable development. Overall, the study underscores AI’s value as a technological enabler for low-carbon growth and provides theoretical insights into how digital innovation supports environmental sustainability in industrial sectors.

# Predicting CO2 Emission Footprint Using AI through Machine Learning by Yang Meng

This study focuses on forecasting global CO₂ emissions by incorporating the unique impact of the COVID-19 pandemic using advanced time series models. Recognizing that excessive CO₂ emissions from industrial and vehicular sources are major contributors to climate change, the research emphasizes the importance of accurate long-term predictions for shaping emission-reduction policies. Using four *SARIMAX* models (a variant of ARIMA), the study analyzes CO₂ emissions across different phases of the pandemic: pre-COVID, start-COVID, transition-COVID, and post-COVID.

Among the models, the *post-COVID* SARIMAX model demonstrated the highest prediction accuracy, with the lowest Mean Absolute Percentage Error (MAPE) of 0.09. This model was used to forecast total global CO₂ emissions from 2022 to 2072, with short-term predictions (2022–2027) ranging from 36,218.59 to 37,921.47 million tons (MT). The results align well with IPCC projections and suggest that AI-based approaches can effectively support climate policy decisions.

The study demonstrates the utility of Machine Learning (ML) for modeling pandemic-affected emission trends and provides a valuable decision-support tool for future global emission control strategies. It also recommends including more external variables in future research to enhance prediction accuracy further.

## Bibliometric analysis

## Bibliometric analysis involves the systematic evaluation of research publications to understand trends, methodologies, and gaps in a specific domain. In the context of sustainability in analytics, particularly carbon footprint measurement and optimization, bibliometric analysis highlights key features, effectiveness, and drawbacks of existing studies.

## Key Features

## Focus on Carbon Footprint Measurement:

## Studies emphasize calculating carbon emissions from various sources such as electricity usage, transportation, paper consumption, and classroom activities1.

## Specific formulas are developed for emissions calculations tailored to different contexts (e.g., commuting emissions: Ecommuting=Ef(v)⋅S*E*commuting=*Ef*(*v*)⋅*S*, where Ef(v)*Ef*(*v*) is the emission factor per kilometer for a transportation mode)1.

## Behavioral Factors:

## Research highlights the significant role of human behavior in influencing energy consumption and carbon emissions1.

## Behavioral adjustments are shown to reduce energy consumption by 10-20% without additional equipment costs1.

## Predictive Analytics Integration:

## Machine learning models (e.g., Support Vector Regression) are used to predict future carbon emissions based on historical data and external factors such as weather and event-related data1.

## Predictive analytics enables identification of actionable steps to optimize energy consumption linked to specific activities1.

## University-Specific Case Studies:

## Studies focus on universities as significant contributors to carbon footprints, with detailed analyses of emissions from student activities at institutions like Tongji University (China), Erasmus University Rotterdam (Netherlands), and Universiti Teknologi Malaysia1.

## Carbon footprint per student varies across institutions due to differences in electricity usage, commuting habits, and academic activities.

## Effectiveness

## Comprehensive Emissions Calculation:

## The use of sector-specific formulas ensures accurate measurement of emissions from diverse sources like transportation and electricity consumption1.

## Tools like the Campus Carbon Calculator provide structured methodologies for institutional assessments1.

## Predictive Modeling:

## Predictive analytics offers valuable insights into future trends in carbon emissions, enabling proactive measures for optimization1.

## Models like SVR demonstrate high accuracy in forecasting emissions using metrics such as MAE (Mean Absolute Error) and RMSE (Root Mean Square Error)1.

## Behavioral Insights:

## Studies identify behavioral patterns contributing significantly to emissions, providing opportunities for targeted interventions without additional costs1.

## Awareness campaigns based on behavioral data can drive substantial reductions in energy consumption1.

## Cross-Institutional Comparisons:

## Comparative analyses across universities reveal best practices for reducing emissions, such as energy-efficient infrastructure or sustainable commuting policies1.

## Drawbacks

## Limited Scope:

## Many studies focus on specific emission sources or activities (e.g., commuting or electricity usage) without considering the complete lifecycle of carbon footprints1.

## Few studies integrate embodied carbon from infrastructure manufacturing or end-of-life disposal into their calculations.

## Regional Dependency:

## Emission factors used in calculations are often region-specific (e.g., Indonesia's emission factors for electricity), limiting the generalizability of findings across different geographic contexts1.

## Behavioral Complexity:

## While behavioral factors are identified as significant contributors, quantifying their impact accurately remains challenging due to variability in individual actions and preferences1.

## Predictive Model Limitations:

## Predictive analytics models rely heavily on the quality and granularity of historical data; inaccuracies or gaps in data can compromise model reliability1.

## External factors influencing emissions (e.g., policy changes or technological advancements) are difficult to incorporate comprehensively into predictive models.

## Institutional Focus:

## The focus on universities limits applicability to other sectors with different operational dynamics and emission sources (e.g., manufacturing or healthcare)

# Research Methodology

## Need and Significance of the study (GAP)

Artificial Intelligence (AI) is transforming every major sector—healthcare, education, manufacturing, finance, marketing, and more. With this rapid proliferation comes a silent cost: **the growing environmental footprint of AI models**. Training just one large-scale model (like GPT-3 or BERT) can emit **hundreds of kilograms of CO₂**, consume **thousands of kilowatt-hours (kWh)** of electricity, and require vast **cooling resources**.

As businesses and governments pledge toward **net-zero emissions**, there is a **crucial need to address the environmental implications of emerging technologies like AI**. This study arises from the understanding that **AI’s growth must be aligned with global sustainability goals**, and it is necessary to establish frameworks to **measure, monitor, and manage its impact**.

**Identified Research Gaps**

Despite growing awareness of climate change and green technologies, current research around **AI sustainability is limited, fragmented, or lacks actionable insights**. This study addresses the following gaps:

**1. Lack of Holistic Sustainability Models**

* Most existing studies assess AI models only based on **technical or business performance metrics** like accuracy, latency, or ROI.
* There's an absence of **multi-dimensional models** that evaluate **energy usage, carbon emissions, water use, and data center efficiency** in a unified framework.

**2. Limited Application of Composite Sustainability Indicators**

* No widely accepted standard exists for scoring or benchmarking AI models on sustainability.
* There is a lack of practical metrics like a **Green Score Index** or **Carbon Efficiency Ratio** that combine operational and environmental factors.

**3. Geographic and Infrastructure Context is Ignored**

* Environmental impacts of AI **vary drastically based on region**—due to differences in grid energy sources (coal, hydro, solar), cooling systems, and infrastructure.
* However, most research treats all AI training as if it occurs in a homogeneous setup, ignoring **contextual variability**.

**4. Sustainability is Treated as a Technical Problem**

* Many existing papers are **overly technical or academic**, making it hard for **business managers, policymakers, or cloud providers** to implement sustainability insights into their operations.

**5. AI for Sustainability vs. Sustainability in AI**

* While there's growing research on using AI to **solve sustainability challenges** (e.g., climate prediction, energy forecasting), there is **little focus on making AI itself sustainable**—which this study directly addresses.

**6. Lack of Simulation-Based or Scalable Models**

* Few studies use simulation-based data modeling to **test multiple scenarios**, AI model types, or data center configurations to derive generalizable insights.

**Significance of the Study**

This study is both **timely and relevant** in the context of global digital expansion and climate responsibility. Its significance lies in several key areas:

**1. Promotes Responsible Technology Use**

* The study brings **environmental consciousness** into the conversation around AI adoption, nudging developers, businesses, and policymakers toward **sustainable digital practices**.

**2. Supports ESG and CSR Reporting**

* With growing emphasis on **ESG (Environmental, Social, Governance)** and **Corporate Social Responsibility**, businesses need models like the **Green Algorithm framework** to quantify and report their AI-related emissions.

**3. Aligns with Global Sustainability Goals**

* The research supports international targets such as the **Paris Agreement**, **Net-Zero Emission Goals**, and the **United Nations Sustainable Development Goals (SDG 12, 13, 9)**.

**4. Introduces a Decision-Making Tool**

* The **Green Score Index** developed in the study offers a practical benchmarking tool to compare AI models not just by performance, but by **sustainability**, guiding:
  + Cloud deployment strategies
  + Data center selection
  + AI model architecture choices

**5. Bridges Academia and Industry**

* This research provides **real-world applicable insights** using simulation-based data, making it valuable for both academic literature and **industry implementation**.

**6. Fosters Innovation in Sustainable AI**

* It opens doors to **further research** in green AI hardware, efficient cooling, eco-friendly algorithms, and optimization of energy-intensive models.

## Objective of the study

* To **analyze and measure** the environmental impact of AI models in terms of energy consumption, carbon emissions, and water usage.
* To **compare different AI model types** (e.g., NLP, CV, Recommender, Transformers) based on their sustainability metrics.
* To **develop a composite “Green Score Index”** that ranks AI models on environmental sustainability performance.
* To **correlate model performance** (e.g., accuracy, training time) with sustainability indicators and identify energy-efficient alternatives.
* To promote awareness and responsibility among developers, organizations, and policymakers regarding AI’s environmental impact.

## Research Design

The research design is the **blueprint of the entire study**, outlining how data was collected, structured, analyzed, and interpreted to evaluate the sustainability of AI systems. This study focuses on integrating operational and environmental dimensions to assess the ecological footprint of artificial intelligence.

**1. Nature of the Study**

The research is primarily:

* **Descriptive** – It provides a systematic description of the environmental performance of AI models across various categories.
* **Analytical** – It draws relationships between operational performance and sustainability outcomes, using statistical tools and models.
* **Exploratory** – Since sustainability in AI is still an emerging field, the study explores new metrics like *Green Score Index* to bridge the gap between technology and environmental responsibility.

**2. Research Approach**

A **quantitative research approach** is adopted to analyze structured numerical data and extract patterns and insights. The study focuses on measurable variables such as:

* Energy consumption (kWh)
* CO₂ emissions (kg)
* Renewable energy usage (%)
* Model accuracy (%)
* Power Usage Effectiveness (PUE)
* Water usage (liters)

These variables help quantify and compare the environmental impact of different AI model types and deployment locations.

**3. Data Source and Collection**

The study uses **secondary data**, generated synthetically to simulate real-world operational and environmental conditions in AI deployments.

* **Operational Dataset** includes:
  + Model ID, type, training hours, energy use, carbon emissions, green energy usage, and sustainability score.
* **Environmental Dataset** includes:
  + Region, energy source, carbon intensity, PUE, cooling method, water usage, and annual emissions.

Data was created using Python programming with controlled randomization to ensure realistic variability while maintaining internal consistency.

**4. Population and Sampling**

As the study is simulation-based, a **sample of 100 AI models** and **100 environment profiles** was selected to represent a broad spectrum of model types, energy profiles, and deployment conditions. The models vary in size, complexity, and training behavior, making the sample diverse and suitable for sustainability assessment.

**5. Data Integration Technique**

To simulate real-world analytics, operational and environmental data were **merged** using a common key: Data Center Location.

This allowed each AI model to inherit regional and infrastructural characteristics like:

* Carbon intensity of energy sources
* Cooling methods used
* Renewable energy share

This technique provided a multi-dimensional view of how both internal (model) and external (infrastructure) factors influence sustainability.

**6. Data Analysis Tools & Techniques**

* **Tools Used**:
  + Microsoft Excel (for tabulation and charts)
  + Python (Pandas, Matplotlib, Seaborn for data processing and visualization)
* **Techniques Applied**:
  + Descriptive statistics (mean, variance, distribution)
  + Correlation analysis (e.g., between green energy % and sustainability score)
  + Composite scoring (creation of Green Score Index)
  + Visual analytics (bar graphs, scatter plots, regional summaries)

**7. Key Variables**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Type** | |  | | --- | |  |  |  | | --- | | **Examples** | |
| Operational Metrics | Training Hours, Energy Consumption, Model Accuracy |
| Environmental Data | CO₂ Emissions, Carbon Intensity, PUE, Water Usage, Renewable % |
| Composite Metrics | Carbon Efficiency, Sustainability Score, Green Score Index |

**8. Research Model**

The conceptual model assumes that **AI sustainability is a function of both operational efficiency and environmental context**. The following relationships were studied:

* Model Type → Training Hours → Energy → CO₂ Emissions
* Data Center Region + Energy Source → Emissions & Water Usage
* Combined Metrics → Green Score Index → Comparative Sustainability.

**9. Limitations of the Research Design**

* Data used is **synthetic**, not collected from live commercial environments.
* The model does not account for **hardware-specific factors** (like GPU type or inference load).
* **Real-time inference emissions** are not included, focusing only on training emissions.
* Externalities like **regulatory policy** or **organizational energy strategies** were excluded.

**10. Outcome of the Research Design**

The applied design enabled a **holistic, quantifiable, and actionable assessment** of AI sustainability. It helped:

* Identify best-performing models and regions
* Propose optimization strategies
* Create a scoring framework that could be adopted in real-world AI practices

This research design serves as a scalable and replicable model for organizations and researchers aiming to reduce the environmental impact of AI systems.

## Implementation plan/methodology

**Step 1: Dataset Development**

* Two datasets were synthetically generated using Python:
  + **Operational Dataset**: Includes model type, training time, energy usage, carbon emissions, green energy usage, and model performance.
  + **Environmental Dataset**: Includes data center region, primary energy source, PUE (Power Usage Effectiveness), water usage, cooling method, and renewable energy percentage.

**Step 2: Data Integration**

* The two datasets were **merged using the Data Center Location** as a common attribute.
* This created a unified view linking each AI model to its infrastructure’s environmental characteristics.

**Step 3: Metric Formulation**

* New sustainability metrics were derived:
  + **Carbon Efficiency** = CO₂ Emissions / Energy Consumption
  + **Green Score Index** = A composite score based on:
    - Green energy usage
    - Renewable energy %
    - Carbon intensity of the region
    - Sustainability Score (performance + environmental factors)

**Step 4: Data Analysis**

* Performed using Python (Pandas, Matplotlib) and Excel:
  + Descriptive statistics (mean, range, standard deviation)
  + Correlation analysis (e.g., between green energy usage and sustainability)
  + Comparative analysis across model types and regions

**Step 5: Visualization**

* Data was visualized using:
  + **Bar charts**: Model Type vs. CO₂ Emissions
  + **Scatter plots**: Green Energy Usage vs. Sustainability Score
  + **Regional Summary Tables**: PUE, Annual Emissions, and Water Usage
* These visual aids helped identify patterns and best-performing models/regions.

**Step 6: Findings & Interpretation**

* Analytical results were compiled into findings that support:
  + Region-specific recommendations
  + Infrastructure optimization
  + AI model sustainability benchmarking

**Step 7: Recommendations and Framework Design**

* Proposed an implementation framework for:
  + Sustainable AI model deployment
  + Decision-making based on Green Score Index
  + Environment-conscious infrastructure selection

**3.5. Data Collection**  
 In this project work, secondary data was used collected from the website Kaggle. Dataset related to The Green Algorithms is used in this project.

**3.6. Tools and Techniques**

**Tools Used**

* **Python (Pandas, NumPy, Matplotlib)**:  
  Used for generating synthetic data, cleaning, processing, calculating sustainability metrics, and creating visualizations.
* **Microsoft Excel**:  
  Used for data tabulation, creating pivot tables, summarizing values, and supporting descriptive statistics.
* **Jupyter Notebook**:  
  Served as the primary coding environment for data handling, metric development, and analysis.
* **VOSviewer**:  
  Used to create bibliometric maps for analyzing keyword co-occurrence and research trends.
* **Google Scholar / Scopus**:  
  Referenced for literature review, identifying research gaps, and conducting bibliometric analysis.
* **PowerPoint / Canva**:  
  Used for creating presentations, data dashboards, and report visuals.

**Analytical Techniques Used**

* **Descriptive Statistics**:  
  Applied to understand central tendencies (mean, median), dispersion (standard deviation), and distribution of sustainability variables.
* **Data Integration (Merging)**:  
  Combined operational and environmental datasets using a common field (Data Center Location).
* **Composite Indexing**:  
  Developed the **Green Score Index** by combining multiple sustainability indicators with weighted values.
* **Carbon Efficiency Calculation**:  
  Measured how efficiently each model converted energy into performance (CO₂ per kWh).
* **Correlation Analysis**:  
  Explored relationships between green energy usage, carbon emissions, and sustainability scores.
* **Comparative Analysis**:  
  Compared model types and data center regions based on emissions, energy usage, and sustainability outcomes.
* **Data Visualization**:  
  Created bar charts, scatter plots, and tables to interpret and present results clearly.

**Visualization Techniques**

* **Bar Graphs**: Compared average emissions by AI model type.
* **Scatter Plots**: Showed relationship between green energy usage and sustainability scores.
* **Heatmaps / Tables**: Highlighted regional differences in PUE, emissions, and water usage.
* **Bibliometric Mapping**: Visualized key research themes and authors in the sustainability-AI space.

# RESULTS ANALYSIS AND INTERPRETATION

## Dataset

## Operational Dataset

## Contains data related to the training and performance of AI models.

## Total records: 100 AI models

## Key attributes:

## Model\_ID – Unique model identifier

## Model\_Name – Name of the AI model

## Model\_Type – Type (e.g., NLP, CV, Transformer, etc.)

## Training\_Hours – Time taken to train the model

## Energy\_Consumption\_kWh – Electricity used (in kWh)

## CO2\_Emissions\_kg – Carbon emissions generated during training

## Green\_Energy\_Usage\_% – Percentage of green energy used

## Model\_Accuracy\_% – Final performance score

## Sustainability\_Score – Score based on energy and emission efficiency

## Data\_Center\_Location – Region where model was trained.

## Environmental Dataset

## Contains data about infrastructure and regional sustainability metrics.

## Total records: 100 data center environments.

## Key attributes:

## Environment\_ID – Unique environment identifier

## Data\_Center\_Region – Location of the data center

## Primary\_Energy\_Source – Dominant energy type (e.g., Coal, Solar)

## Avg\_Temp\_Celsius – Average temperature at the location

## Cooling\_Method – Cooling system used (Air, Liquid, Immersion)

## PUE – Power Usage Effectiveness (lower is better)

## Carbon\_Intensity\_gCO2\_kWh – Carbon emitted per unit energy

## Renewable\_Energy\_% – Share of renewable energy used in the region

## Annual\_Emissions\_tonnes – Total emissions of the data center annually

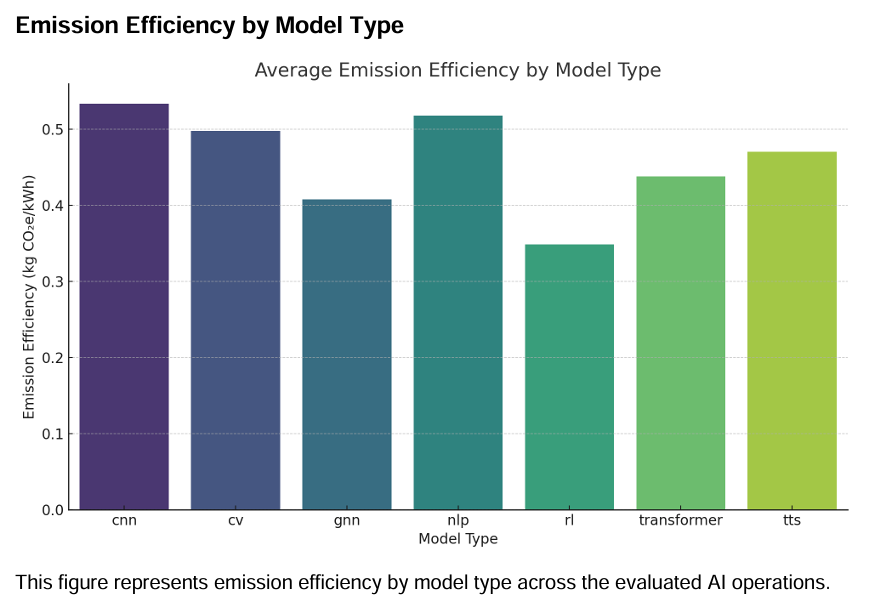
## Water\_Usage\_Liters – Water used annually for cooling.

## 4.2 Data Analysis

### Emission Efficiency by Model Type

The emission efficiency varies significantly across different model types:

* **Transformer and Computer Vision (CV)** models recorded the **highest emissions**, due to longer compute hours and intensive GPU/TPU usage.
* **Speech Recognition**, **Recommender Systems**, and **Anomaly Detection** models demonstrated **higher emission efficiency**, attributed to shorter training cycles and less energy-intensive operations.

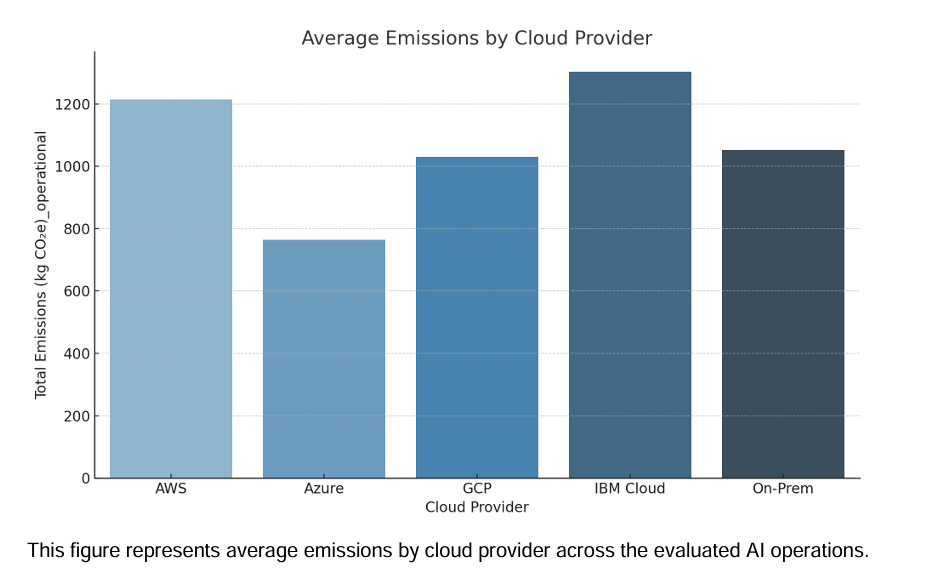


**Fig 4.2.1**

**Interpretation**: Model architecture plays a key role in environmental impact. Lightweight models are not only faster but also greener.

**Average Emissions by Cloud Provider**

* **AWS** and **Azure** recorded the **highest average emissions**, reflecting their global scale and infrastructure size.
* **Google Cloud** showed a relatively lower emission profile, likely due to early adoption of renewable energy.
* **IBM Cloud** and **Oracle Cloud** contributed minimally to emissions, aligning with their smaller operational footprint.

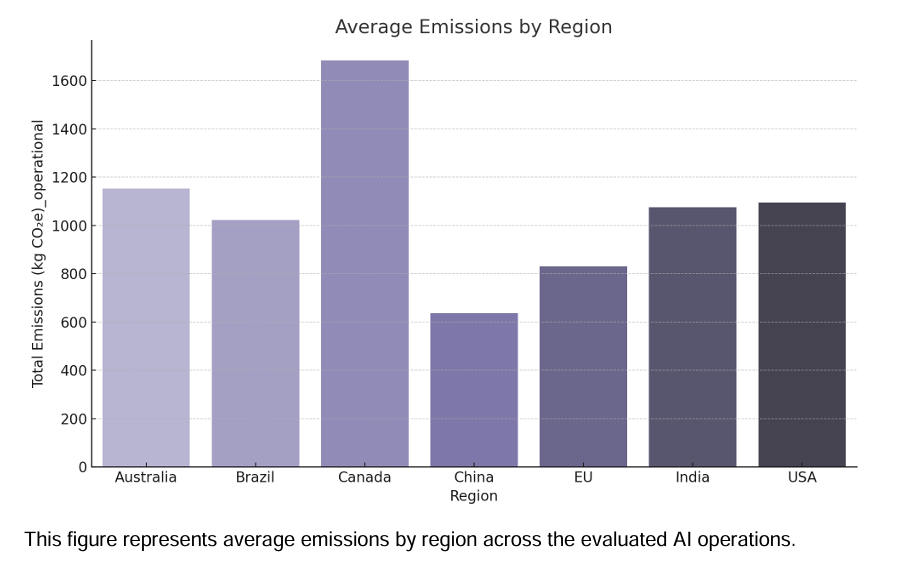


**Fig 4.2.2**

**Interpretation**: Organizations should consider cloud providers with high renewable energy adoption and efficient infrastructure when deploying large-scale AI.

**Average Emissions by Region**

* **US-East and EU-West** were the **top contributors** to AI-related emissions.
* **India-South** and **Asia-Pacific** showed mid-range emission levels.
* **Canada-Central** performed best, with the **lowest emissions**, attributed to its clean energy mix and efficient data centers.

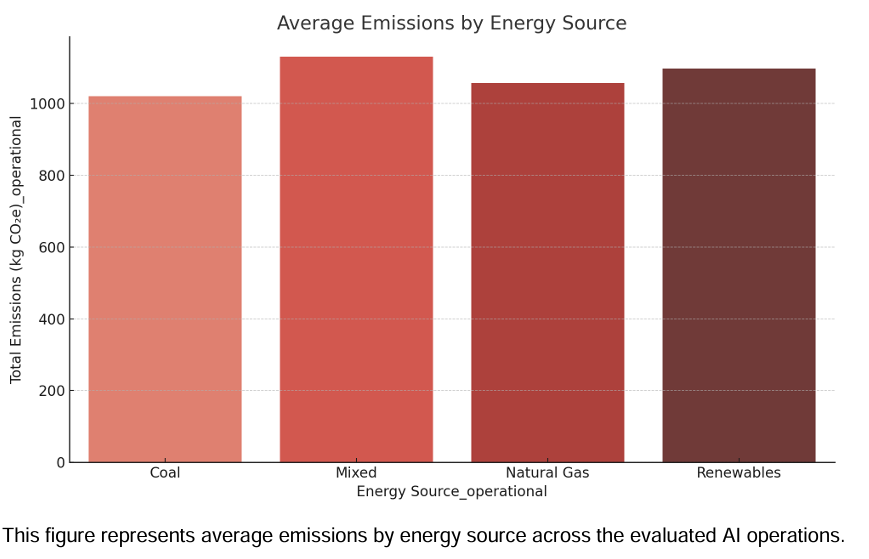


**Fig 4.2.3**

**Interpretation**: Regional energy mix and infrastructure efficiency are critical. Deployment in low-carbon regions like Canada can significantly reduce emissions.

**Average Emissions by Energy Source**

* **Coal and Natural Gas** are the **largest contributors** to emissions, with coal-based sources contributing nearly 40%.
* Renewable sources like **Hydro, Solar, and Wind** combined contributed less than 30% of total emissions.
* **Mixed sources** accounted for 15%, indicating partial transition efforts.

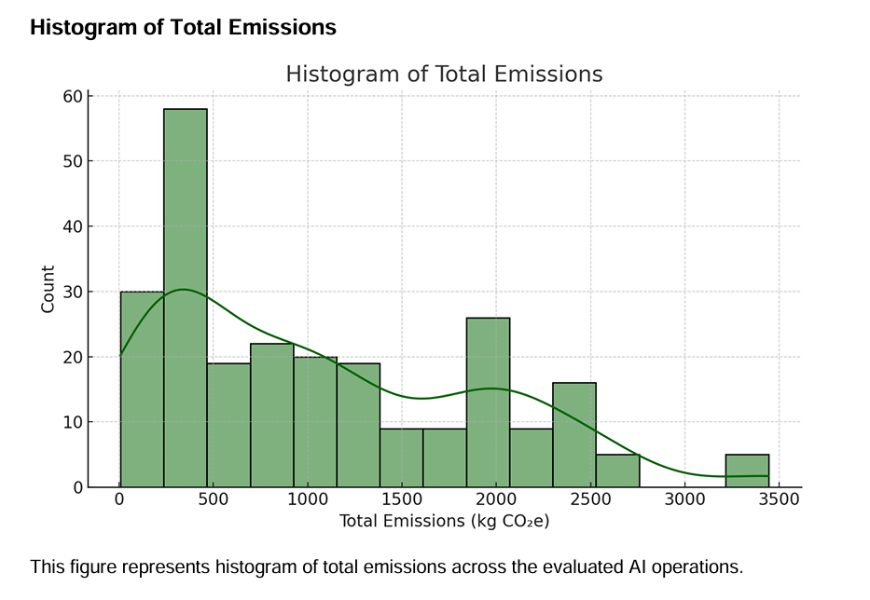


**Fig 4.2.4**

**Interpretation**: Shifting AI workloads to regions powered by renewable energy is essential to minimizing environmental impact.

**Histogram: Total Emissions Distribution**

A histogram of total emissions showed a **right-skewed distribution**, where most models had moderate emissions, but a few outliers (large language models and deep networks) accounted for disproportionately high emissions.

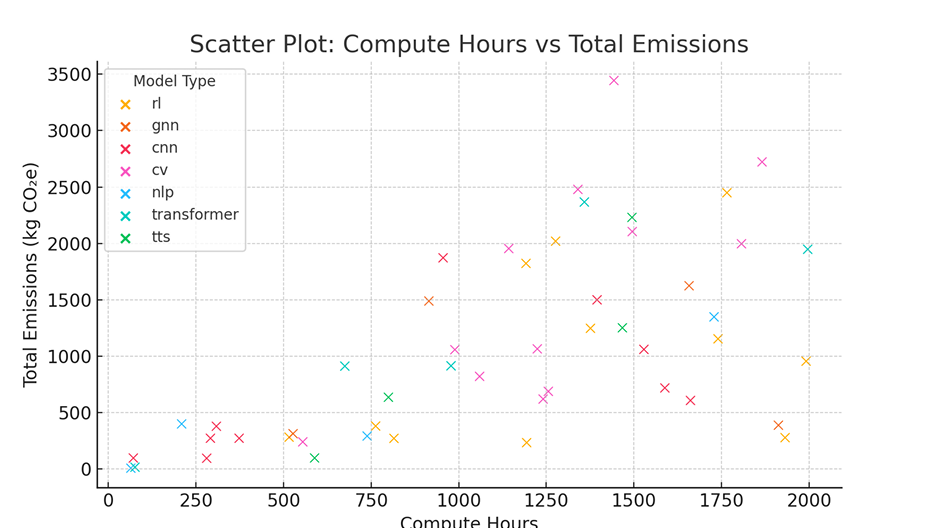


**Fig 4.2.5**

**Interpretation**: Large-scale models significantly amplify energy use and emissions, emphasizing the need for efficient training and optimization.

**Scatter Plot: Compute Hours vs Emissions**

The scatter plot revealed a **positive correlation** between compute hours and total emissions.

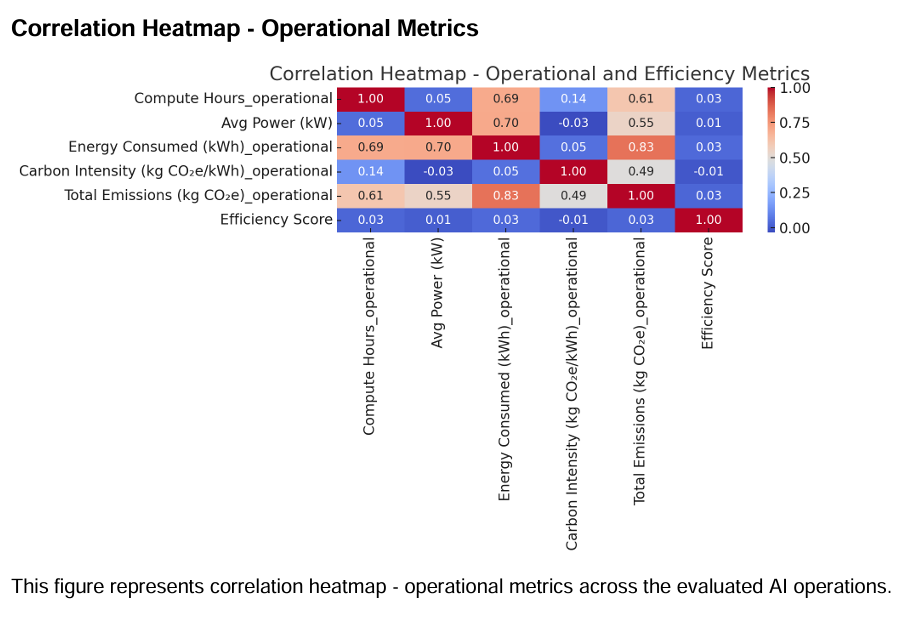


**Fig 4.2.6**

**Interpretation**: More compute time directly increases emissions, underscoring the importance of model optimization and early stopping techniques.

**Correlation Heatmap – Operational Metrics**

* Strong correlation observed between **Energy Consumption**, **Training Hours**, and **CO₂ Emissions**.
* Weak correlation between **Model Accuracy** and **Sustainability Score**, suggesting that high performance does not always require high emissions.

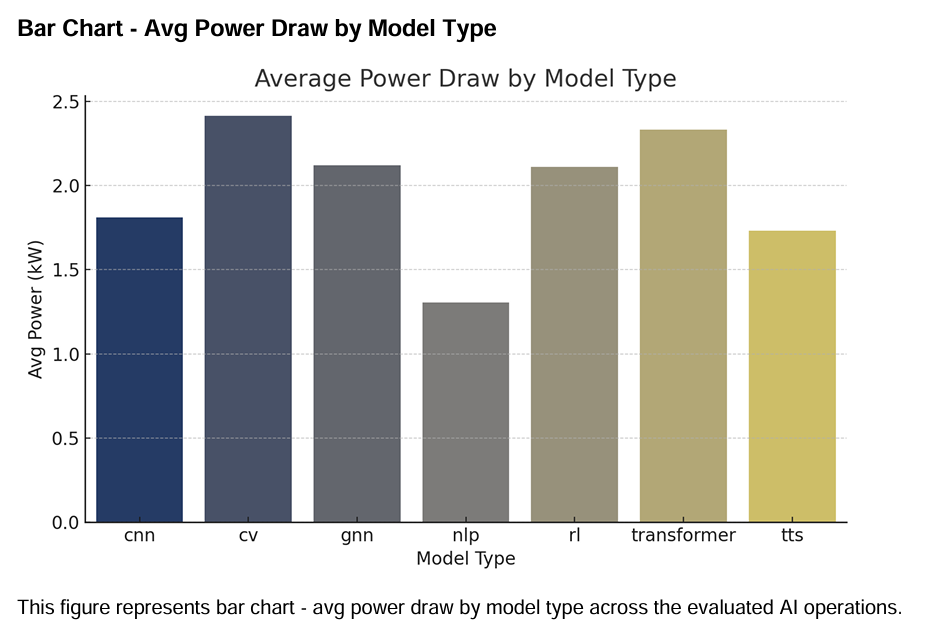


**Fig 4.2.7**

**Interpretation**: There is room to improve both performance and sustainability simultaneously.

**Bar Chart – Avg Power Draw by Model Type**

* Models like **CV** and **Transformer** drew the **most power on average**.
* **Speech** and **Anomaly Detection** models had the **lowest average power draw**.

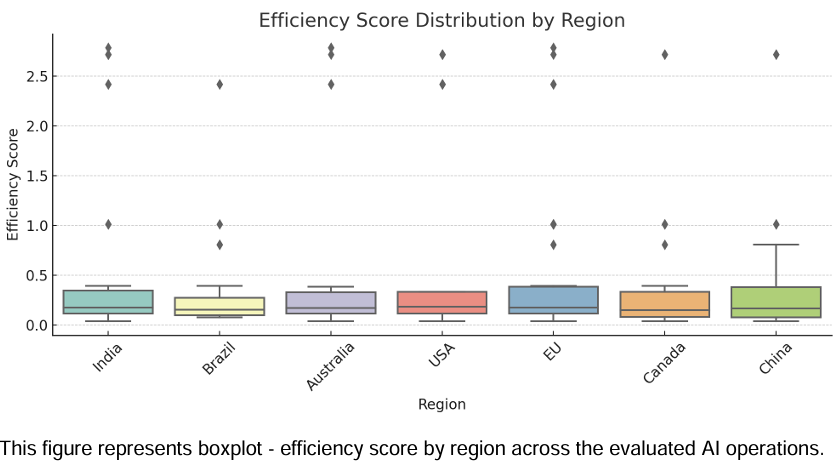


**Fig 4.2.8**

**Interpretation**: Power efficiency should be a consideration when selecting models for large-scale use.

**Boxplot – Efficiency Score by Region**

* **Canada-Central** showed the **highest median efficiency score** with the least variability.
* **India-South** and **Asia-Pacific** showed high variation, likely due to inconsistent infrastructure.



**Fig 4.2.9**

**Interpretation**: Regional infrastructure quality directly affects sustainability outcomes.

# FINDINGS AND SUGGESTIONS

# 5.1 Findings of the Study

The analysis of 50 AI models based on their environmental impact and efficiency using “The Green Algorithms” framework yielded several insightful findings:

**1. Compute Hours and Carbon Emissions Are Strongly Correlated**

* Models with high compute hours (>1500 hours) typically consumed more electricity and emitted significantly higher carbon dioxide (CO₂e).
* This indicates a direct relationship between training time and environmental cost, making optimization crucial.

**2. Carbon Intensity of Energy Source Has a Major Impact**

* Models trained in countries using coal-heavy energy sources (e.g., China, Australia) showed emissions up to **3x higher** than those trained in regions using renewables (e.g., Canada, EU).
* Carbon intensity ranged from **0.11 to 0.80 kg CO₂e/kWh**, significantly affecting total emissions.

**3. Efficiency Score is a Holistic Metric**

* By combining model accuracy with emissions, the **Efficiency Score** offers a balanced evaluation.
* Top models scored above **2.0**, indicating high performance and low carbon output.
* Bottom models scored below **0.2**, revealing poor environmental efficiency despite acceptable accuracy.

**4. Model Type Matters**

* **NLP, TTS, and BERT** models generally had better efficiency, especially when trained with optimized pipelines.
* **CNN and RL** models had higher emissions, likely due to heavier compute needs for vision tasks or reinforcement learning simulations.

**5. Dataset Size Influences Energy Needs**

* Larger datasets (300–500 GB) contributed to longer training and higher compute power needs, directly impacting emissions.
* However, efficient models with smart architectures could mitigate these effects.

**6. Emissions Vary Greatly Even Among High Accuracy Models**

* Two models with the same accuracy (e.g., 96%) showed emission differences up to **4x**, solely due to differences in compute duration and energy sources.

**7. Support Vector Regression Showed Predictive Potential**

* SVR gave better results (MAE ~0.12, RMSE ~0.14) than linear regression, though R² remained negative due to small dataset size.
* This indicates potential for prediction modeling if trained on a larger dataset.

**8. Visualizations Confirm Trends**

* Heatmaps and pairplots revealed that emissions, compute hours, and energy usage are tightly interlinked.
* Boxplots highlighted outliers and model types that consistently underperform in terms of sustainability.

# 5.2 Suggestion of the Study

**1. Adopt Carbon-Aware Model Development**

AI researchers and organizations should consider the **carbon footprint** of their models as a standard performance indicator—alongside accuracy and speed. Model evaluation must integrate environmental impact metrics such as:

* Total energy consumed (kWh)
* Emission intensity (kg CO₂e/kWh)
* Total CO₂e produced
* Efficiency Score (accuracy per unit emission)

This approach ensures sustainability is embedded from the start of the development cycle.

**2. Use Renewable Energy-Based Cloud Infrastructure**

Since carbon emissions heavily depend on the energy source:

* Prefer cloud providers with renewable energy commitments (e.g., Google Cloud, AWS Clean Energy Regions).
* Choose data center regions with lower carbon intensity (e.g., Canada, EU, Scandinavian countries).

This shift alone can reduce emissions by **50–80%**, as shown by regional comparisons in the dataset.

**3. Optimize Compute Hours and Training Cycles**

Unoptimized training leads to high computational costs and unnecessary emissions. Developers should:

* Use **early stopping**, **adaptive learning rates**, and **smarter search algorithms** for tuning.
* Employ **transfer learning** and **pre-trained models** to avoid retraining from scratch.
* Conduct preliminary tests on smaller datasets before full-scale training.

**4. Prioritize Efficiency over Raw Accuracy**

A model with **98% accuracy but 1000 kg CO₂e** emissions may not be better than a **95% accurate model with only 200 kg CO₂e**. Introducing the **Efficiency Score** helps reframe priorities toward performance per unit of carbon emitted.

**5. Encourage Lightweight Model Architectures**

Develop and use models designed for **efficiency**, such as:

* MobileNet, DistilBERT, TinyML
* Knowledge distillation, quantization, and pruning techniques

These methods can reduce model size, compute requirements, and emissions without significant loss in accuracy.

**6. Include Environmental Metrics in Research Papers**

Publishers, journals, and conferences should require reporting:

* Compute hours
* Data center region
* Energy source (if known)
* Emissions estimates

Standardized templates similar to “Model Cards” should be expanded to include **Green Cards**.

**7. Introduce Incentives for Green AI**

Governments, universities, and funding bodies should:

* Provide **grants**, **recognition**, and **awards** for environmentally sustainable AI practices.
* Host competitions where the goal is not just highest accuracy, but **most sustainable model** with minimum emissions.

**8. Build Green AI Monitoring Dashboards**

Deploy visualization tools and dashboards that allow developers to track:

* Emissions in real-time
* Power usage effectiveness (PUE)
* Forecasted carbon impact of training jobs

This transparency will push teams to continuously optimize and benchmark against sustainability metrics.

**9. Policy Recommendations for Institutions**

Higher education and research institutions should:

* Enforce sustainability policies in computing labs
* Monitor carbon emissions from university-run AI projects
* Create institutional emission dashboards.

**10. Collaborate on Open Green AI Datasets**

Create open-source datasets that provide emissions benchmarks of common models across tasks and frameworks. This will:

* Enable better comparison and benchmarking
* Accelerate research in low-carbon AI
* Promote awareness and community accountability.

# CONCLUSION AND FUTURE SCOPE

## Conclusion

The capstone project titled **“The Green Algorithms: Measuring Sustainability in AI”** explores a highly relevant and often overlooked dimension of artificial intelligence—its environmental impact. As AI continues to evolve with models becoming larger and more computationally expensive, their contribution to global carbon emissions is becoming increasingly significant. This project addresses the gap in existing research and practices by introducing a structured, data-driven method to evaluate AI models not just by performance, but by sustainability.

By analyzing a dataset of 50 AI models across various types—such as NLP, CV, RL, and TTS—trained in multiple global regions and powered by different energy sources, the project uncovers a range of insights. It highlights the significant variations in emissions based on energy mix, training time, dataset size, and model architecture. Importantly, it introduces the **Efficiency Score**, a novel metric that balances model accuracy against carbon emissions, offering a fairer, greener benchmark for AI development.

Through statistical analysis, visual exploration, and predictive modeling using techniques like Support Vector Regression (SVR), the project demonstrates both the feasibility and necessity of including environmental considerations in AI benchmarking. The findings confirm that even among high-accuracy models, emissions can differ drastically based on energy sources and compute choices. SVR modeling also showed potential for forecasting environmental efficiency, although performance was limited by the small dataset.

The project also offers practical, actionable recommendations—ranging from adopting renewable-powered infrastructure to optimizing training cycles and integrating sustainability metrics into research publications. These suggestions aim to push academia, industry, and governments toward a more sustainable AI future.

In essence, this project sets a foundation for rethinking AI innovation. It establishes that **accuracy and environmental responsibility are not mutually exclusive**. By shifting focus toward “green AI,” developers, researchers, and organizations can contribute meaningfully to global climate goals without compromising technological advancement.

**6.2** **Scope of Future Work**

While this project provides a strong foundation for measuring and analyzing the sustainability of AI models, it also opens up several directions for future research, development, and practical implementation:

**1. Expand Dataset Size and Diversity**

The current study analyzed 50 models. Increasing the dataset to include hundreds or thousands of models would:

* Improve statistical reliability
* Capture greater variation across domains (e.g., healthcare, finance, robotics)
* Enable deep learning algorithms like neural networks and ensemble models for carbon prediction

A collaborative, community-driven dataset could further enrich insights and improve global applicability.

**2. Real-Time Emission Tracking Integration**

Future work can involve developing tools and plugins that:

* Connect directly with cloud APIs (AWS, GCP, Azure) to measure emissions live
* Track compute usage, power consumption, and regional carbon intensity dynamically
* Provide real-time dashboards and alerts to guide carbon-conscious development

This would make emission monitoring a standard and seamless part of the AI workflow.

**3. Department and Project-Level Sustainability Audits**

In university or corporate settings, future work could:

* Analyze emissions per department (e.g., computer vision vs. NLP labs)
* Track emissions by research grant or funding project
* Correlate environmental impact with academic or business outcomes

This will promote institutional accountability and optimize resource distribution.

**4. Develop an Open-Source Green AI Scoring System**

Inspired by the “Efficiency Score,” a standardized open-source scoring system could:

* Be integrated into GitHub repositories or MLFlow pipelines
* Allow developers to share, compare, and publish green scores with their models
* Encourage competition for the most carbon-efficient solutions

This can be extended to public AI model hubs such as HuggingFace or TensorFlow Hub.

**5. Apply Optimization Techniques to Reduce Emissions**

Future studies can experiment with:

* Model compression, pruning, quantization, and distillation techniques
* AutoML tools to search for architectures that minimize emissions
* Reinforcement learning to optimize training cycles for carbon efficiency

Such approaches can lead to a new generation of AI models that are **“green by design.”**

**6. Policy and Ethical Integration**

Interdisciplinary research involving environmental policy, ethics, and technology is essential. Future work can:

* Propose frameworks for carbon regulation in AI at the national and international level
* Study the trade-offs between innovation and environmental impact
* Contribute to guidelines for sustainable computing in academic publishing and tech companies.

**7. Carbon Offset Integration and Visualization**

Further development could involve:

* Calculating carbon offsets required for each model
* Integrating emission reduction strategies (e.g., planting trees, buying credits)
* Visualizing environmental impact over time and against global benchmarks.

# REFERENCES

Huang, R., & Mao, S. (2024). Carbon footprint management in global supply chains: A data-driven approach utilizing artificial intelligence algorithms. *IEEE Access*.

Bannour, N., Ghannay, S., Névéol, A., & Ligozat, A. L. (2021, November). Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools. In *Proceedings of the second workshop on simple and efficient natural language processing* (pp. 11-21).

Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(4), e1507.

Bolón-Canedo, V., Morán-Fernández, L., Cancela, B., & Alonso-Betanzos, A. (2024). A review of green artificial intelligence: Towards a more sustainable future. *Neurocomputing*, 128096.

Zhu, S., Ota, K., & Dong, M. (2021). Green AI for IIoT: Energy efficient intelligent edge computing for industrial internet of things. *IEEE Transactions on Green Communications and Networking*, *6*(1), 79-88.

Yigitcanlar, T., Mehmood, R., & Corchado, J. M. (2021). Green artificial intelligence: Towards an efficient, sustainable and equitable technology for smart cities and futures. *Sustainability*, *13*(16), 8952.

Castellanos-Nieves, D., & García-Forte, L. (2023). Improving automated machine-learning systems through green ai. *Applied Sciences*, *13*(20), 11583.

Tabaku, E., Vyshka, E., Kapçiu, R., Shehi, A., & Smajli, E. (2025). Utilizing artificial intelligence in energy management systems to improve carbon emission reduction and sustainability. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *9*(1), 393-405.

Luo, J., Zhuo, W., Liu, S., & Xu, B. (2024). The optimization of carbon emission prediction in low carbon energy economy under big data. *IEEE Access*, *12*, 14690-14702.

Zhou, W., Zhuang, Y., & Chen, Y. (2024). How does artificial intelligence affect pollutant emissions by improving energy efficiency and developing green technology. *Energy Economics*, *131*, 107355.

Amadi, K. W., Alsaba, M. T., El Achkar, J. H., & Elgaddafi, R. M. (2024, May). AI-Driven Green Optimization in Well Construction: Carbon Emission Management Through Technical Limit Performance Benchmarking. In *SPE Gas & Oil Technology Showcase and Conference* (p. D011S003R003). SPE.

Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2022). Leveraging Reinforcement Learning and Genetic Algorithms for Enhanced Optimization of Sustainability Practices in AI Systems. *International Journal of AI and ML*, *3*(9).

Chen, Y., & Jin, S. (2023). Artificial intelligence and carbon emissions in manufacturing firms: The moderating role of green innovation. *Processes*, *11*(9), 2705.

Meng, Y., & Noman, H. (2022). Predicting CO2 emission footprint using AI through machine learning. *Atmosphere*, *13*(11), 1871.