TWT

THE ENVIRONMENTAL IMPACT OF COMPUTING & LLMs

Tina Vartziotis[1,2,3], Michael Keckeisen[3], Patrick Vaudrevange[3], Francesca Dominici[2]

[1] National Technical University of Athens (NTUA), Greece
[2] Harvard University, USA
[3] TWT Science & Innovation, Stuttgart, Germany

June 29th - July 4th, 2026

SummerSOC 2026

SummerSoc
Service Oriented Computing

Image Source: TWT, SummerSOC

TWT

CONTENT

TWT SCIENCE & INNOVATION

01

| WE TRANSFER KNOWLEDGE – INTO MARKET-READY TECHNOLOGIES |

TECHNICAL AND SCIENTIFIC TRANSFER

AI & Data Science, Simulation, Processes,
Neuromorphic & Quantum Computing,
Sustainability Engineering

Highly qualified professionals

300

from 17 nations – academic, interdisciplinary, excellent

Partner organizations

100+

from leading universities, global industry players, and our trusted partner in Greece

Self-financed, very strong credit rating

100%

independent – for over 25 years

Research Projects

100+

innovative, application-oriented, and future-oriented

Years as a technology pioneer

40

in technical and scientific transfer

Professors

20+

former TWT staff now in academia and research

MANAGEMENT: Dr. Dimitris Vartziotis, Dipl.-Ing. Frank Beutenmüller, Dr. Michael Keckeisen.

| ACTIVE RESEARCH PROJECT |

QAIAC: QUANTUM ARTIFICIAL INTELLIGENCE FOR THE AUTOMOTIVE INDUSTRY AND CLEAN ENERGY

Goal

Comparative Performance Evaluation of the Potential Quantum Advantage of Quantum Machine Learning in the Automotive Value Chain and Renewable Energy Generation.

Approach

Using computationally intensive optimization problems as a case study, this research aims to narrow the gap between the theoretically achievable and the practically realizable quantum advantage.

Duration and Funding Agency

Duration: 03/2026 – 02/2028

Federal Ministry
of Research, Technology
and Space

TWT Contribution

- Quantum-assisted AI surrogate modeling for chassis design
- Comparison with classical machine learning algorithms
- Testing Quantum Machine Learning (QML) algorithms on NISQ quantum hardware
- Evaluation of the advantages of QML for vehicle route planning with EVERIS

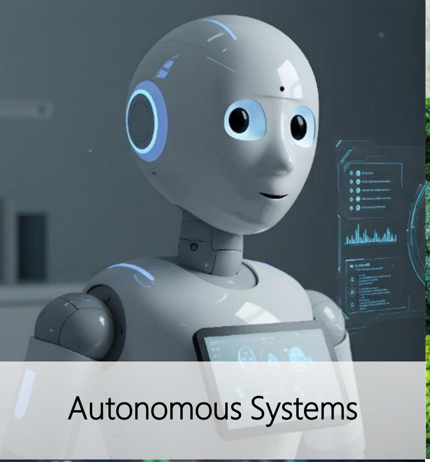
Image Source: Mercedes-Benz, FZJ

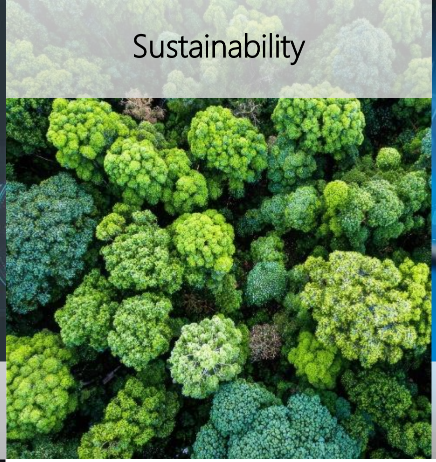
Partners

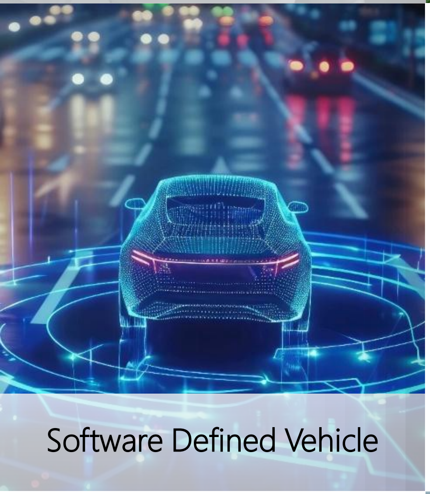
e·on

JÜLICH
Forschungszentrum

Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

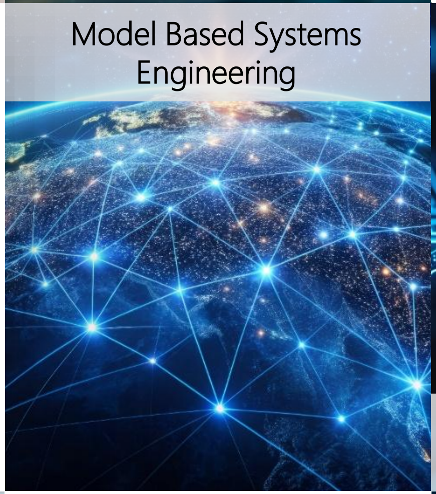


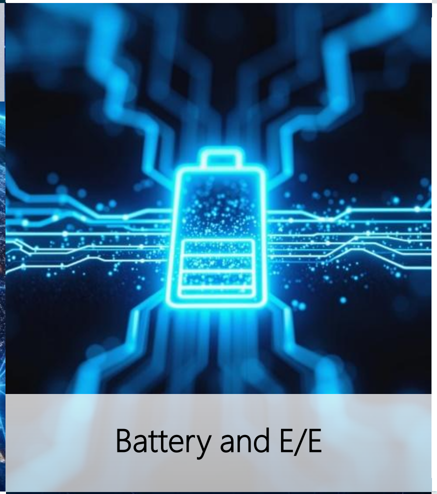


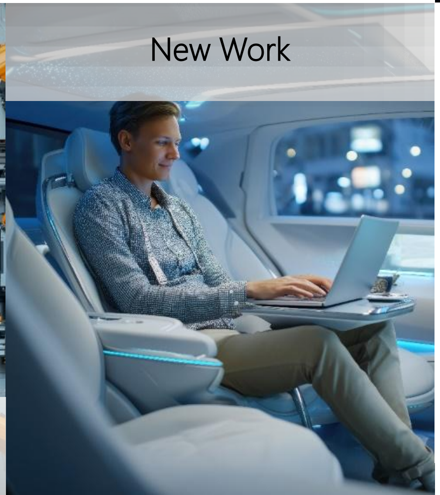


PORTFOLIO

Interdisciplinary range of services







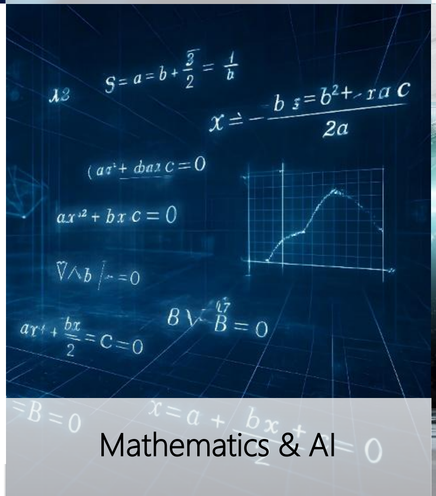


Image Source: TWT

ADDED VALUE THROUGH RESEARCH & AWARDS

Framework Conditions for Application and Funding

- Application volume: approx. 200,000€ to 1,000,000€
- Duration: approx. 12 to 36 months
- Application processing time: approx. 30 to 60 days
- Effort for writing the application: approx. 120 hours
- Define the topic: Identify a pre-competitive idea
- Identify the appropriate funding program
- Formulate text for dissemination
- Perform the cost calculation
- Write and submit the application
- Receive funding and initiate the project

AITOC
Winner ITEA
Award of Excellence
Business Impact
2024

SPEAR
Winner Eureka
Innovation Awards
„Innovation and Sustainability"
2022

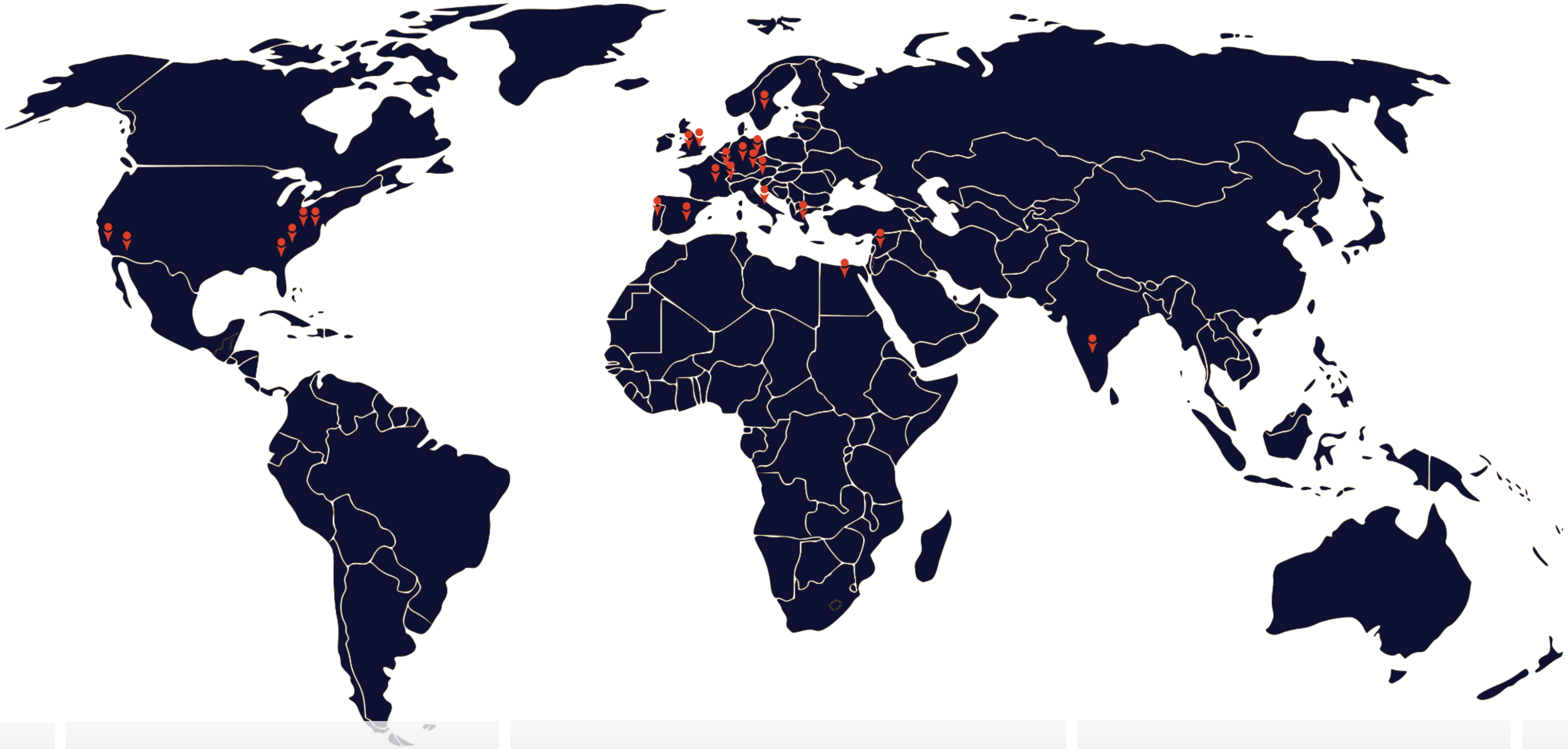
PRODUCTIVE 4.0
Winner ECSEL
JU Technology
Pioneer Award
2021

ENTOC
Winner ITEA
Award of Excellence
Innovation
2020

MODELISAR
Winner ITEA
Achievement Award
Silver
2012

TWT – HOW WE WORK TOGETHER

From Research to Industrial Applications with over 200 Academic Partners across more than 50 Institutions.



Partnerships in Research Projects	Consulting in Industrial Projects	Student Positions (Interns; Master's Thesis)	Graduates (Full-Time, PhD, Freelancers)	Lectures and Laboratory Practicals

Image Source: Adobe Stock

University of Stuttgart · TUM · KIT · Universitätsbund Tübingen e. V. · Universität Augsburg University · tu · ETH zürich · University of Ioannina · TU Wien · GJU German Jordanian University · University of Oxford · UC Berkeley · KTH · École Polytechnique · University of Virginia · Columbia University in the City of New York · MIT · Harvard University

TWT

TWT – SELECTED PUBLICATIONS

Integrating Environmental Health in the AI Era: An Urgent Imperative

Francesca Dominici[1] , Tina Vartziotis[2]

[1] Harvard University, Cambridge, MA , USA

[2] TWT GmbH Science & Innovation, Stuttgart, Germany

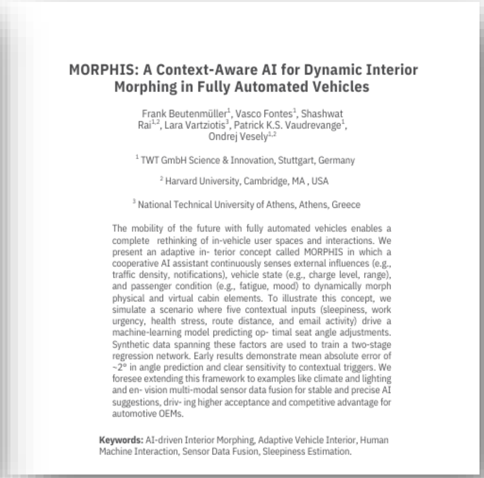
Artificial intelligence is growing rapidly, and so are the large data centers that are needed to support it. This progress brings real benefits, but it also has hidden costs: more electricity use, more demand on local water supplies, and increased levels of exposure to air pollution, often in places where people are already facing health burdens. This commentary lays out a practical plan for responsible AI for people and the planet. We call for clear and public reporting from every data center on energy use, emissions, and water; for honest accounting of health risks before projects get approved; for siting decisions that favor cleaner power and cleaner grids; and for open, trustworthy data and tools so communities, journalists, and researchers can verify claims. The message is simple: We can have innovation without sacrificing public health. By setting smart rules, engaging local communities, and measuring what matters - not just carbon but also air quality, water, and who bears the impacts - we can build AI in a way that protects lives, reduces harm, and shares the benefits more fairly.

Keywords: Environmental Intelligence, Artificial intelligence, Public Health, Green AI, Green Coding, Sustainability.

Integrating Environmental Health in the AI Era: An Urgent Imperative

Harvard University
TWT Science & Innovation

Harvard Data Science Review, 7(4), 2025



MORPHIS: A Context-Aware AI for Dynamic Interior Morphing in Fully Automated Vehicles

Frank Beutenmüller[1], Vasco Fontes[1], Shashwat Rai[1,2], Lara Vartziotis[3], Patrick K.S. Vaudrevange[1], Ondrej Vesely[1,2]

[1] TWT GmbH Science & Innovation, Stuttgart, Germany

[2] Harvard University, Cambridge, MA , USA

[3] National Technical University of Athens, Athens, Greece

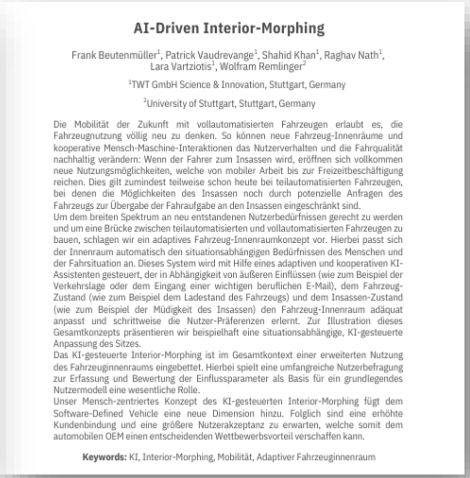
The mobility of the future with fully automated vehicles enables a complete rethinking of in-vehicle user spaces and interactions. We present an adaptive in- terior concept called MORPHIS in which a cooperative AI assistant continuously senses external influences (e.g., traffic density, notifications), vehicle state (e.g., charge level, range), and passenger condition (e.g., fatigue, mood) to dynamically morph physical and virtual cabin elements. To illustrate this concept, we simulate a scenario where five contextual inputs (sleepiness, work urgency, health stress, route distance, and email activity) drive a machine-learning model predicting op- timal seat angle adjustments. Synthetic data spanning these factors are used to train a two-stage regression network. Early results demonstrate mean absolute error of ~2° in angle prediction and clear sensitivity to contextual triggers. We foresee extending this framework to examples like climate and lighting and en- vision multi-modal sensor data fusion for stable and precise AI suggestions, driv- ing higher acceptance and competitive advantage for automotive OEMs.

Keywords: AI-driven Interior Morphing, Adaptive Vehicle Interior, Human Machine Interaction, Sensor Data Fusion, Sleepiness Estimation.

MORPHIS: A Context-Aware AI for Dynamic Interior Morphing in Fully Automated Vehicles

TWT Science & Innovation
Harvard University
NTUA

Driving the Future Symposium 2025, 2025



AI-Driven Interior-Morphing

Frank Beutenmüller[1], Patrick Vaudrevange[1], Shahid Khan[1], Raghav Nath[1], Lara Vartziotis[1], Wolfram Remlinger[2]

[1] TWT GmbH Science & Innovation, Stuttgart, Germany

[2] University of Stuttgart, Stuttgart, Germany

Die Mobilität der Zukunft mit vollautomatisierten Fahrzeugen erlaubt es, die Fahrzeugnutzung völlig neu zu denken. So können neue Fahrzeug-Innenräume und kooperative Mensch-Maschine-Interaktionen das Nutzerverhalten und die Fahrqualität nachhaltig verändern: Wenn der Fahrer zum Insassen wird, eröffnen sich vollkommen neue Nutzungsmöglichkeiten, welche von mobiler Arbeit bis zur Freizeitbeschäftigung reichen. Dies gilt zumindest teilweise schon heute bei teilautomatisierten Fahrzeugen, bei denen die Möglichkeiten des Insassen noch durch potenzielle Anfragen des Fahrzeugs zur Übergabe der Fahraufgabe an den Insassen eingeschränkt sind.
Um dem breiten Spektrum an neu entstandenen Nutzerbedürfnissen gerecht zu werden und um eine Brücke zwischen teilautomatisierten und vollautomatisierten Fahrzeugen zu bauen, schlagen wir ein adaptives Fahrzeug-Innenraumkonzept vor. Hierbei passt sich der Innenraum automatisch den situationsabhängigen Bedürfnissen des Menschen und der Fahrsituation an. Dieses System wird mit Hilfe eines adaptiven und kooperativen KI-Assistenten gesteuert, der in Abhängigkeit von äußeren Einflüssen (wie zum Beispiel der Verkehrslage oder dem Eingang einer wichtigen beruflichen E-Mail), dem Fahrzeug-Zustand (wie zum Beispiel dem Ladestand des Fahrzeugs) und dem Insassen-Zustand (wie zum Beispiel der Müdigkeit des Insassen) den Fahrzeug-Innenraum adäquat anpasst und schrittweise die Nutzer-Präferenzen erlernt. Zur Illustration dieses Gesamtkonzepts präsentieren wir beispielhaft eine situationsabhängige, KI-gesteuerte Anpassung des Sitzes.
Das KI-gesteuerte Interior-Morphing ist im Gesamtkontext einer erweiterten Nutzung des Fahrzeuginnenraums eingebettet. Hierbei spielt eine umfangreiche Nutzerbefragung zur Erfassung und Bewertung der Einflussparameter als Basis für ein grundlegendes Nutzermodell eine wesentliche Rolle.
Unser Mensch-zentriertes Konzept des KI-gesteuerten Interior-Morphing fügt dem Software-Defined Vehicle eine neue Dimension hinzu. Folglich sind eine erhöhte Kundenbindung und eine größere Nutzerakzeptanz zu erwarten, welche somit dem automobilen OEM einen entscheidenden Wettbewerbsvorteil verschaffen kann.

Keywords: KI, Interior-Morphing, Mobilität, Adaptiver Fahrzeuginnenraum

AI-Driven Interior-Morphing

TWT Science & Innovation
University of Stuttgart

VDI conference AI in Vehicles & Software Engineering 2025, 2025

Konzept des mobilen Office in (teil-) automatisierten Fahrzeugen, erlebt in einer VR-Simulation

F. Beutenmüller[1], Dr. N. Reining[1], M. Schmidt[1], S. Ferrari[1], L. Vartziotis[2], Dr. Ing. M. Bues[3]

[1] TWT GmbH Science & Innovation, Stuttgart, Germany

[2] National Technical University of Athens, Athens, Greece

[3] Fraunhofer IAO, Stuttgart, Germany

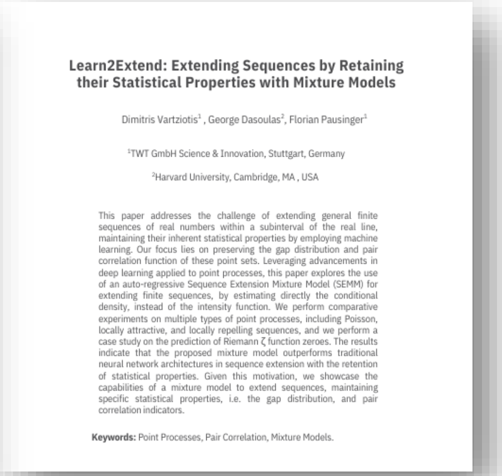
The mobility of the future with fully automated vehicles enables a complete rethinking of in-vehicle user spaces and interactions. We present an adaptive in- terior concept called MORPHIS in which a cooperative AI assistant continuously senses external influences (e.g., traffic density, notifications), vehicle state (e.g., charge level, range), and passenger condition (e.g., fatigue, mood) to dynamically morph physical and virtual cabin elements. To illustrate this concept, we simulate a scenario where five contextual inputs (sleepiness, work urgency, health stress, route distance, and email activity) drive a machine-learning model predicting op- timal seat angle adjustments. Synthetic data spanning these factors are used to train a two-stage regression network. Early results demonstrate mean absolute error of ~2° in angle prediction and clear sensitivity to contextual triggers. We foresee extending this framework to examples like climate and lighting and en- vision multi-modal sensor data fusion for stable and precise AI suggestions, driv- ing higher acceptance and competitive advantage for automotive OEMs.

Keywords: AI-driven Interior Morphing, Adaptive Vehicle Interior, Human Machine Interaction, Sensor Data Fusion, Sleepiness Estimation.

Konzept des mobilen Office in (teil-) automatisierten Fahrzeugen, erlebt in einer VR-Simulation

TWT Science & Innovation
NTUA
Fraunhofer IAO

38. VDI-Tagung „Fahrerassistenzsysteme und Automatisiertes Fahren“, 2025



Learn2Extend: Extending Sequences by Retaining their Statistical Properties with Mixture Models

Dimitris Vartziotis[1] , George Dasoulas[2], Florian Pausinger[1]

[1] TWT GmbH Science & Innovation, Stuttgart, Germany

[2] Harvard University, Cambridge, MA , USA

This paper addresses the challenge of extending general finite sequences of real numbers within a subinterval of the real line, maintaining their inherent statistical properties by employing machine learning. Our focus lies on preserving the gap distribution and pair correlation function of these point sets. Leveraging advancements in deep learning applied to point processes, this paper explores the use of an auto-regressive Sequence Extension Mixture Model (SEMM) for extending finite sequences, by estimating directly the conditional density, instead of the intensity function. We perform comparative experiments on multiple types of point processes, including Poisson, locally attractive, and locally repelling sequences, and we perform a case study on the prediction of Riemann ζ function zeroes. The results indicate that the proposed mixture model outperforms traditional neural network architectures in sequence extension with the retention of statistical properties. Given this motivation, we showcase the capabilities of a mixture model to extend sequences, maintaining specific statistical properties, i.e. the gap distribution, and pair correlation indicators.

Keywords: Point Processes, Pair Correlation, Mixture Models.

Learn2Extend: Extending Sequences by Retaining their Statistical Properties with Mixture Models

TWT Science & Innovation
Harvard University

Experimental Mathematics 34 (2): 270–281, 2024

TWT

TWT – SELECTED PUBLICATIONS WITH PARTNERS / CUSTOMERS

Introduction of Systems Engineering on a global scale for commercial vehicle development

Carsten Henkelmann[1], Martin Obstbaum[2]

[1]Daimler Truck AG, Baden-Württemberg, Germany

[2]TWT GmbH Science & Innovation, Stuttgart, Germany

The challenges for developing commercial vehicles in terms of mastering complexity, reducing time-to-market, and total cost of ownership are ever growing. At the same time technology advances raise the potential of new features to be engineered into vehicles. For a company integrating business units and brands all over the world the described sources of complexity exponentiate. For highest quality development and innovation of products which span the whole variety of busses and trucks an overarching globally harmonized systems engineering approach is needed. In this talk we present how a transformation from an established component-based into a function-oriented system-based development process can be successfully implemented to foster commonality of parts across all vehicle platforms. The presentation will outline a consistent framework of essential process steps and process roles going from market space through solution neutral towards solution specific specification and design and will show how verification and validation may be integrated. This will be done using tangible feature examples addressing the functional synthesis from features towards systems and how to integrate the different engineering domains needed to develop globally harmonized vehicle platforms. We will also outline the benefits and challenges of introducing systems engineering for both commercial vehicle development in general and worldwide distributed brands with different time-zones.

Keywords: ystems Engineering; Global Commercial Vehicle Development; Engineering Processes

Introduction of Systems Engineering on a global scale for commercial vehicle development

Daimler Truck AG
TWT Science & Innovation

ProSTEP iViP Symposium 2023, 2023

Investigation of Stator Cooling Concepts of an Electric Machine for Maximization of Continuous Power

Mike Reinecke[1], Akif Karayel[1], Hendrik von Schöning[2], Uwe Schaefer[3], Matthias Moullion[2], Victor Faessler[2], Robert Lehmann[1]

[1]Mercedes-Benz AG, Sindelfingen, Germany

[2]TWT GmbH Science & Innovation, Stuttgart, Germany

[3]TU Berlin, Berlin, Germany

With the automotive industry's increasing focus on electromobility and the growing share of electric cars, new challenges are arising for the development of electric motors. The requirements for torque and power of traction motors are constantly growing, while installation space, costs and weight are increasingly becoming limiting factors. Moreover, there is an inherent conflict in the design between power density and efficiency of an electric motor. Thus, a main focus in today's development lies on space-saving and yet effective and innovative cooling systems. This paper presents an approach for a multi-physical optimization that combines the domains of electromagnetics and thermodynamics. Based on a reference machine, this simulative study examins a total of nine different stator cooling concepts varying the cooling duct positions and end-winding cooling concepts. To ensure the highest possible comparability, the rotor geometry as well as the overall dimensions in terms of outer diameter and length of the electric machine remain unchanged. The stator design is slightly adjusted to achieve same maximum torque and winding cross-section. Initially, the electromagnetic effects of various cooling slot positions are investigated and compared with respect to efficiency and individual loss distribution. Subsequently, the thermal performance is analyzed by means of fluid-dynamical simulations to quantify the heat transfer and assess the cooling effectivity. Eventually, these results are merged in a lumped parameter thermal network model. Accounting for both the distinguished electromagnetic and thermal benefits and disadvantages, a final study is presented evaluating the continuous power capability of the different concepts at equal boundary conditions.

Keywords: Electric Vehicles, Electric Motors, Heat Transfer, Thermodynamics, Measurements, Optimization, Logistics, Propellers and Rotors

Investigation of Stator Cooling Concepts of an Electric Machine for Maximization of Continuous Power

Mercedes-Benz AG
TWT Science & Innovation
TU Berlin

SAE International Journal of Advances and Current Practices in Mobility 7(3):1187-1206, 2025

Carbon Footprint Evaluation of Code Generation through LLM as a Service

Tina Vartziotis[1,2], Maximilian Schmidt[1], George Dasoulas[3], Ippolyti Dellatolas[4], Stefano Attademo[1], Viet Dung Le[5], Anke Wiechmann[5], Tim Hoffmann[5], Michael Keckeisen[1] & Sotirios Kotsopoulos[2]

[1] TWT GmbH Science & Innovation, Stuttgart, Germany

[2] National Technical University of Athens, Athens, Greece

[3]Harvard University, Cambridge, MA , USA

[4]Massachusetts Institute of Technology, Cambridge, MA, USA

[5]Mercedes-Benz AG, Sindelfingen, Germany

Due to increased computing use, data centers consume and emit a lot of energy and carbon. These contributions are expected to rise as big data analytics, digitization, and large AI models grow and become major components of daily working routines. To reduce the environmental impact of software development, green (sustainable) coding and claims that AI models can improve energy efficiency have grown in popularity. Furthermore, in the automotive industry, where software increasingly governs vehicle performance, safety, and user experience, the principles of green coding and AI-driven efficiency could significantly contribute to reducing the sector's environmental footprint. We present an overview of green coding and metrics to measure AI model sustainability awareness. This study introduces LLM as a service and uses a generative commercial AI language model, GitHub Copilot, to auto-generate code. Using sustainability metrics to quantify these AI models' sustainability awareness, we define the code's embodied and operational carbon.

Keywords: Computational Linguistics, Computational Intelligence, Computing Milieux, Ecological Modelling, Machine Learning, Natural Language Processing (NLP)

Carbon Footprint Evaluation of Code Generation through LLM as a Service

TWT Science & Innovation
NTUA
Harvard University
MIT
Mercedes-Benz AG

2024 Stuttgart International Symposium on Automotive and Engine Technology, 2024

SpecBook-Copilot
Efficient Formalization of Requirements using Artificial Intelligence in the Development of MB.OS

Dr. Martin Obstbaum[1], Dipl.-Ing. Matthias Staib[2], Dipl.-Ing. Ireneus Wior[1], Dr. Johanna Kasper[1], Dr. habil. Patrick Vaudrevange[1], Dr. Shahid Khan[1], Dipl.-Ing. Rainer Freitag[2], Dr. Michael Keckeisen[1], Dr. Jutta Schneider[2]

TWT GmbH Science & Innovation, Stuttgart, Germany

Mercedes-Benz AG, Sindelfingen, Germany

Mercedes-Benz is committed to the continuous development of modern, sustainable luxury vehicles, a vision that is embodied in MB.OS, a chip-to-cloud electronic architecture that seamlessly connects the vehicle with the cloud and the Internet of Things (IoTs). A key to delivering the desired customer experience is a versatile electronics platform, driven by precise, unique requirements that guide architecture and design. These requirements must cover critical aspects like electromagnetic compatibility (EMC), functional safety, and cybersecurity. Ensuring these requirements are explicit and consistently followed requires formalization to be machine-readable and interpretable. Traditional methods are often inefficient, highlighting the need for an advanced tool like a copilot for requirements engineers. Large Language Models (LLMs) and generative AI, such as ChatGPT, offer a promising solution. This paper presents a proof-of-concept for the SpecBook-Copilot, demonstrating how LLMs can enhance requirements engineering efficiency and reliability. We introduce quality factors for evaluating AI-generated content and discuss how Mercedes-Benz's cloud-AI supports scalable deployment while protecting intellectual property. Additionally, we explore potential applications and interfaces for AI copilots and how formalized requirements can streamline the generation of artifacts like graphical models and test cases.

Keywords: AI-driven Interior Morphing, Adaptive Vehicle Interior, Human Machine Interaction, Sensor Data Fusion, Sleepiness Estimation.

SpecBook-Copilot: Efficient Formalization of Requirements using Artificial Intelligence in the Development of MB.OS

TWT Science & Innovation
Mercedes-Benz AG

***22nd International Congress and Exhibition ELIV 2024*, Bonn, 2024, pp. 221–232.**

Scenario-based Collaborative Robot Evaluation (SCORE) Framework

Ali Al-Yacoub, Alina Peter, Dennis Guck, Felix Fleschhut, and Michael Keckeisen

TWT GmbH Science & Innovation, Stuttgart, Germany

Human-Robot Collaboration Systems (HRCSs) are integral to the automation industry. Robots work alongside humans to enhance productivity and efficiency across an extensive range of applications. In each case, ensuring the safety of HRCSs is crucial, necessitating a robust safety verification process—ideally while maintaining efficiency and optimizing cost. This imposes contradicting requirements. For example, an efficient assembly task might require high robot speed, but high speed can endanger humans. Both the complexity of the, typically manual, safety validation process and the resulting restraints on efficiency can hinder widespread adoption of HRCSs. Optimization in the face of the combinatorial explosion of possible HRCSs implementations and scenarios will require a new approach. Here, we propose a scenario-based safety verification framework for collaborative robots that can systematically generate and analyze scenarios and test simulated control strategies to ensure that human safety is maintained while achieving optimal productivity. The proposed approach consists of a text-based HRCSs operational scenario that describes the scene, hardware setup, and dynamical elements in the scene. Unlike existing methods, the proposed approach is intended to enable automated scenario generation and adaptive risk assessment, providing a scalable and repeatable framework for systematically designing and evaluating collaborative robot safety in dynamic operational settings.

Keywords: collaborative robots, safety verification, scenario-based simulation, human-robot interaction, risk assessment.

Scenario-based Collaborative Robot Evaluation (SCORE) Framework

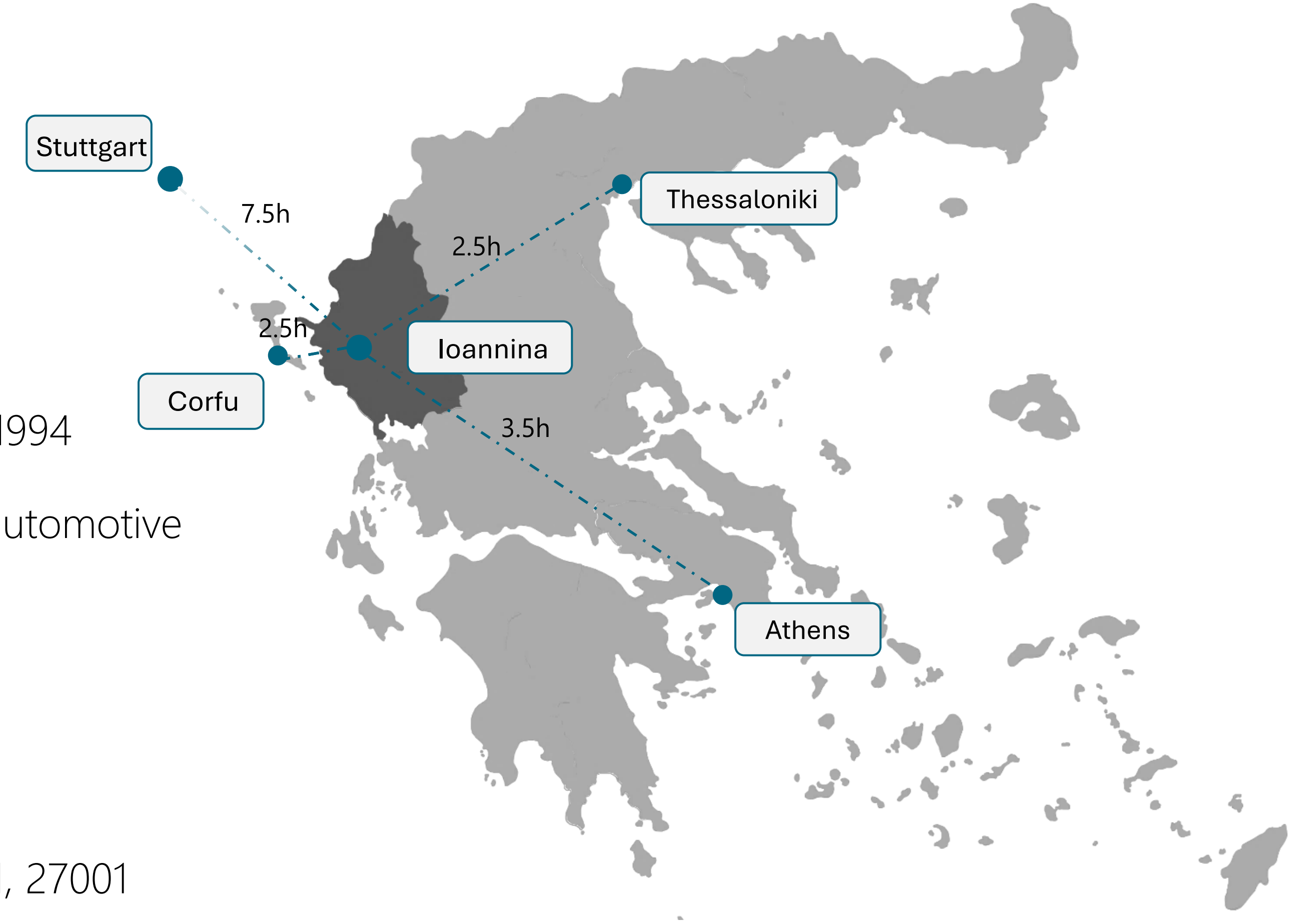
TWT Science & Innovation

10th Changeable, Agile, Reconfigurable and Virtual Production Conference, 2025

YOUR ADVANTAGE – OUR REGIONAL PARTNER NIKI / COOPERATION PARTNER

niki
Digital Engineering

- ✓ Ioannina, Greece · CEO: Dr. D. Vartziotis
- ✓ IT & Digital Engineering · since 1994
- ✓ Projects for TWT at · Premium automotive OEMs
- ✓ European research activities
- ✓ Approx. 60 employees
- ✓ Certifications TISAX, 9001, 14001, 27001
- ✓ Project languages · German and English



30+ years partnership

High number of talents from technical Universities

Silicon Valley of Greece

Mutual visits to promote collaboration

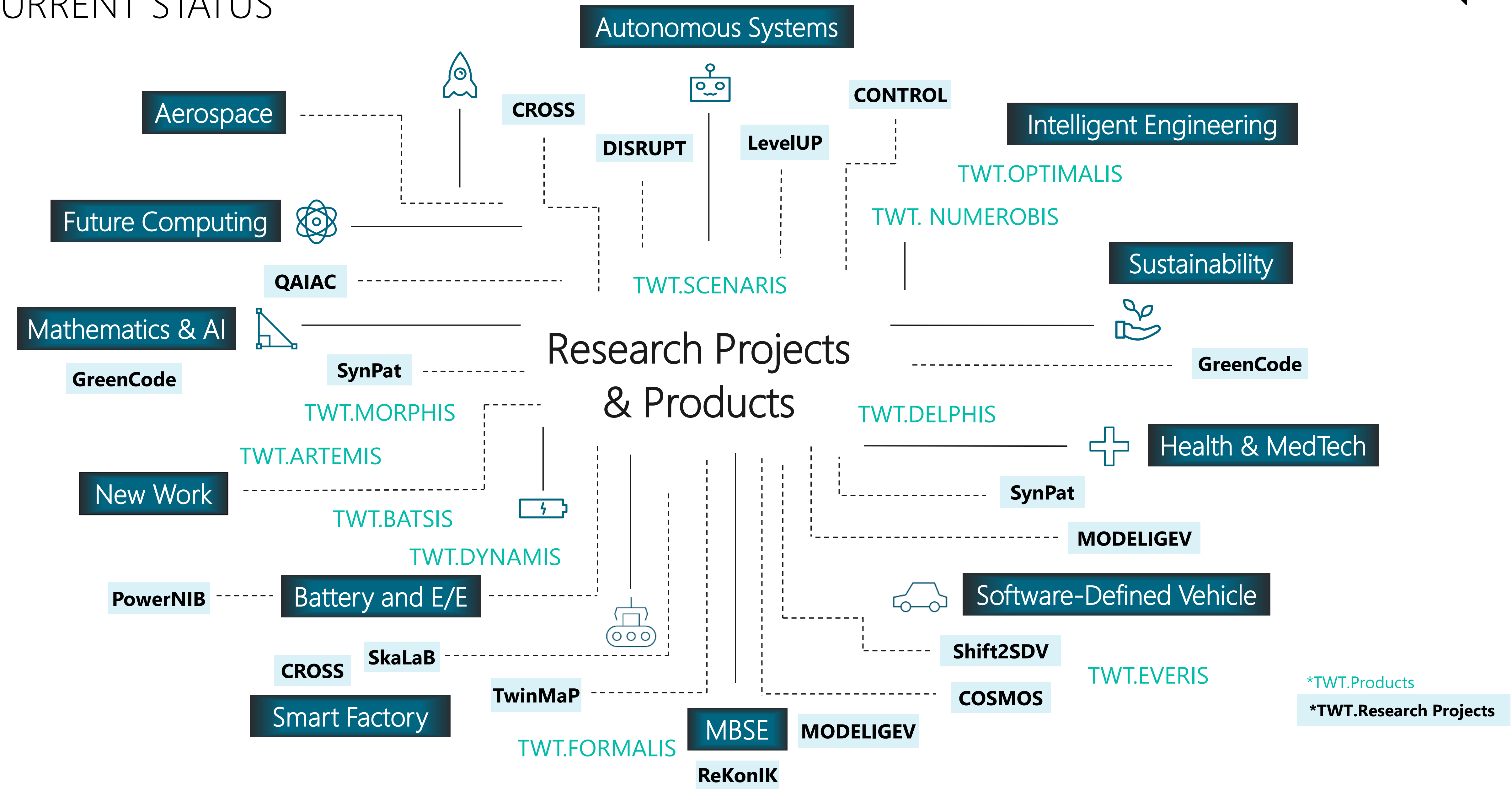
Easy accessibility

CURRENT STATUS

02

TWT

CURRENT STATUS



GREENCODE RESEARCH PROJECT

03

TWT @ GREENCODE RESEARCH PROJECT

Goals TWT

- New methods for green coding in industrial applications and R&D
- Exploring or setting energy standards
- Combining green coding with AI / SLM / LLM

GREEN CODE

Partners

Beia CONSULT INTERNATIONAL
SIEMENS
DigitalTactics
Fraunhofer IESE
FH AACHEN UNIVERSITY OF APPLIED SCIENCES
ZAL
Universität Augsburg University
Trier University of Applied Sciences | HOCHSCHULE TRIER

a.o.

ITEA4

Bundesministerium für Bildung und Forschung

Image Source: ITEA4, BMBF, GreenCode

GREEN ICT: OVERVIEW

Green IT is a broader concept encompassing various disciplines aiming to reduce the environmental impact of technology, including energy-efficient hardware, sustainable data centers, and responsible e-waste management.

Green coding refers to the practice of writing code in a way that minimizes its environmental impact. It is the impact of the code's modeled scenarios, not the code itself, that determines its ecological footprint.

GreenAI in our coding context refers to using the most appropriate small -LLM to develop code with the least environmental impact.

Image Source: https://www.ratbacher.de/blog/green-it-massnahmen/

GREEN IT KEY CONTRIBUTORS **OUR FOCUS**

GREEN CODING

SUSTAINABILITY METRICS AND MEASUREMENT METHODS

LLM INFERENCE COST, GREENER AI

ENVIRONMENTAL IMPACT OF DATA CENTERS

TRENDS OF GREEN ICT IN EUROPE

WHAT IS GREEN CODE IN OUR CONTEXT

TWT

Code Measurement

- ICT growth increases energy use and carbon emissions.
- Study reviewed 181 sources (2010–2025).
- Energy measured at all levels: software → cloud → data center.
- Main optimization methods:
  - Power management
  - Resource optimization
  - VM/container consolidation
  - Energy-aware software design
- LLMs can help generate more efficient code.
- Key gaps:
  - No standard metrics
  - Inconsistent reporting
  - Limited network energy analysis

Future goal: unified, cross-layer energy optimization.

LLM Inference Cost

LLM Inference

- Profiling GPU Power Modeling
- Token-Level Energy Analysis Carbon Footprint Quantification
- Analytical vs. Measurement-Based Estimation

Data Centers

– AI & Energy Use

AI consumes large amounts of electricity.
Data centers increase energy demand.

– Clean AI Claims

Companies promote renewable-powered AI.
Renewable credits ≠ actual clean energy use.

– Reality vs Reporting

Energy is measured yearly.
Location and timing are often ignored.

– Location & Timing Issues

Renewable energy may not match data center needs.
AI runs continuously, even during high-emission periods.

– Transparency Problem

Limited data on AI energy impact.
Sustainability claims are difficult to verify.

– Solution

Track energy by:
Where
When
Emissions

– Conclusion

AI is not automatically clean.
True sustainability needs transparency.

Future of Green ICT in Europe

– European Technology Stakeholders

– Pan-European roadmap study for sustainable digital infrastructure — repeat of Dutch study, now scaled across EU

– Four dimensions captured: solutions, adoption factors, impediments, open problems

– Three-phase methodology: individual interviews → focus group validation → questionnaire confirmation

– Multi-stakeholder perspective: industry practitioners, consultants, and researchers across software, cloud, and hardware domains

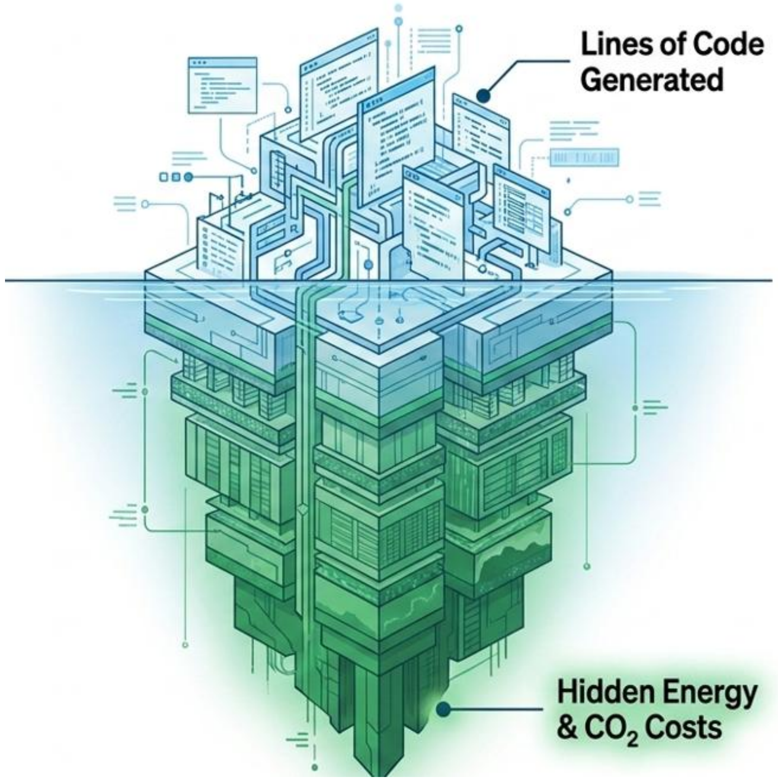
– Iterative co-creation: preliminary findings validated and refined by expert focus groups before finalisation

MOTIVATION FOR GREENCODE

MEASURING CODE'S ENVIRONMENTAL IMPACT

MEASURING VIBE CODES ENVIRONMENTAL IMPACT

UNDERSTANDING DATA CENTERS ENVIRONMENTAL IMPACT



EXISTING TOOLS OVERVIEW

TWT

Tool matrix (1-10)

Tools matrix

Power/energy consumption, efficiency and environmental-impact coverage by tool

Tool / metric	Power / energy	Efficiency	Environmental impact	Hardware / scope	Software / flow
ENRICO	Yes	Yes	Yes	CPU, GPU, full-node via wattmeter and RAPL	CI/CD GitLab integration
CodeCarbon	Yes	Yes	Yes	CPU/GPU/RAM/Cloud	Python Package
Green Cost Explorer	Yes	Yes	Yes	AWS Regions	Cost + Carbon Overlay
Kube Green	Yes	Yes	Yes	Kubernetes Nodes	Kubernetes Add-on
JoularJX	Yes	Yes	Yes	Java Applications	Java Agent
Kepler	Yes	Yes	Yes	Kubernetes Nodes	eBPF + Prometheus
Website Carbon Calculator	Yes	No	Raises awareness	Web Hosting Servers	Front-end/Website Analyzer
Customer Carbon Footprint Tool	Yes	Yes	Yes	AWS	Cloud Usage Analyzer
Emissions Impact Dashboard	Yes	Yes	Yes	Azure/Microsoft Cloud	Dashboard Integration
Cloud Carbon Footprint	Yes	Yes	Yes	AWS, GCP, Azure	Open-source CLI & Dashboard

1/3

Tool matrix (11-20)

Tools matrix

Power/energy consumption, efficiency and environmental-impact coverage by tool

Tool / metric	Power / energy	Efficiency	Environmental impact	Hardware / scope	Software / flow
Green Metrics Tool	Yes	Yes	Yes	CPU/Software Systems	Software Lifecycle Analyzer
Carbon Footprint (reporting tool)	Yes	Yes	Yes	Cloud Providers	Reporting Dashboard
EcoCode	No	Yes	Yes	Android Devices	Static Analyzer for Android
Ec0lint	No	Yes	Yes	Web Browsers	Static Code Analysis
Fruggr	Yes	Yes	Yes	Full IT Stack	Automated Evaluation
Power Distribution Unit (PDU)	Yes	No	No	Anything with external power supply	—
Watts Up Pro	Yes	No	No	Anything with external power supply	—
IPMI	Yes	No	No	Server board	—
Intel RAPL	Yes	No	No	CPU and Memory energy consumption	—
nvidia-smi	Yes	No	No	Nvidia GPU	—

2/3

Tool matrix (21-30)

Tools matrix

Power/energy consumption, efficiency and environmental-impact coverage by tool

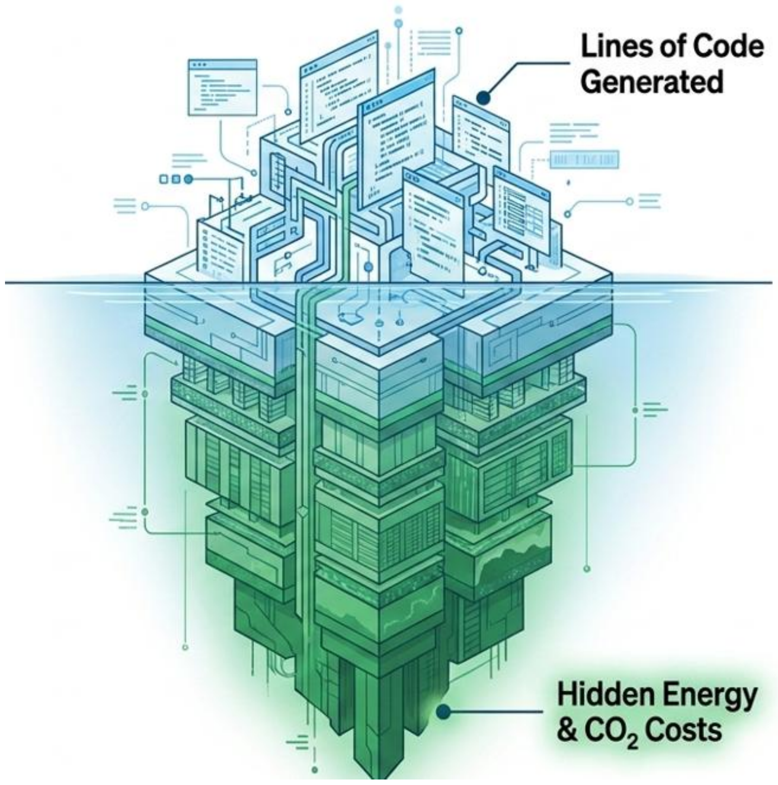
Tool / metric	Power / energy	Efficiency	Environmental impact	Hardware / scope	Software / flow
PowerTop	Yes	No	No	General-purpose	—
Joulemeter	Yes	No	No	General-purpose	—
vMeter	Yes	—	—	General-purpose	—
GreenDroid	Yes	Yes	Yes	Android Devices	Android Test Suite
GreenHub	Yes	No	Yes	Android Devices	Community Power Profiling
Carbon Tracker	Yes	Yes	Yes	GPU/CPU	DL Training Monitor
GSF Green Maturity Matrix	No	Yes	Yes	Organizational Tools	Self-Assessment Framework
GCPM (GreenCode-PowerMeter)	Yes	Yes	No	Any CPU (model-based, host-independent)	VS Code extension
PowerNode	Yes	No	No	Servers, Data Store, Switches (Cloud)	GreenWave Gateway
eco-ci-energy-estimation	Yes	Yes	Yes	CPU	Integrated into CI pipeline

3/3

GAP ANALYSIS

MEASURING CODE'S ENVIRONMENTAL IMPACT

Existing software energy measurement tools primarily focus on monitoring and reporting consumption metrics, offering limited support for energy-aware optimization and validation activities.

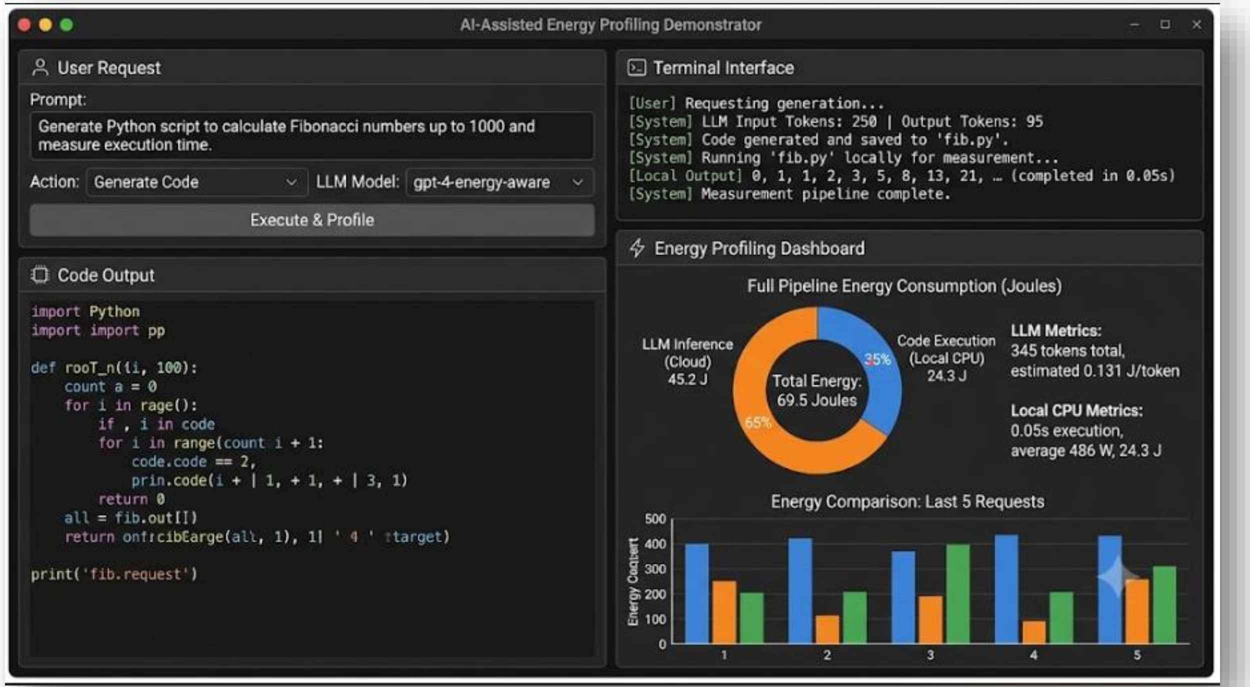


→ The goal is to develop an AI-assisted end-to-end platform that enables developers to measure, validate, and improve softwares energy efficiency within their existing development workflows.

DEMONSTRATOR OVERVIEW

1. Generate AI-assisted code improvements Measure inference cost.

2. Measure software energy consumption



TWT

SOFTWARE LIFE CYCLE ASSESSMENT (LCA)

LCA ANALYSIS OF SOFTWARE DEVELOPMENT PROCESS

LCA (ISO 14040/14044) measures environmental impact across a product's full life cycle. Already standard in automotive, construction, and packaging, it's now being applied to software.

Why are we doing an LCA?

Software's carbon footprint isn't obvious. It lives in the servers it runs on, the laptops it's built with, and the AI tools queried during development. Most teams have no visibility into this.

Our goal: produce a single GWP-100 score (kg CO_2-eq) for a real software product, combining electricity, hardware, cloud services, and waste into one comparable number.

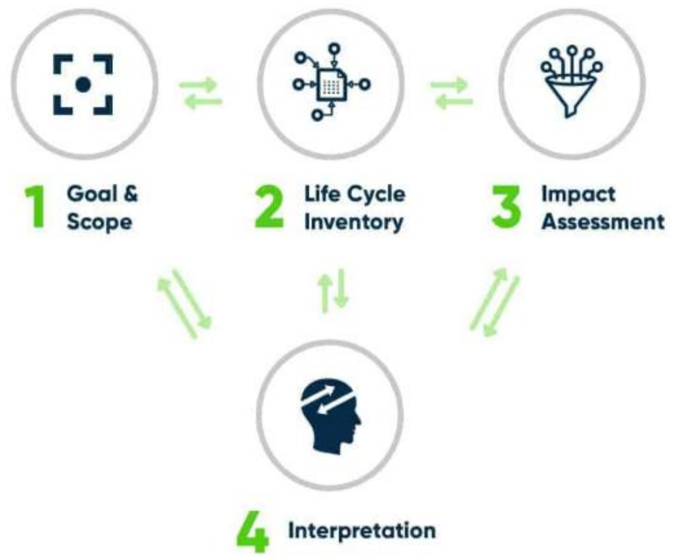
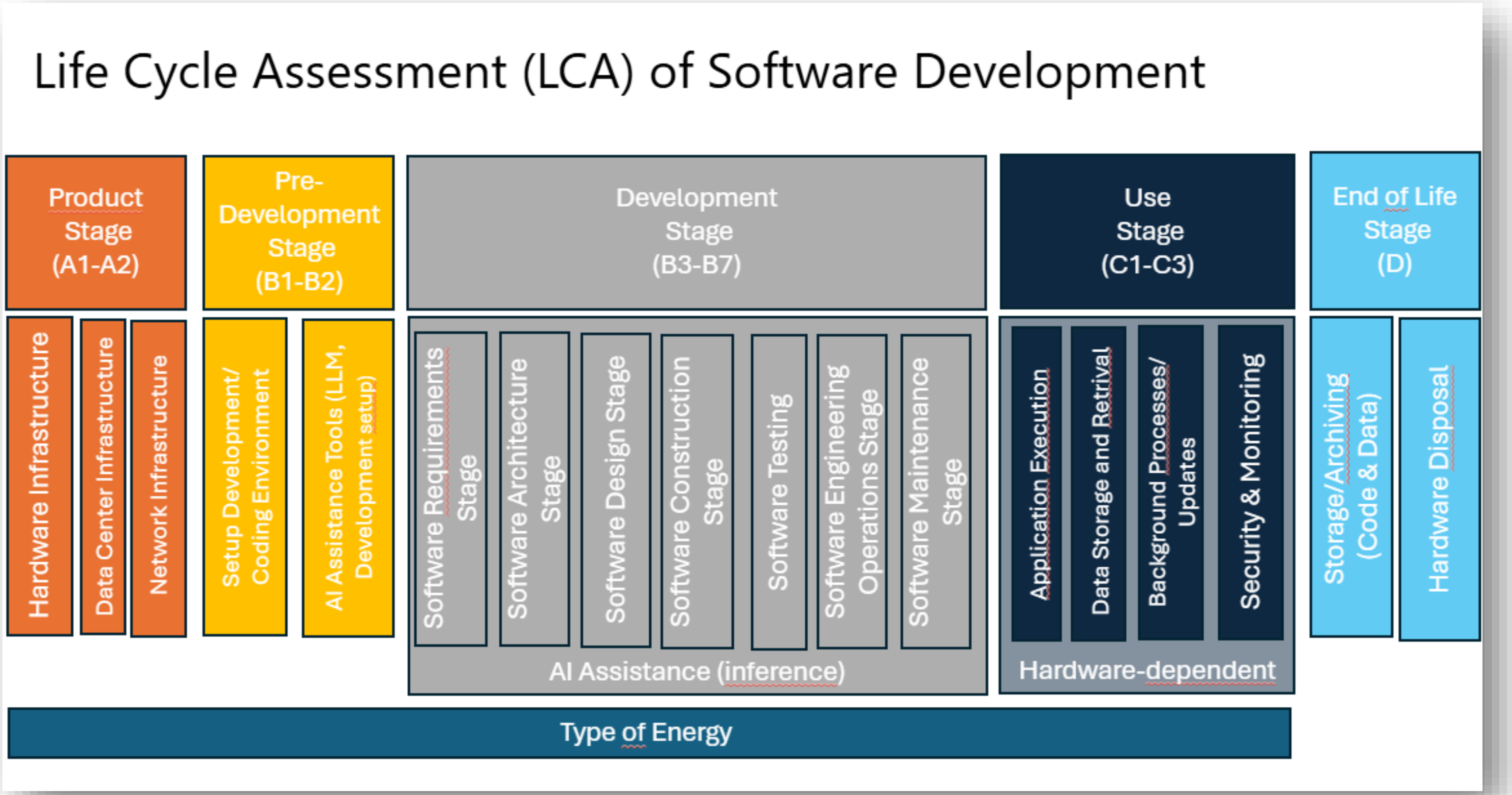


Image Source:

WHOLE PICTURE OF SOFTWARE DEVELOPMENT



TWT

LCA CALCULATION PIPELINE

software_lci.csv

↓

Step 1
brightway_import.py

Reads the CSV and builds two databases in Brightway2:

- my_biosphere catalogs all environmental flows (electricity, CO_2, waste)
- software_lci creates one process activity per life cycle stage, each linked to its flows with the corresponding amounts

→

Step 2
brightway_add_method.py

Registers the GWP-100 characterization method.
Maps each flow to its GWP_factor, which tells the model how to convert raw flow amounts — kilowatt-hours of electricity, kilograms of waste — into a single unit: kg CO_2-equivalent.

→

Step 3
brightway_product_system.py

Assembles MySoftware v1.0 as the functional unit, linking all life cycle stages as inputs.
Triggers the matrix LCA calculation: Brightway2 builds the technosphere and biosphere matrices, multiplies by the characterization vector, and sums to produce the final score.
Also re-runs the calculation per stage to rank contributions.

→ output

Image Source:

EXAMPLE OUTPUT

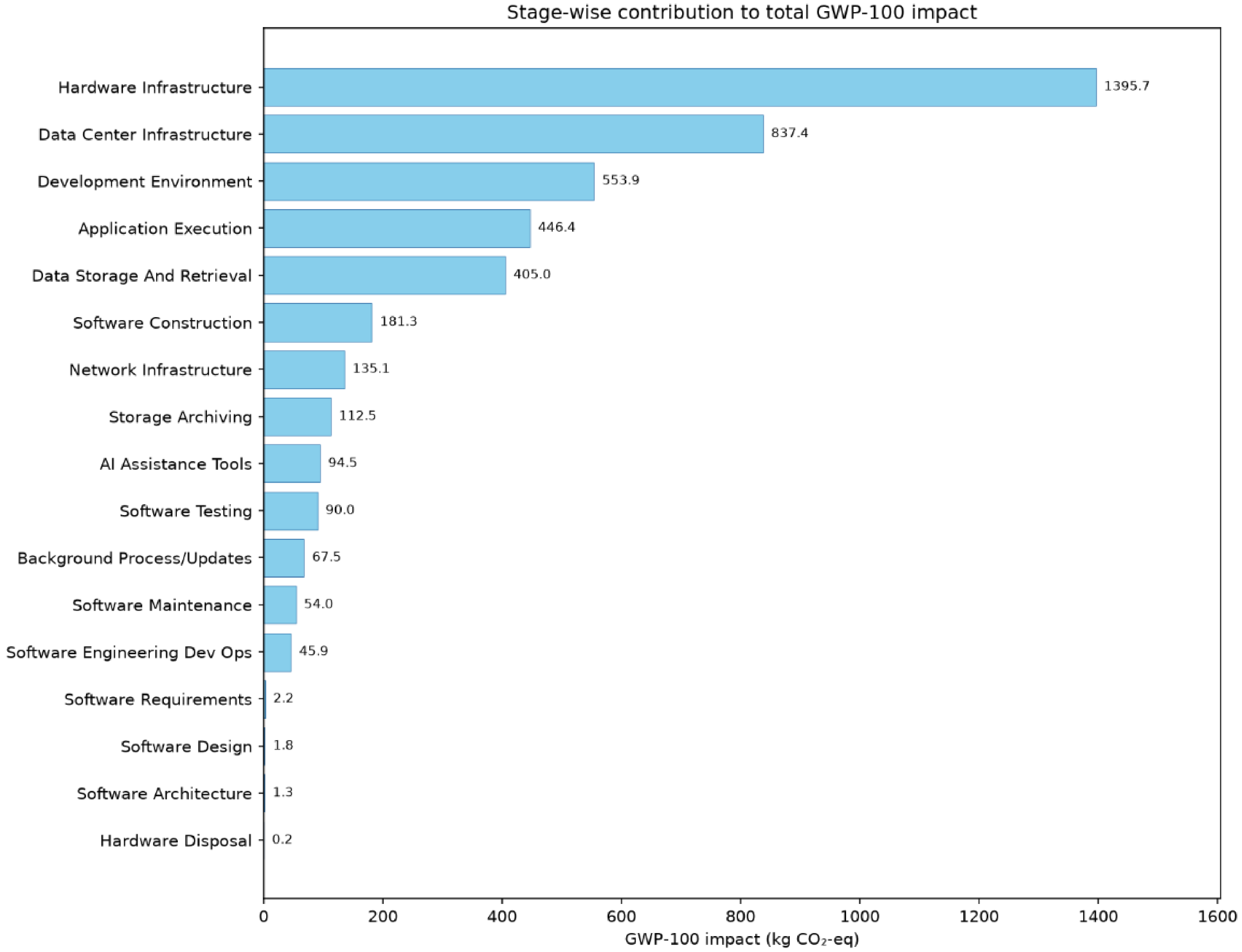


Image Source:

LCA PIPELINE

TWT

1. Define the software-development life-cycle (stages, sub-stages, processes).
2. Assemble a custom LCI by harvesting, cleaning and merging data from existing LCI sources to match the defined stages.
3. Explore & study LCA tools (openLCA, Brightway2, lca-calc) to build a Python-driven workflow that
    1. parses the LCI,
    2. runs impact calculations
    3. aggregates results by stage/sub-stage.

    We selected to mainly use Brightway as it is an open source library that can be easily integrated to the Python workflow.
4. Create visual outputs with a dedicated Python visualisation script.

Tools in the market

LCA=CALC SMART ECO-LIGHTING METRICS

Brightway

openLCA

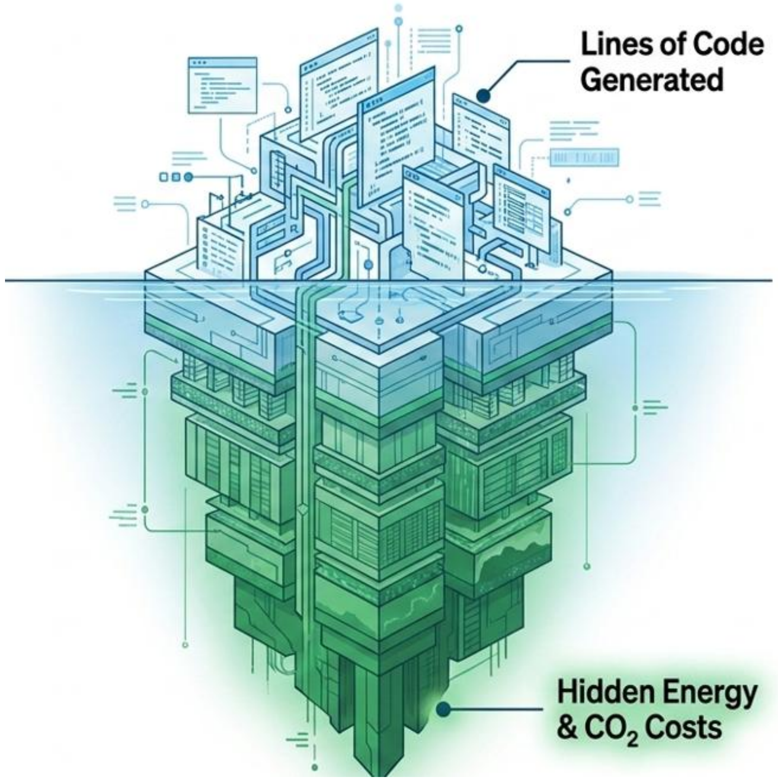
```python
import brightway2 as bw
import pandas as pd
import csv
import os
import sys
import traceback

# ------------------------------------------------------------------
# CONFIGURATION
# ------------------------------------------------------------------
CSV_FILE = "software_lci.csv"
PROJECT_NAME = "software_lci_project"
DB_NAME = "software_lci"

# ------------------------------------------------------------------
# 0️⃣  Show where we are running from
# ------------------------------------------------------------------
print("🔎 Current working directory:", os.getcwd())

# ------------------------------------------------------------------
# 1️⃣  Verify the CSV exists
# ------------------------------------------------------------------
if not os.path.isfile(CSV_FILE):
    sys.exit(f"❌ Error: '{CSV_FILE}' not found in the current folder ({os.getcwd()}).")

# ------------------------------------------------------------------
# 2️⃣  Load the CSV (handles quoted fields automatically)
# ------------------------------------------------------------------
try:
    df = pd.read_csv(
        CSV_FILE,
        sep=",",
        engine="python",
        encoding="utf-8-sig",
        quoting=csv.QUOTE_MINIMAL,
        quotechar='"',
        header=0
    )
except Exception as e:
    sys.exit(f"❌ Failed to read CSV: {e}")

# Clean column names (remove stray spaces)
df.columns = df.columns.str.strip()

# ------------------------------------------------------------------
```

MOTIVATION FOR GREENCODE

MEASURING CODE'S ENVIRONMENTAL IMPACT

MEASURING VIBE CODES ENVIRONMENTAL IMPACT

UNDERSTANDING DATA CENTERS ENVIRONMENTAL IMPACT



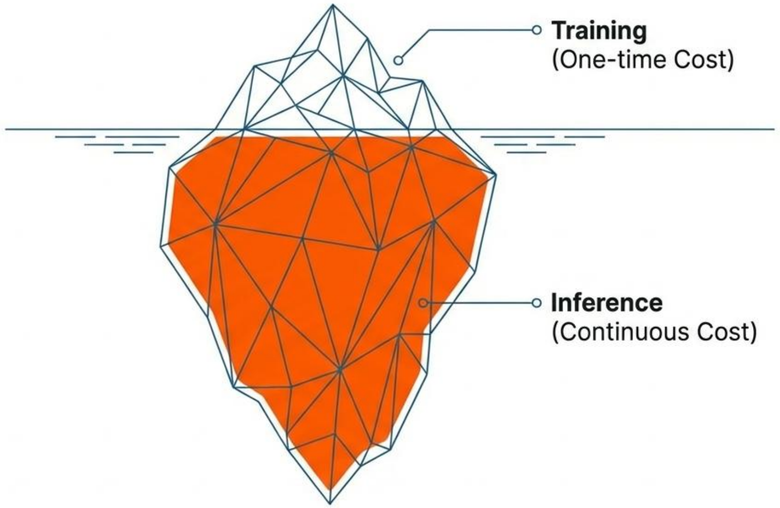
MOTIVATION AND STRATEGY FOR GREENCODE

TWT

INFERENCE, NOT TRAINING, IS THE LONG-TERM ENERGY (COST) DRIVER

While training grabs headlines, deployment (inference) dominates the total lifecycle energy.

As AI integrates into daily workflows (coding agents, chatbots), the cumulative cost of billions of requests skyrockets.



TWT

GREENCODE GENERATION AND INFERENCE COST ANALYSIS ON LOCAL SERVERS

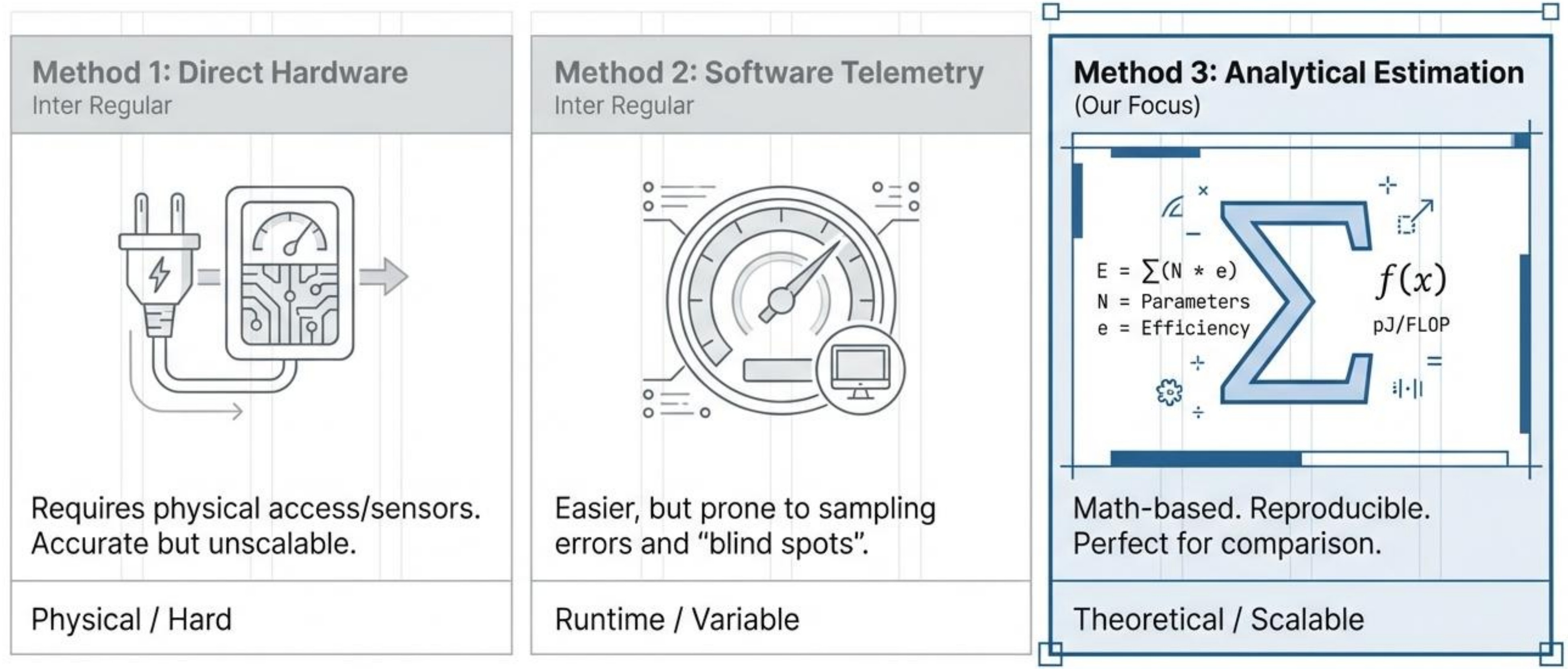
LLM/SLM EVALUATION

- CO2 aware integration of LLMs into vibe code pipelines
- Inference Cost
- On premise
- Multiple models
- Model selection
- Hardware aware analytical estimation

LLM/SLM INTEGRATION

- Multi-Agentic System
- Efficient and robust integration of LLMs into vibe code pipelines
- Code Evaluation
- Structured communication
- Review code output
- Test validation

LLM-SLM EVALUATION

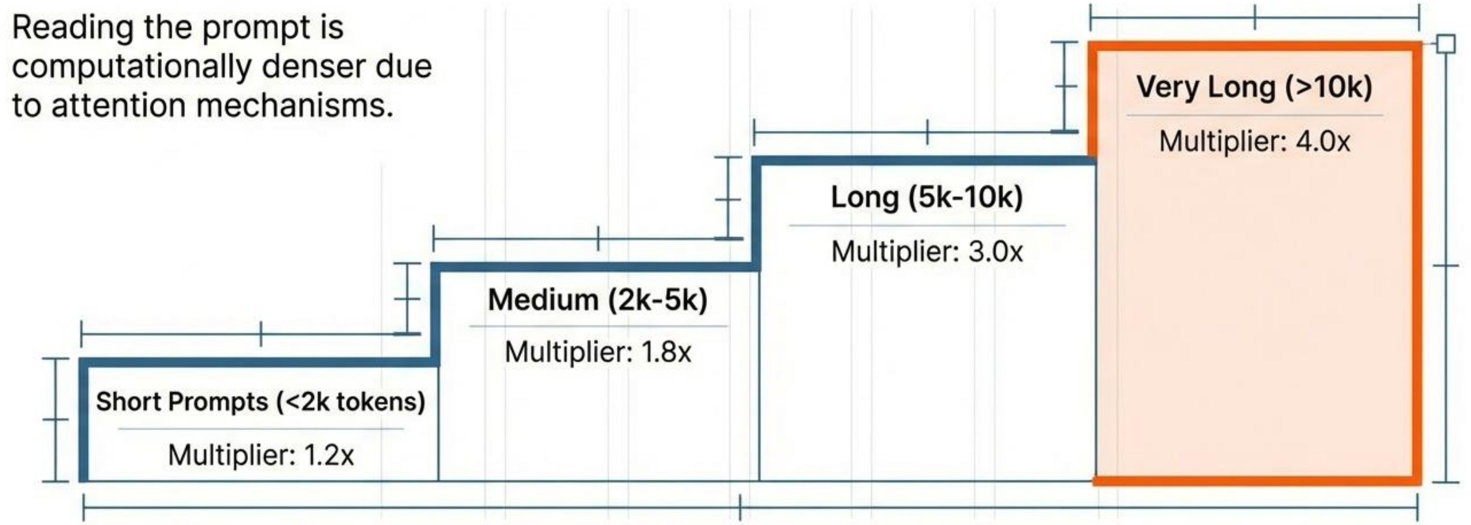


Why we can't always just "read the meter."

LLM-SLM EVALUATION

Parameters in our methodology

- Size of the prompt
- Type of the hardware (GPU)
- Input and Output token

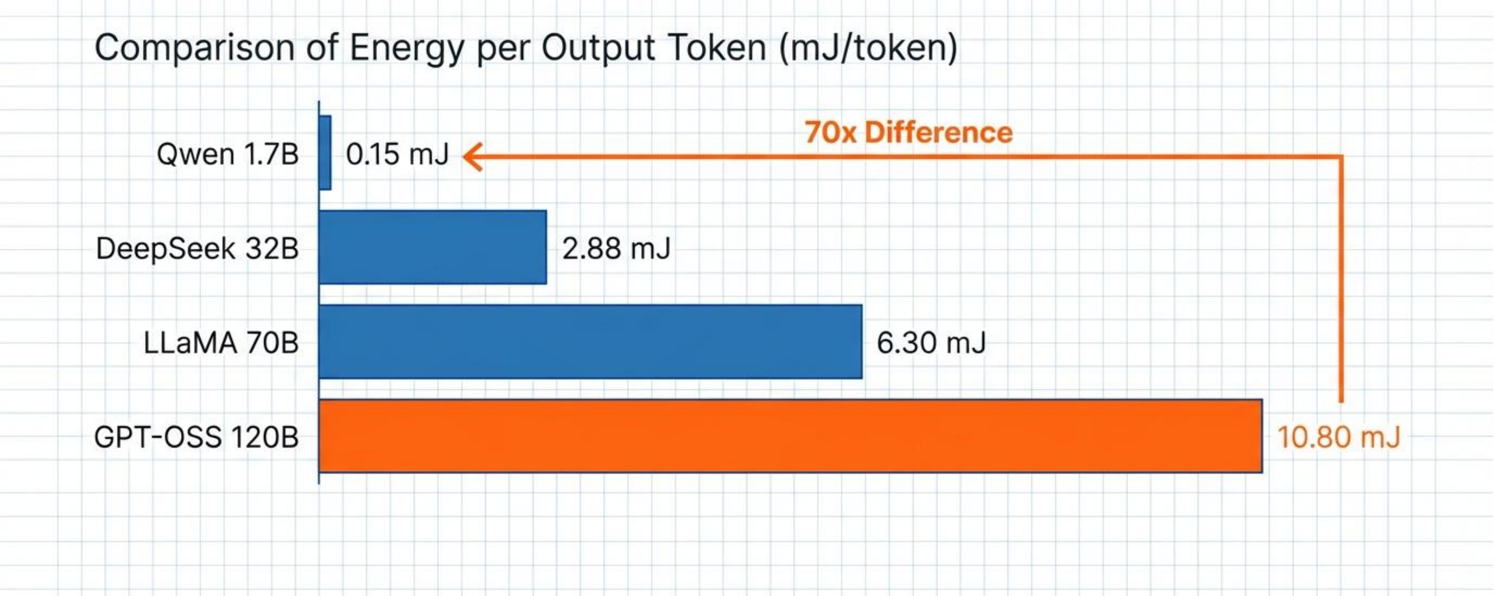


FLOPs(prefill) = Multiplier(L) × Decode Cost

LLM-SLM EVALUATION

Findings

- Energy scales with model size
- Identified values for Energy per output tokens and per input tokens for 15 models.
- Model size varied from 0.308 to 120Billion parameters.



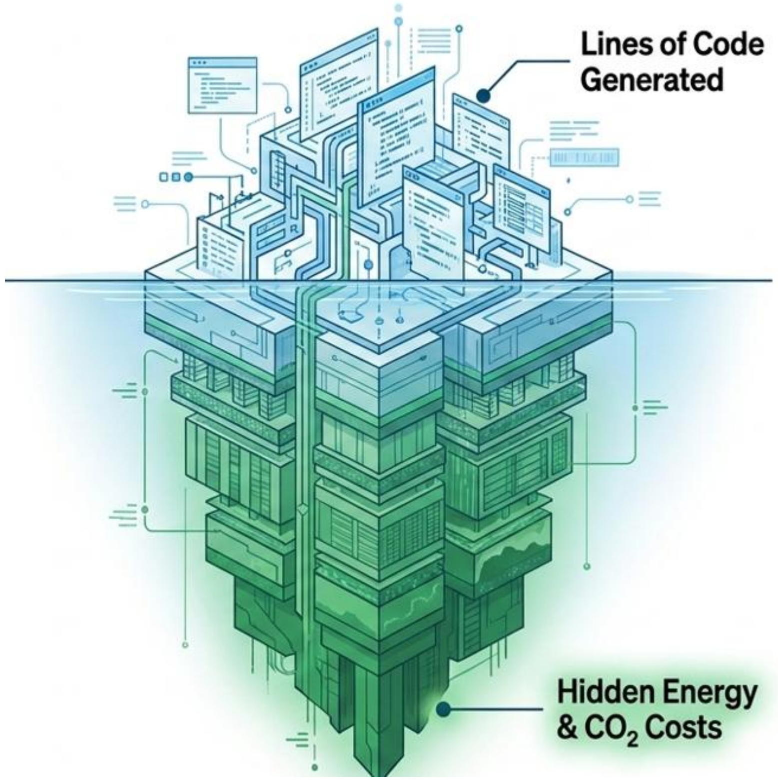
MOTIVATION FOR GREENCODE

TWT

MEASURING CODE'S ENVIRONMENTAL IMPACT

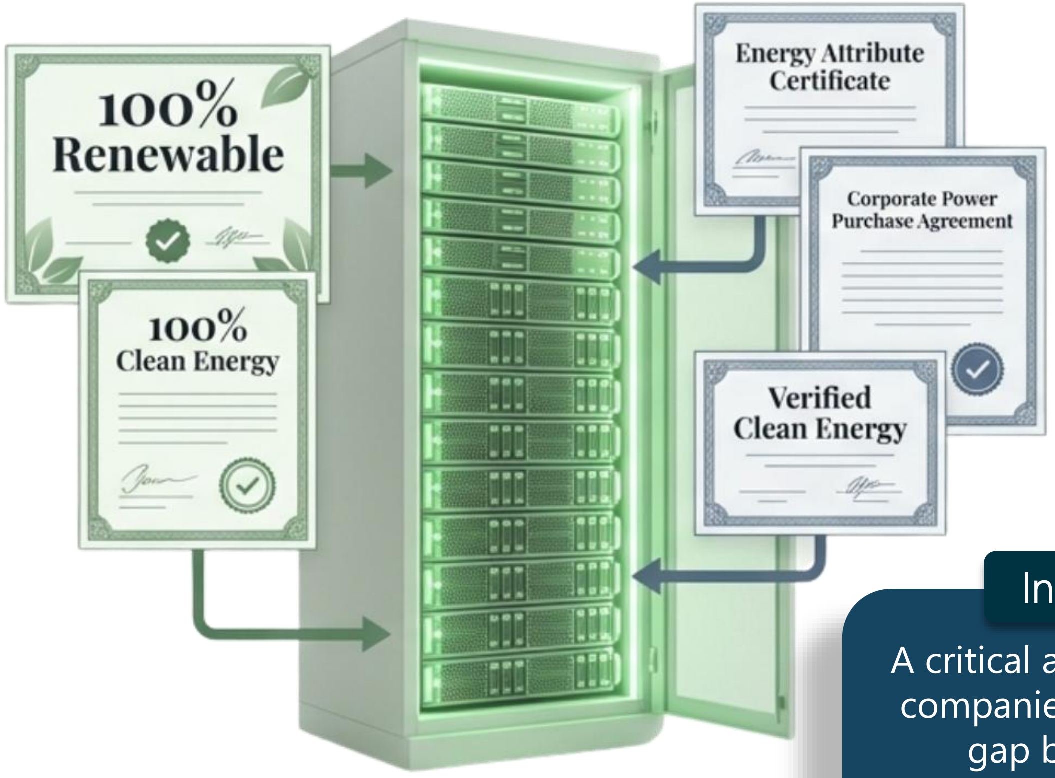
MEASURING VIBE CODES ENVIRONMENTAL IMPACT

UNDERSTANDING DATA CENTERS ENVIRONMENTAL IMPACT



THE ILLUSION OF CLEAN AI

TWT



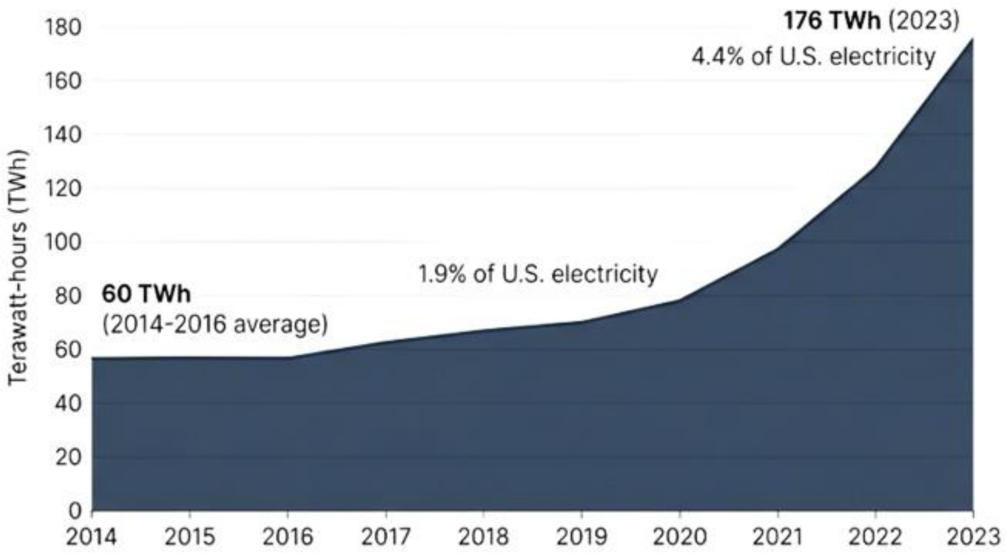
Insight Callout

A critical audit of the Big 5 tech companies reveals a structural gap between financial sustainability accounting and physical grid physics.

Image Source: NotebookLM

TWT

AI WORKLOADS TRANSFORMING THE POWER GRID



- **Spatially Concentrated:** Large-scale model training is executed on massive, geographically dense GPU clusters.
- **Temporally Inflexible:** Inference services and training runs require continuous, sustained high-load availability 24/7.

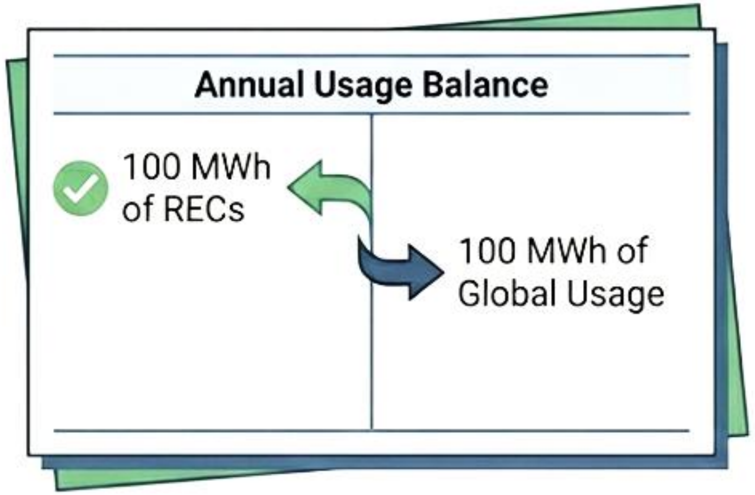
Because AI cannot easily shift its load across hours or regions without severe trade-offs, its true climate impact depends entirely on where and when the computation occurs.

Image Source: NotebookLM

CORPORATE SUSTAINABILITY NARRATIVES

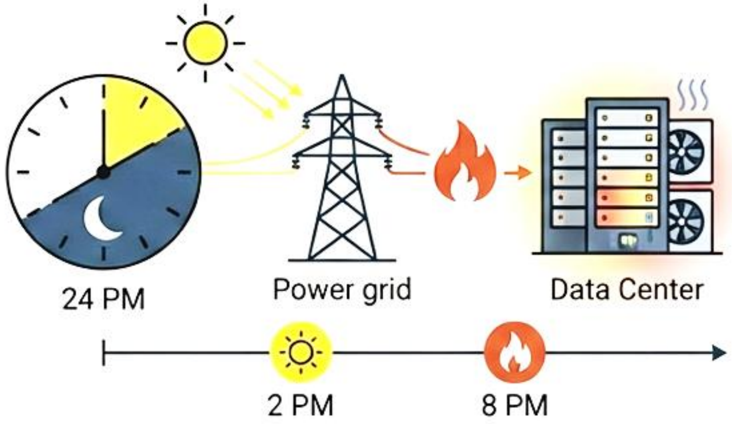
TWT

The Procurement Narrative



Annual Financial Matching

The Operational Reality



Hourly Physical Dispatch

Annual balancing does not require temporal or spatial coincidence. A firm can claim ‘100% renewable’ electricity while relying heavily on fossil-based generation during specific hours or in specific regions.

Image Source: NotebookLM

CONTACT PERSONS

TWT

Michael Keckeisen
michael.keckeisen@twt-gmbh.de

Tina Vartziotis
tina.vartziotis@twt-gmbh.de

Patrick Vaudrevange
patrick.vaudrevange@twt-gmbh.de

WE KNOW HOW.

TWT

WE KNOW HOW.

TWT

METHODOLOGY FOR GREEN CODING: OVERVIEW

- How to assess the software "greenness". We focus on greener auto-generated code for stimulating the energy efficiency awareness.
- Investigate the trend of sustainability metrics how metrics measure resources, and what type of metrics are more appealing.
- Generate greener code using automated code generation tools

STEPS

Investigation of Automated Code Generation for greener code

Sustainability Metrics to measure "greenness" of the code.

Generation of more Sustainable Code

SUSTAINABILITY METRICS IN SOFTWARE DEVELOPMENT AND THEIR IMPORTANCE

TWT

Objectives

Sustainability metrics in software development are becoming an increasingly crucial consideration. With the growing emphasis on environmental impact and efficiency, software developers need to understand the sustainability metrics and how to apply them effectively.

Approach

We perform **literature reviews** and **analysis** on the existing **sustainability metrics** and ways to quantify. We identify sustainability metrics and we **categorize code related metrics** and distinguish from other practices.

Findings

1. **Energy efficiency:** Amount of energy required to execute the code efficiently.
2. **Performance:** Number of service requests served at a given.
3. **Utilization:** Memory, CPU and Storage
4. **Carbon footprint:** (Gross or Net CO2e): Amount of carbon dioxide and other greenhouse gases emitted during the code's lifecycle. (Gross or Net CO2e): Amount of carbon dioxide and other greenhouse gases emitted during the code's lifecycle.
5. **Code quality:** Quality of the code- including its correctness, maintainability, scalability, and modularity.

TWT

SUSTAINABILITY METRICS IN SOFTWARE DEVELOPMENT AND THEIR IMPORTANCE

Energy Efficiency

This metric pertains to the amount of energy required to execute code efficiently. High energy efficiency reduces operational costs and contributes to environmental sustainability by minimizing energy use.

Performance

This metric reflects the number of service requests the software can serve within a given time period. High-performing software allows for more efficient use of resources, which can reduce power usage and strain on hardware resources.

Utilization

This includes measures of memory, CPU, and storage usage. By optimizing the utilization of these resources, we can lower energy consumption, reduce costs, and extend the lifespan of our hardware resources, thus contributing to overall sustainability.

Carbon Footprint

This metric is a measure of the amount of carbon dioxide and other greenhouse gases emitted during the code's lifecycle. Tracking and working to minimize a software's carbon footprint is essential to meet corporate sustainability goals and decrease the environmental impact.

Code Quality

This includes maintainability, scalability, and modularity. High-quality code reduces the need for constant updates and revisions, which in turn decreases the energy and resources spent on these processes. In summary, a commitment to sustainable software development practices is not only environmentally responsible, but also contributes to more efficient and cost-effective operations. As a result, understanding and optimizing these sustainability metrics should be an integral part of our management strategy.

CLASSIFICATION OF SUSTAINABILITY METRICS BASED ON CONTEXT

TWT

How can green metrics be classified? [4]
Based on Context:

- Application is related to software programs or pieces of code.
- Architecture that looks into the energy consumption at the design stage.
- Service that is related to execution and the development of software service, such as Dropbox, Google Workspace, and Salesforce.
- Service Center that focuses on the service execution on a service center. For instance, the energy consumption generated by both active servers – which are executing services – and idle servers that are part of a service center.
- Virtual Machine
- Data Center is related to the impact of data storing and retrieval on a data center.
- Embedded Software is software that interacts directly with the physical world, such as Traffic control systems found in traffic lights.
- Server is related to application, service or data processing on a server machine and the Data Center
- DBMS database management system context describes the energy cost of data storing and retrieving operations.

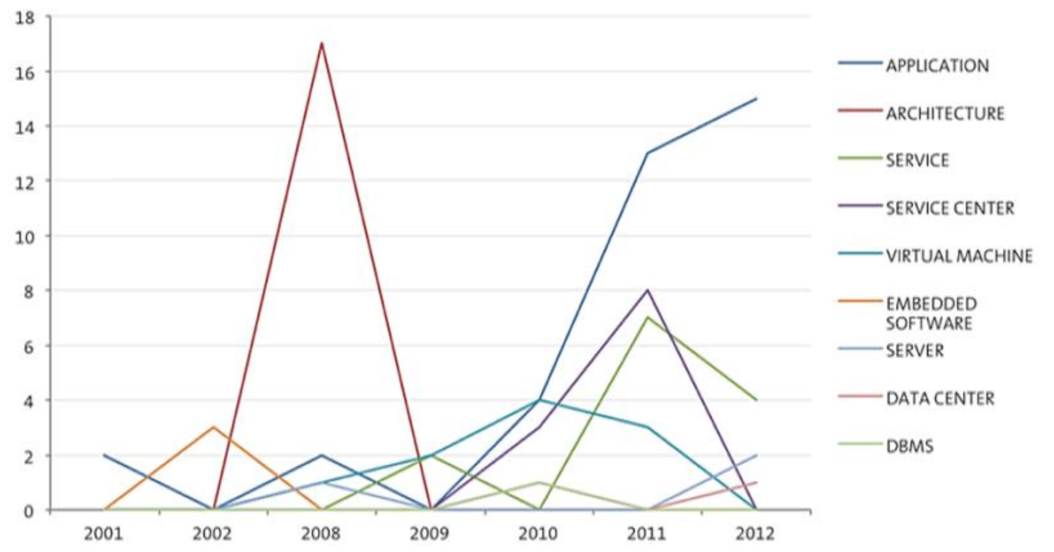


Figure 2: Trend of metrics contexts from 2001 to 2012.

„A SYSTEMATIC LITERATURE REVIEW ON GREEN SOFTWARE METRICS“ [2]

Objectives

1. Describe and classify metrics related to software "greenness" present in the software engineering literature.
2. Evolution of metrics in terms of type and context

Metrics Type	Total	Measurement Unit(s)
Energy	48	Joule (J), Index, Watt (W), Ampere (A) Kilowatt-hour (kWh), Number, byte/kWh
Performance	19	GFLOPS/kWh, Computing Unit/kWh, Percentage (%), Seconds (s), Index, Number
Utilization	17	Percentage (%), Megabyte (MB), Megahertz (MHz), GB/s
Economic	9	Dollars ($)
Performance / Energy	2	GFLOPS/Watt, Index
Pollution	1	CO_2 units

Table 1. Found metrics designed for measuring software energy consumption. Table taken from [2]

2] Lago, P., Gu, Q., & Bozzelli, P. (2014). A systematic literature review of green software metrics. VU Technical Report.

„A SYSTEMATIC LITERATURE REVIEW ON GREEN SOFTWARE METRICS“ [2]

TWT

Findings

1. They verified that research community is focusing on metrics strictly related to energy consumption and saving dimension.
2. Utilization of IT resources is relevant for measuring energy consumption and how much energy may affect system performance.
3. Showed which resources are more measured and how the metrics calculate their energy consumption.
4. Showed which metric types are more attractive for defined contexts.

Which metric contexts have been explored during the last decade? [4]

Metrics Type	Years 2001	2002	2008	2009	2010	2011	2012	Total
Energy	2	3	21	4	8	5	5	48
Performance	0	0	0	0	2	10	7	19
Utilization	0	0	0	0	3	10	4	17
Economic	0	0	0	0	0	4	5	9
Performance / Energy	0	0	0	0	0	1	1	2
Pollution	0	0	0	0	0	1	0	1

Table 2. Number of metrics claimed during the last decade, sorted by type. Table taken from [2]

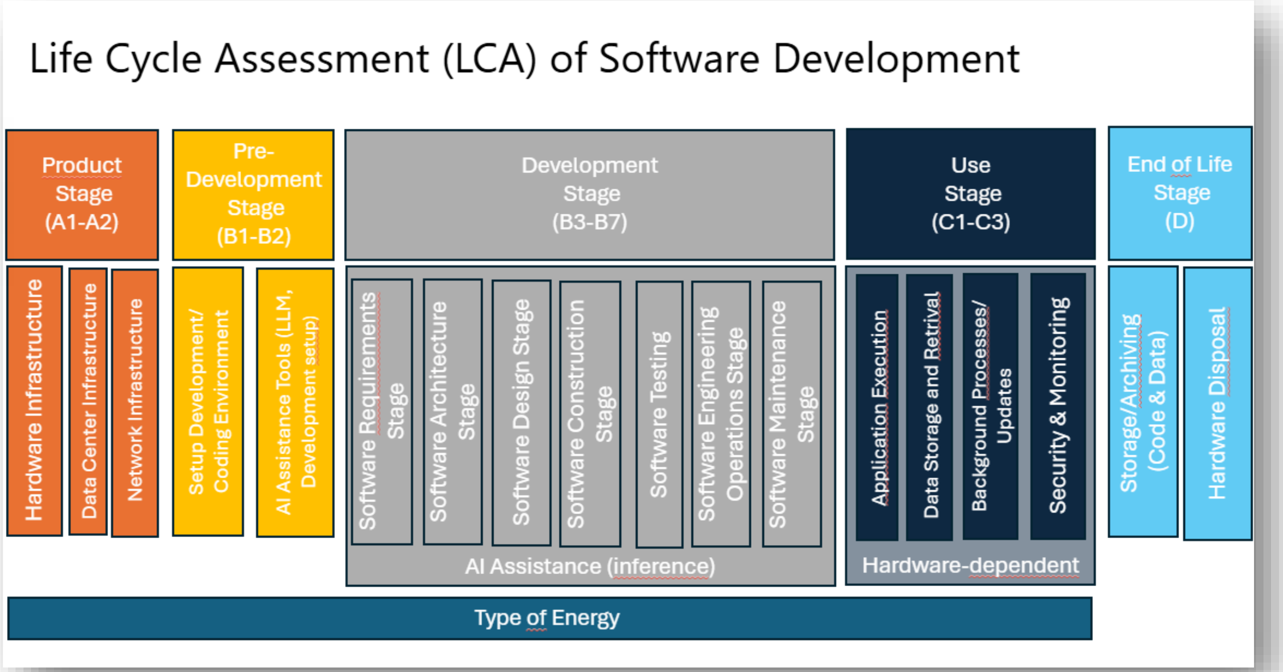
CONCEPTS / ARCHITECTURE / DISSEMINATION / MARKET

TWT

General TWT involvement and goals

Collaboration with Informatics Europe - Green ICT (Prof. Marco Aiello). → In progress

Analysis of lightweight scheduler for multi-GPU LLM inference. → Upcoming



UPCOMING CONFERENCES/PAPERS

TWT

How Expensive Is AI Coding? Measuring Energy and Emissions

Tina Vartzioti[1], Laura Casella[1], Michael Herrnberger[1], Michael Keckeisen[1]

[1]TWT GmbH Science & Innovation, Industriestraße 6, 70565 Stuttgart, DE

Abstract

Large language models (LLMs) are increasingly adopted for code generation in industrial software development, promising productivity gains but introducing new operational costs related to inference compute, energy consumption, and carbon emissions. While prior work has largely focused on the environmental impact of model training, the cumulative footprint of inference—especially for interactive and large-scale code generation—remains poorly quantified. This contribution presents an empirical analysis of the inference cost of LLM-based code generation, focusing on energy use and associated CO_2 emissions. We report measurements across multiple model sizes, prompt structures, and code generation workloads representative of real-world development scenarios. The analysis distinguishes between the prefill and decode phases of inference and examines how factors such as context length, token throughput, and hardware configuration influence power draw and energy per generated token. We further compare estimator-based approaches commonly used in sustainability reporting with direct system-level measurements, highlighting sources of uncertainty and their implications for reproducibility and comparability. Based on these results, we derive practical metrics such as energy per request and emissions per generated code artifact, enabling more meaningful assessments of LLM-assisted development workflows.

Our findings indicate that larger models often provide higher-quality code but at a disproportionately higher inference energy cost, particularly in interactive settings. In contrast, smaller or task-adapted models can achieve more favorable energy–quality trade-offs for specific use cases. The talk concludes with practical guidance for selecting and deploying LLMs for code generation in a cost- and emissions-aware manner, supporting more sustainable use of generative AI in industrial software engineering.

Keywords

Large Language Models (LLMs), Code Generation, GPU Inference, Inference Energy Modeling, Energy Consumption, Carbon Emissions, Green AI

Aachen Automotive Colloquium 2026

Theme: Mobility meets AI

05.10.2026 – 07.10.2026

From Tokens to Watt-hours: Analytical Energy Estimation for LLM Inference on Modern GPUs

Tina Vartzioti[1,2,3,*,†], Rodopi Kosteli[4,†], Elli Danae Vartziotis[3,6], George Dasoulas[2], Michael Keckeisen[3], Konstantinos Skianis[5], Sotirios Kotsopoulos[1,7] and Francesca Dominici[2]

[1]National Technical University of Athens, Patission Complex 42, 10682 Athens, Greece
[2]Harvard University, 1350 Massachusetts Avenue, 02138 Cambridge, MA, USA
[3]TWT GmbH Science & Innovation, Industriestraße 6, 70565 Stuttgart, DE
[4] NIKI Ltd Digital Engineering, 205 National Resistance Street, 45500 Ioannina, Greece
[5] University of Ioannina, Campus, 451 10 Ioannina, Greece
[6]National and Kapodistrian University of Athens, Panepistimiou 30, 106 79 Athens, Greece
[7]Massachusetts Institute of Technology, 77 Massachusetts Avenue, 02139 Cambridge, MA, USA

Abstract

The operational energy consumption of large language model (LLM) inference is becoming an increasingly important component of the environmental footprint of deployed AI systems. However, direct measurement of inference energy often requires hardware telemetry, power instrumentation, or infrastructure-specific monitoring, limiting its applicability in comparative studies, early-stage system design, and sustainability reporting. This report presents an analytically structured, empirically calibrated, GPU-level methodology for estimating LLM inference energy on NVIDIA H100-class accelerators without direct runtime measurement. The proposed estimator combines parameter-scaled transformer FLOP accounting, calibrated memory-traffic factors, and hardware-specific energy coefficients for FP16/BF16 tensor-core computation and high-bandwidth-memory movement. It explicitly separates prompt prefill from autoregressive decoding, enabling energy estimates for input tokens, output tokens, and complete inference requests. The methodology further decomposes total energy into compute, parameter-access, key–value-cache write, and attention-read components, allowing the scaling behavior with model size, context length, and generated-token count to be analyzed. The resulting estimates are not intended to replace physical power measurements; rather, they provide transparent, reproducible, and assumption-explicit approximations suitable for model comparison, green-coding analysis, and design-time evaluation of LLM inference workloads.

Keywords

Large Language Models, GPU Inference, Inference Energy Modeling, Transformer Systems, HBM Memory Traffic, Tensor-Core Computing, Energy-Efficient AI, Green AI, Sustainable Computing, High-Performance Computing

GreenAI Workshop at ECML-PKDD 2026

Conference held in Naples, Italy

07.09.2026 – 11.09.2026

Image Source: TWT

CONTACT PERSONS

Michael Keckeisen

michael.keckeisen@twt-gmbh.de

Tina Vartziotis

tina.vartziotis@twt-gmbh.de

Patrick Vaudrevange

patrick.vaudrevange@twt-gmbh.de

WE KNOW HOW.