Green AI: Evaluating the Environmental Cost of Training Machine Learning Models

Sian O'Briain
School of Enterprise Computing and Digital Transformation, TU Dublin, Ireland
X00218714@myTUDublin.ie

OLLSCOIL TEICNEOLAÍOCHTA BHAILE ÁTHA CLIATH
DUBLIN
TECHNOLOGICAL UNIVERSITY DUBLIN

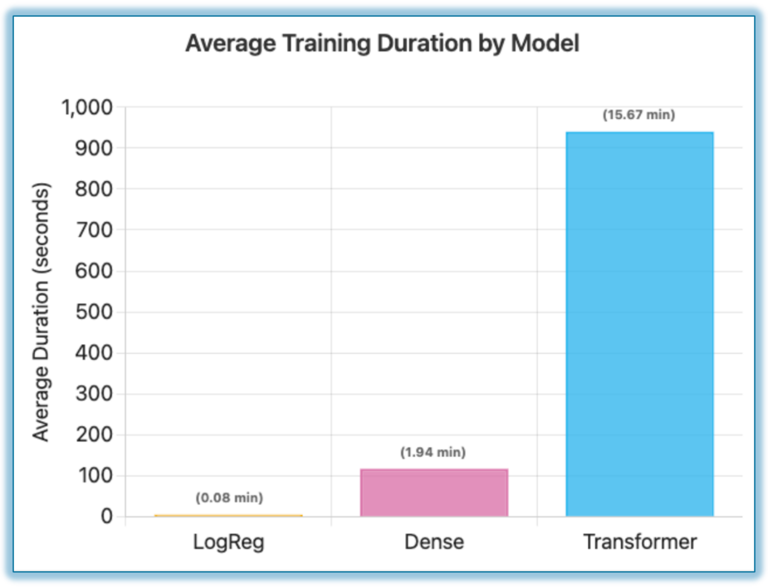
Introduction

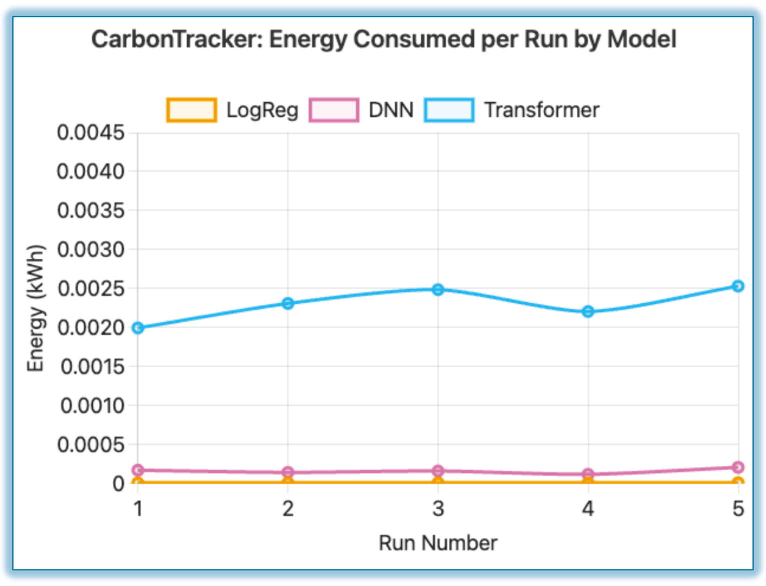
As the use of AI continues to rise so does its corresponding carbon footprint. While much of the current research focuses on monitoring machine learning workloads running in large-scale cloud environments, this study aims to investigate the relationship between model complexity and associated carbon emissions during training in a local controlled environment. Three model types of varying complexity - Logistic Regression, a Dense Neural Network and a Tiny Transformer are trained using the same IMDb film review dataset and on the same binary classification task to determine if a film review has a positive or negative sentiment. Each model's performance is then evaluated in order to further explore the trade-off between predicted performance and associated carbon footprint. Energy consumption and carbon emissions are estimated using two carbon tracking tools, CodeCarbon and CarbonTracker which are also compared as part of the analysis.

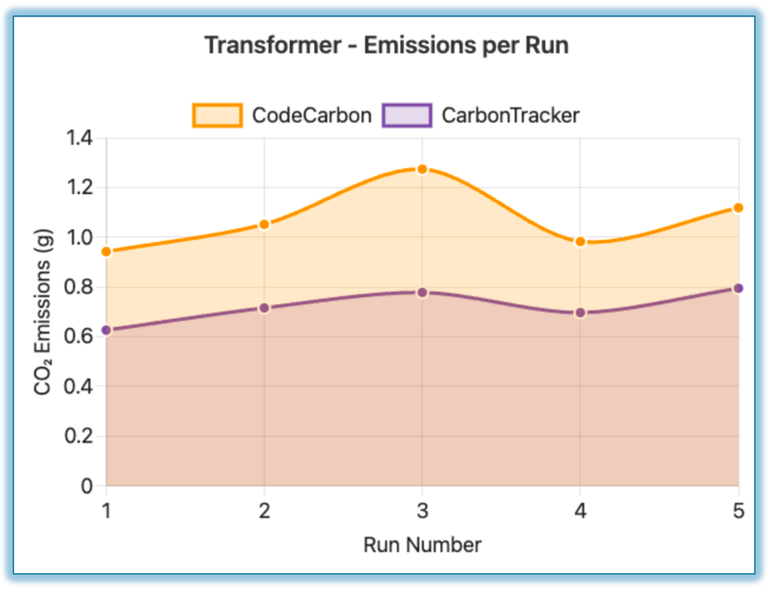
Results

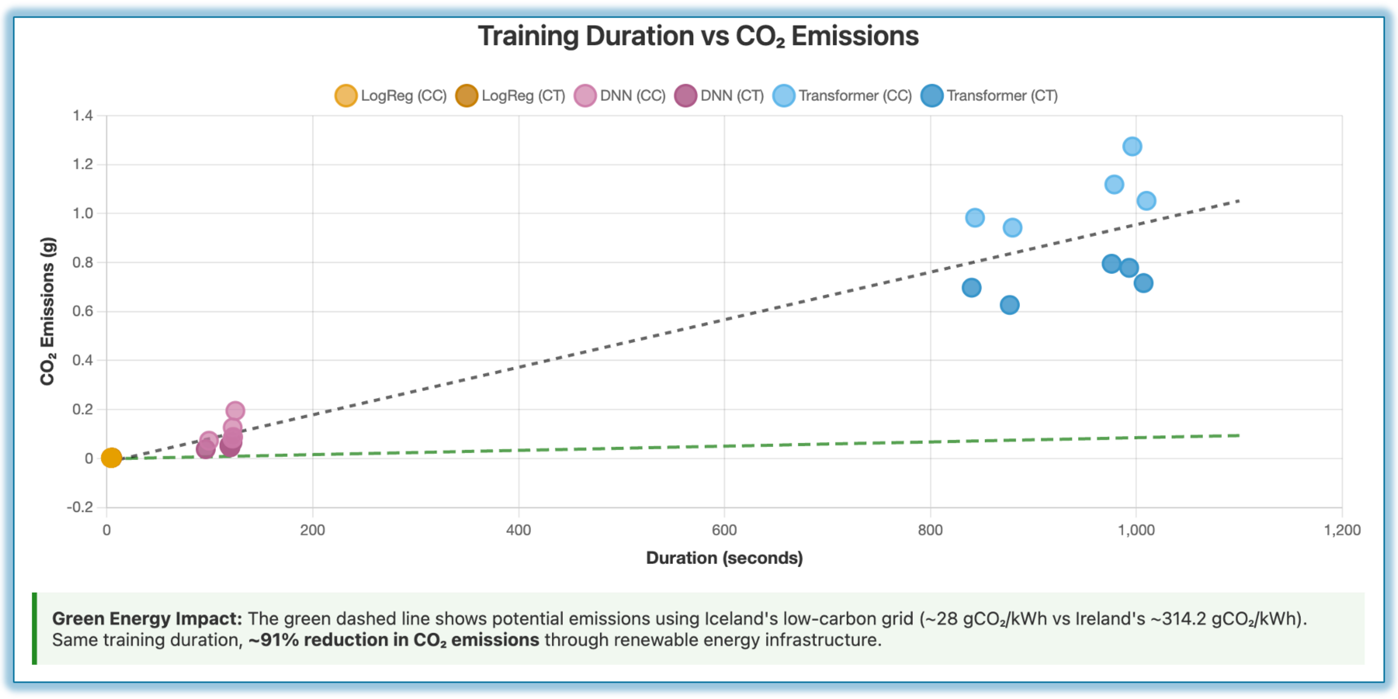
Energy & Emissions

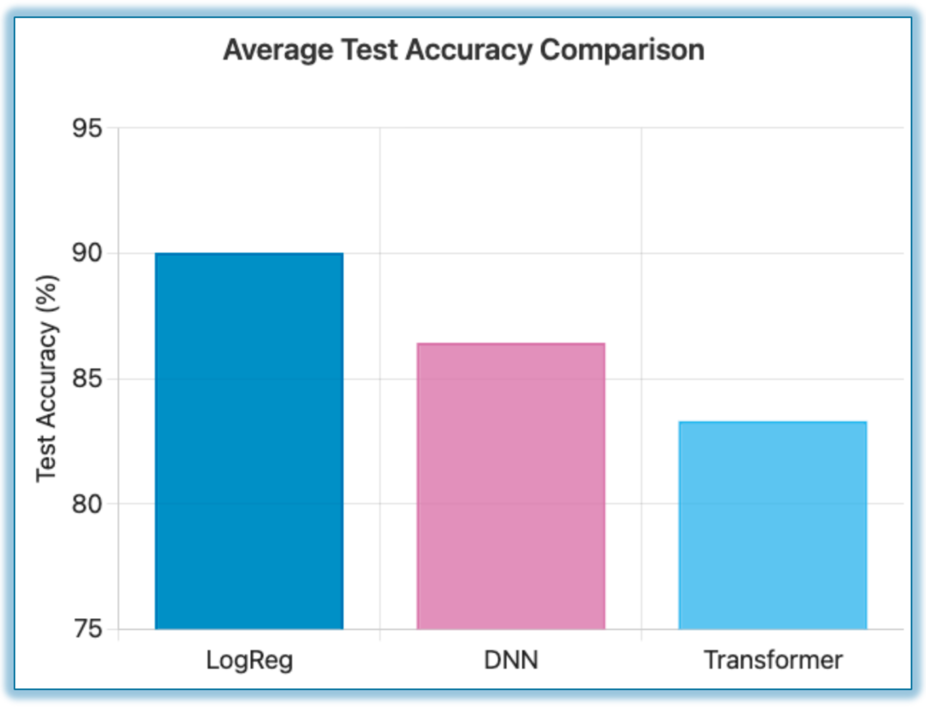
The Transformer model dominated total energy consumption & accounted for >90% overall usage, emitting ≈300× more CO_2e than Logistic Regression, training duration being the primary contributing factor. Discrepancies were observed between carbon tracking tools due to differences in estimation methodologies. Variations in carbon intensity had a significant impact on the resulting emissions.











Performance Vs Energy Cost

Logistic Regression achieved the highest predictive performance with greater than 90% accuracy, while also being ≈300× more carbon-efficient than the Transformer model. For this task, the Transformer performed worse despite its substantially higher environmental footprint. Early stopping and patience avoided unnecessary emissions from continued training and prevented any performance degradation.

Research Questions

RQ1: How does model complexity affect energy consumption and carbon footprint during training?

RQ2: Are the gains in performance from the more complex models worth the trade-offs associated with higher carbon emissions?

Methodology

Task & Data

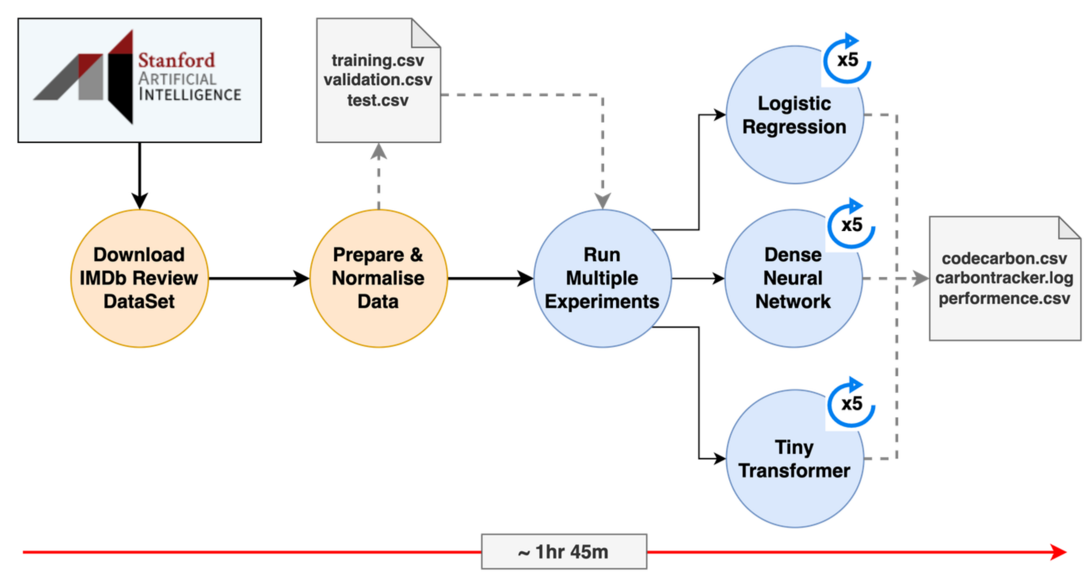
- IMDb binary sentiment classification (positive/negative), 50,000 reviews
- Train/validation/test split: **80/10/10**

Models (increasing complexity)

- Logistic Regression
- Dense Neural Network
- Tiny Transformer

Experiment Setup

- Local controlled environment: MacBook Air (Apple M3, 16GB RAM)
- **5 runs/model**; report mean values
- Track energy & emissions with **CodeCarbon** and **CarbonTracker**



Automated pipeline used to run repeatable experiments.

Conclusions and Future Work

Model choice matters
From a Green AI perspective, bigger is not always better, as larger models incurred significantly higher energy consumption and emissions in this study. These findings highlight that model selection has a direct and substantial impact on environmental cost.

CodeCarbon as the most suitable tracking tool
CodeCarbon was identified as the most suitable tracking tool due to its greater configurability and its consideration of system components beyond CPU and GPU usage. In addition, its structured output formats were well suited for analysis, and its active maintenance and community support contribute to its reliability.

Carbon intensity and geography strongly influence emissions
Identical workloads can result in vastly different emissions depending on the underlying energy mix and carbon intensity of the execution location. As training workloads are geographically flexible, this makes emissions optimisation both practical and feasible.

No single solution is sufficient
Model selection and workload location alone are not enough to address AI's environmental impact. Additional measures are required, including energy-efficient hardware, efficient algorithms, and improved non-AI workload efficiency (e.g. CI/CD), alongside future work on inference-stage emissions, transfer learning, cloud versus IoT assessments, and education.

QR Code for Recording