Farid KHOUDA

**Plan**

**I. Introduction**

A. Contextualization of the carbon footprint problem in Machine Learning (ML)

B. Importance of understanding and reducing the carbon footprint

**II. Theoretical Considerations**

A. Definition of the carbon footprint in ML

B. Distinction between operational emissions and lifecycle emissions

C. Issues related to inaccurate emission estimates

**III. Best Practices for Carbon Footprint Reduction**

A. Presentation of the "4Ms" (Model, Machine, Mechanization, Mapping)

B. Details on each practice and its impact on emission reduction

C. Graphic illustration of improvements over time (Figure 1)

**IV. Case Studies**

A. Case 1: Transformer vs. Evolved Transformer vs. Primer

1. Evolution of models and hardware over time

2. Impact on emission reduction

B. Case 2: GPT-3 vs GLaM

1. Comparison of models in terms of parameters, years of acceleration, energy consumption, and CO2e

2. Impact of data center location on emissions

**V. Global Energy Consumption in ML**

A. Estimates of energy consumption for Google

1. Distribution between training and inference

2. Stability of ML energy share in Google's total consumption

**VI. Additional Factors**

A. Impact of Neural Architecture Search (NAS)

1. Evaluation of the energy cost and utility of NAS

B. Impact of ML model usage on client devices

1. Consideration of the impact on mobile device energy consumption

2. Comparison between server and client consumption

**VII. Future Perspectives**

A. Predictions on the future reduction of the carbon footprint in machine learning

B. Call for widespread adoption of best practices

**VIII. Conclusion**

A. Recapitulation of main conclusions and contributions

B. Importance of transparency on emissions in ML publications

# **I.Introduction**

In recent years, the field of Machine Learning (ML) has experienced unprecedented growth, becoming an integral part of various industries and applications. However, this surge in ML workloads has given rise to a critical concern – the environmental impact associated with the carbon footprint of ML operations. As the demand for ML continues to escalate, there is a growing need to contextualize and address the environmental implications of this technological advancement.

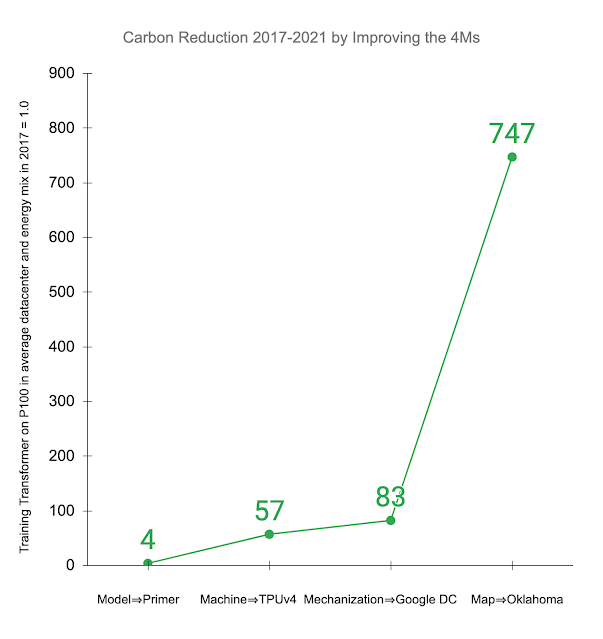


Figure 1 graphic about Google Oklahoma data center

Source : [Good News About the Carbon Footprint of Machine Learning Training](https://blog.research.google/2022/02/good-news-about-carbon-footprint-of.html)

**A. Contextualization of the Carbon Footprint Problem in Machine Learning (ML)**

The rapid expansion of ML workloads has led to an increased focus on the energy consumption and associated carbon emissions of ML training processes. Numerous studies have underscored the environmental repercussions, raising questions about the sustainability of current ML practices. One of the significant challenges is the accurate assessment of the carbon footprint, considering the complexities involved in estimating both operational and lifecycle emissions associated with ML operations.

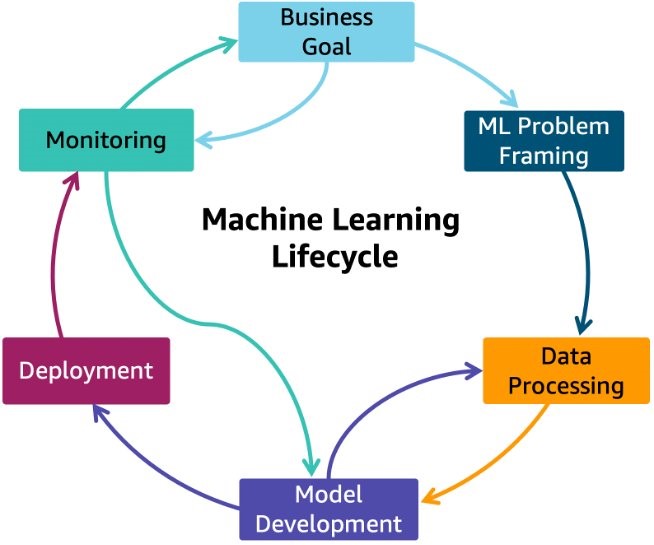


Figure 2. ML lifecycle

Source : [Optimize AI/ML workloads for sustainability: Part 1, identify business goals, validate ML use, and process data](https://aws.amazon.com/fr/blogs/architecture/optimize-ai-ml-workloads-for-sustainability-part-1-identify-business-goals-validate-ml-use-and-process-data/)

**B. Importance of Understanding and Reducing the Carbon Footprint**

In light of the escalating global concerns surrounding climate change and environmental sustainability, the imperative to comprehend and mitigate the carbon footprint of ML operations has never been more crucial. Beyond the ethical responsibility, understanding the environmental impact of ML is essential for fostering responsible technological advancements. As ML practitioners strive to innovate and improve model quality, a parallel commitment to minimizing the environmental footprint becomes imperative. This introduction sets the stage for a comprehensive exploration of best practices and innovative solutions aimed at mitigating the carbon footprint of ML operations.

**II. Theoretical Considerations**

**A.Definition of the Carbon Footprint in ML**

Before delving into the intricacies of the carbon footprint in the context of Machine Learning (ML), it is imperative to establish a clear definition. The carbon footprint in ML encapsulates the total greenhouse gas emissions, particularly carbon dioxide (CO2) and other associated pollutants, attributed to the entire lifecycle of ML operations. This encompasses the energy consumption during model training, inference, and the manufacturing processes of hardware components involved.

**B.Distinction between Operational Emissions and Lifecycle Emissions**

To comprehensively assess the environmental impact of ML, it is crucial to distinguish between operational emissions and lifecycle emissions. Operational emissions encompass the energy costs incurred during the day-to-day operation of ML hardware, including data center overheads. On the other hand, lifecycle emissions extend beyond operational boundaries to incorporate the embedded carbon emitted during the manufacturing of all components involved, from chips to data center buildings. While this study primarily focuses on operational emissions, recognizing the broader lifecycle perspective sets the stage for future comprehensive analyses.

**C. Issues Related to Inaccurate Emission Estimates**

One of the key challenges in addressing the carbon footprint in ML lies in the accurate estimation of emissions. Recent studies have highlighted the prevalence of inaccuracies in emission estimates, leading to potentially misleading extrapolations. Inaccurate assessments can result in unwarranted concerns about the environmental impact of ML workloads. This section will delve into the complexities and nuances associated with carbon accounting in ML, emphasizing the importance of precise and transparent reporting to ensure responsible and informed decision-making.

**III. Best Practices for Carbon Footprint Reduction**

Reducing the carbon footprint in the field of machine learning (ML) is crucial to mitigate the environmental impact of these operations. In this section, we will explore best practices that can significantly decrease energy consumption and CO2 emissions associated with ML work.

**A. Introduction of the "4Ms" (Model, Machine, Mechanization, Map)**

The "4Ms" represent a set of fundamental best practices aimed at optimizing ML processes for a reduced carbon footprint.

Model: Choosing efficient ML model architectures, such as sparse models compared to dense models, can significantly reduce computations, up to a factor of ~5–10.

Machine: Utilizing processors optimized for ML training, such as TPUs or recent GPUs (e.g., V100 or A100), as opposed to general-purpose processors, can improve performance per watt by 2 to 5 times.

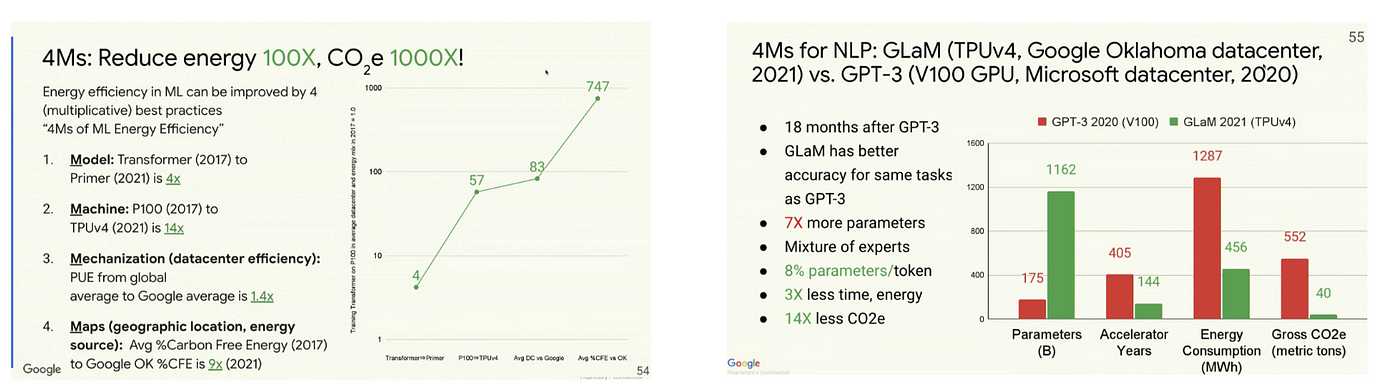


Figure 3 : 4Ms overview

Source : [Stanford ME344S HPC-AI Summer Seminar — Week 1 in Review: Codesigning Computing Systems for Artificial Intelligence](https://romrawin.medium.com/stanford-me344s-hpc-ai-summer-seminar-week-1-in-review-codesigning-computing-systems-for-2e0ddcd9392d)

**B. Details on Each Practice and Its Impact on Emission Reduction**

Model: Choosing more efficient model architectures, such as sparse models, optimizes computations and reduces energy consumption during training.

Machine: The use of specialized processors, like TPUs or recent GPUs, allows a significant improvement in performance per watt, thus reducing energy consumption during training.

Mechanization: Adopting cloud computing has significant benefits in terms of data center energy efficiency, lowering energy costs.

Map: Selecting the data center location based on cleaner energy sources contributes to a substantial reduction in the overall carbon footprint.

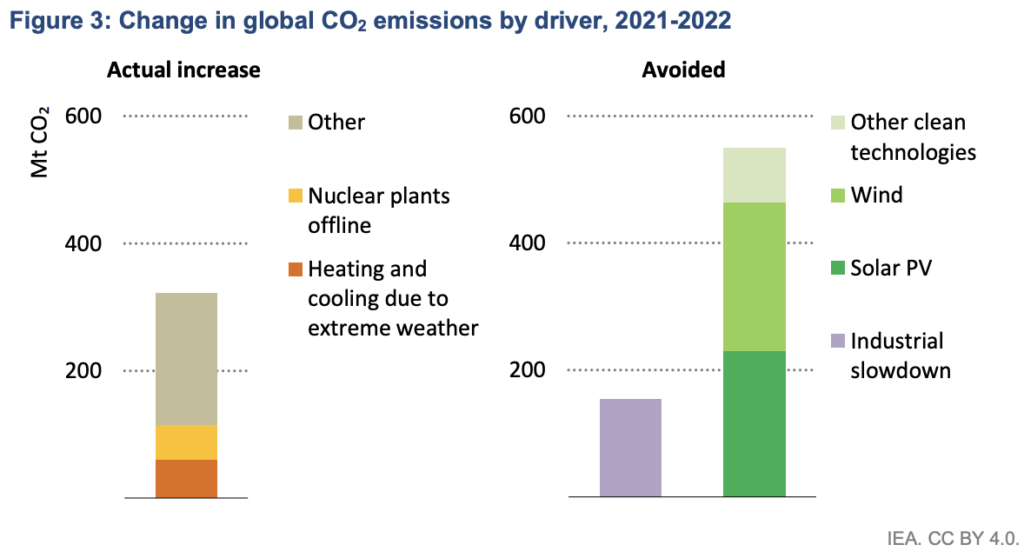


Figure 4 : Charge in global CO2 emissions by driver, 2021-2022

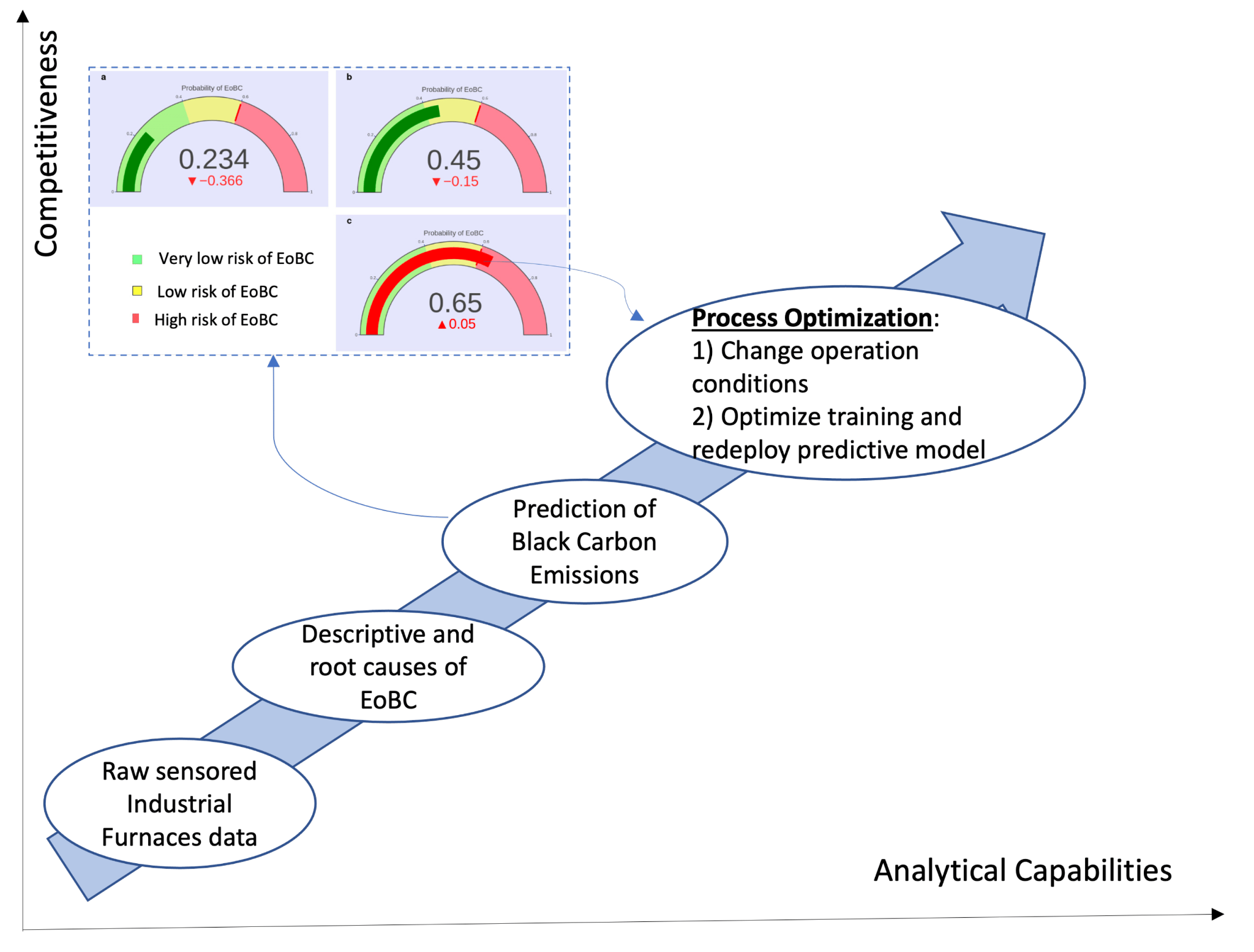
Source : [Reducing Carbon Emissions with AI The Role of Machine Learning in Energy Efficiency](https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.datategy.net%2F2023%2F03%2F07%2Freducing-carbon-emissions-with-ai-the-role-of-machine-learning-in-energy-efficiency%2F&psig=AOvVaw0M9-lFdXC7Co7tPwyLSg91&ust=1705763761740000&source=images&cd=vfe&opi=89978449&ved=0CBUQ3YkBahcKEwiwwo-r3-mDAxUAAAAAHQAAAAAQAw)

**C. Graphic Illustration of Improvements Over Time (Figure 1)**

Figure 1 provides a visual representation of continuous improvement over time resulting from the application of these best practices. It highlights the combined impact of the "4Ms" on reducing energy consumption and CO2 emissions over the years. This graphic illustration tangibly demonstrates the viability of these practices to mitigate the ML carbon footprint, emphasizing their importance in sustainable development initiatives.

**IV. Case Studies**

In this section, we examine two case studies to concretely illustrate the impact of model choices, hardware, and deployment practices on the carbon footprint in the field of Machine Learning (ML).



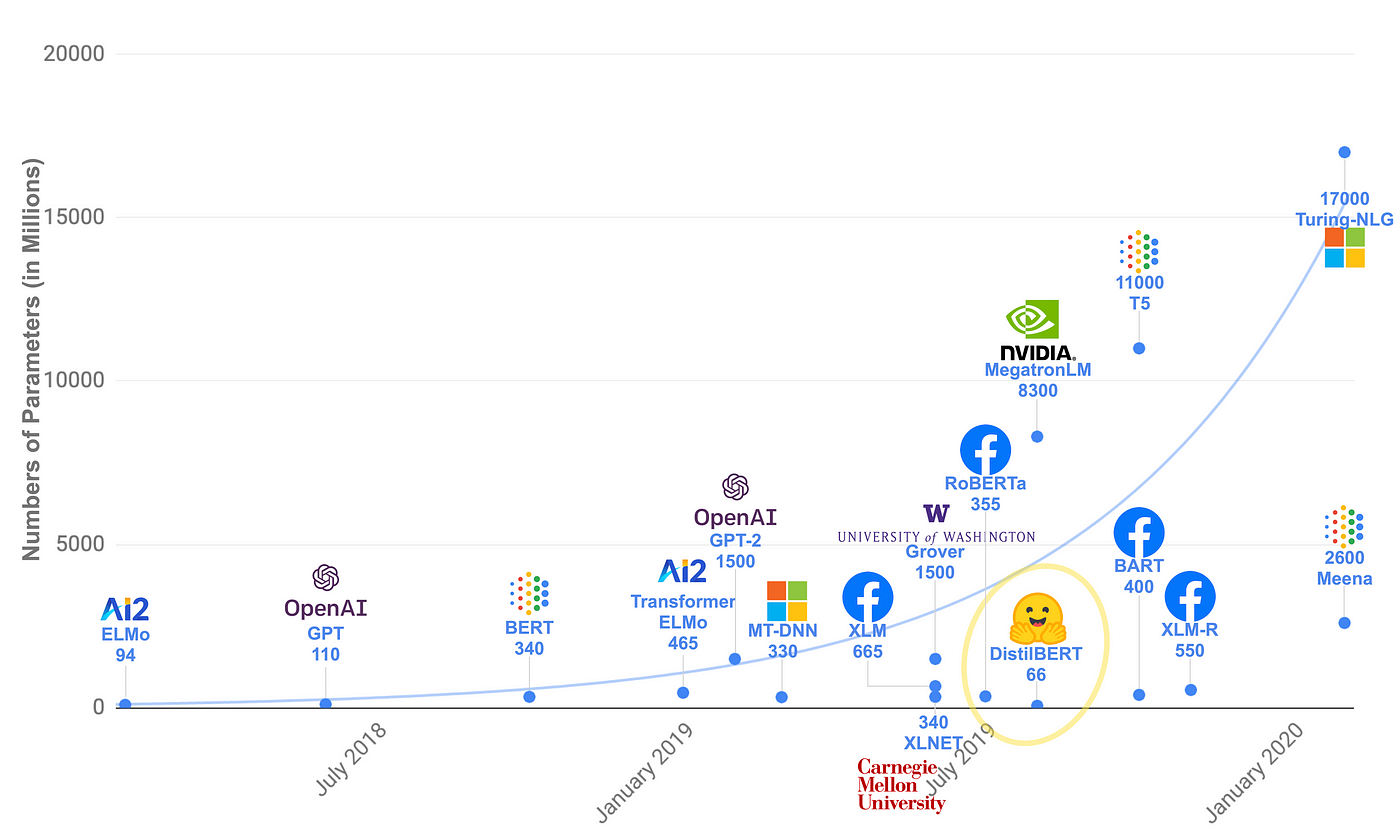
**Figure 6.** Optimization to elevate analytical capabilities and competitiveness.

Source : [Case- Study Research in Predictive Models for Black Carbon Emissions Les images peuvent être soumises à des droits d'auteur. En savoir plus](https://www.mdpi.com/1424-8220/22/10/3947)

**A. Case 1: Transformer vs. Evolved Transformer vs. Primer**

**1. Evolution of Models and Hardware Over Time**

By analyzing the evolution of models such as Transformer, Evolved Transformer, and Primer, along with the hardware used for their training, we can observe a significant trend towards more efficient architectures over time. Improvements in these models, combined with the use of more advanced hardware, demonstrate how the ML industry is progressing towards less energy-intensive solutions



**Figure 7.**Evolution of Transformers. Courtesy: [HuggingFace](https://huggingface.co/)

Source : [Evolution of Transformers — Part 1 | by Sanchit Goel | Medium](https://sanchman21.medium.com/evolution-of-transformers-part-1-faac3f19d780)

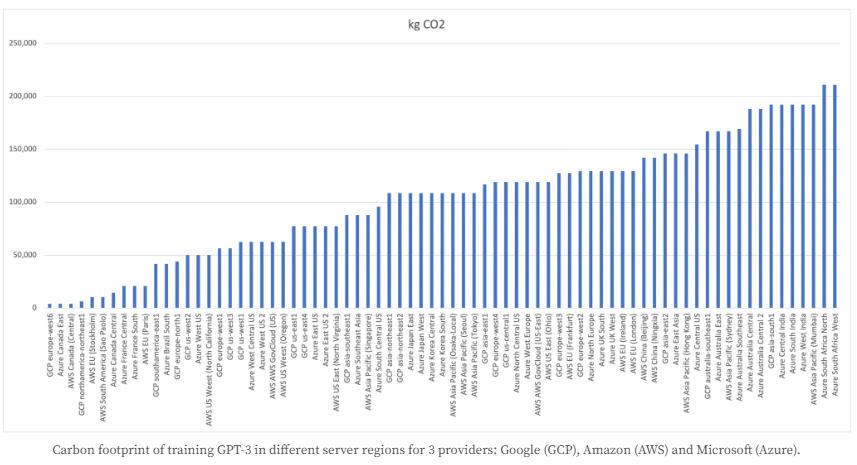
**2. Impact on Emission Reduction (Figure 7)**

Figure 7 graphically illustrates the impact of this evolution on emission reduction. By comparing the emissions generated by training older models with those of more recent models, we can quantify the gains achieved in terms of reducing the carbon footprint. This comparison highlights the importance of adopting newer models and technologies for more sustainable outcomes.

**B. Case 2: GPT-3, Google (GCP), Amazon (AWS) and Microsoft (Azure)**

**1**. Comparison of Models in Terms of Parameters, Acceleration Years, Energy Consumption, and CO2e

A detailed comparison between GPT-3, Google (GCP), Amazon (AWS) and Microsoft (Azure) , considering aspects such as the number of parameters, acceleration years, energy consumption, and CO2 equivalent emissions, provides insightful perspectives. This analysis offers crucial information on how design choices can influence the overall environmental impact of ML models.



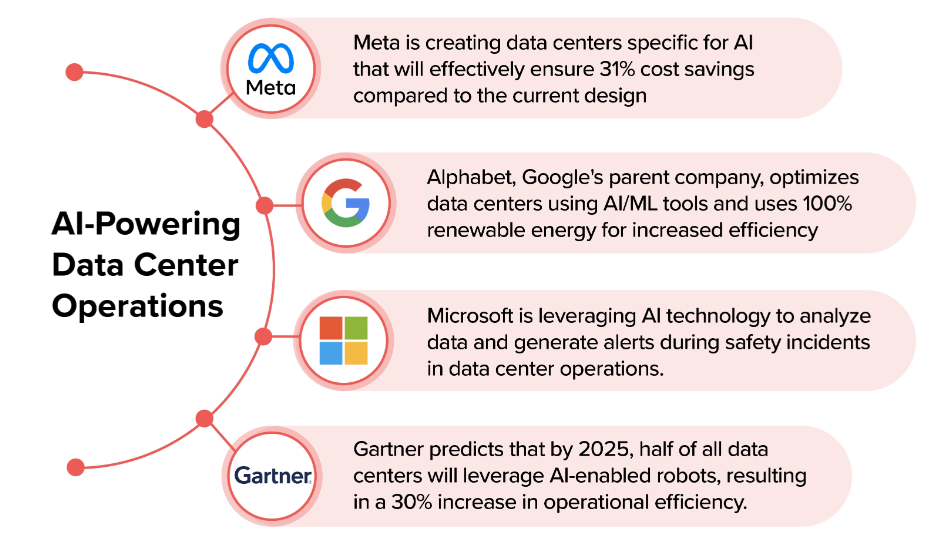
**Figure 8.**Carbon footprint of training GPT-3 in different server regions for 3 providers: Google (GCP), Amazon (AWS) and Microsoft (Azure).

Source : [Carbon footprint of training GPT-3 in different server regions for 3 providers: Google (GCP), Amazon (AWS) and Microsoft (Azure).](https://shrinkthatfootprint.com/carbon-footprint-of-training-gpt-3-and-large-language-models/)

**2. Impact of Datacenter Location on Emissions**

In examining the second case, we assess the impact of the datacenter location on emissions. By choosing data centers powered by renewable or cleaner energy sources, it is possible to significantly reduce the overall carbon footprint of ML. This part of the case study underscores the importance of considering geographical location when deploying ML models at scale.

In summary, these case studies offer a tangible view of the environmental implications of choices made in the ML domain. They demonstrate the importance of informed decision-making to reduce the carbon footprint associated with ML activities and encourage the development of more sustainable technologies.



**Figure 8.**How AI in data center operations is redefining the future of technology

Source : [How AI in data center operations is redefining the future of technology](https://appinventiv.com/blog/ai-in-data-center-operations/)