Day 10, Lecture 1: GHG Impact Assessment of AI

Speakers: **Lynn Kaack, George Kamiya**

TAs: Donna Vakalis and Ying-Jung Chen

CCAI Summer School 2024
July 22, 2024

Climate Change AI

Instructors

Lynn H. Kaack

Assistant Professor of Computer Science & Public Policy, Hertie School
Co-founder and Chair, Climate Change AI

George Kamiya

Previously led the International Energy Agency’s analysis on the energy and climate impacts of digitalisation

Learning objectives

- Get to know the terminology, methods and metrics of **impact assessment**
- Understand how machine learning (ML) has “**direct**” and “**indirect**” effects on climate change
- Understand how ML is responsible for energy use and greenhouse gas (GHG) emissions from **data centres and end user devices**
- Understand how the use of ML can increase and/or decrease GHG emissions **through its application** in different sectors and services
- Understand how different stakeholders can **reduce** ML-related emissions and policy implications

Impact assessment

- Climate change mitigation: **ton CO_2equivalent ($CO_{2\text{-eq}}$)** per unit of time, product, service, etc.
- $CO_{2\text{-eq}}$ accounts for CO_2 and other greenhouse gases and converts them based on their **global warming potential (GWP)** over 100 yrs
- GWP uses the relative potency (by mass) and atmospheric lifetime, and is normalized to CO_2.
- **Example**: Methane (CH4) has a lifetime of about a decade, and is ~80 times as potent as CO_2 over 20 years. It has a GWP of 27-30.

Interactive poll

Have you ever estimated the greenhouse gas (GHG) emissions impact/footprint in the following situations?

[Multiple answers OK]

a. Comparing flights or other travel options
b. Calculating my annual carbon footprint using online tools
c. Estimating emissions from ML training jobs
d. Estimating the emissions impact of a project applying ML
e. Other
f. I’ve never estimated GHG emissions impacts

Product view: Life-cycle assessment (LCA)

- Assessing environmental impacts of a product, process, or service
- Whole value chain from cradle (raw material extraction), over use phase, to grave (disposal)
- Methods rely on goal and boundary definition and inventory analysis
- LCA is family of methods (does not deliver unique and objective result)

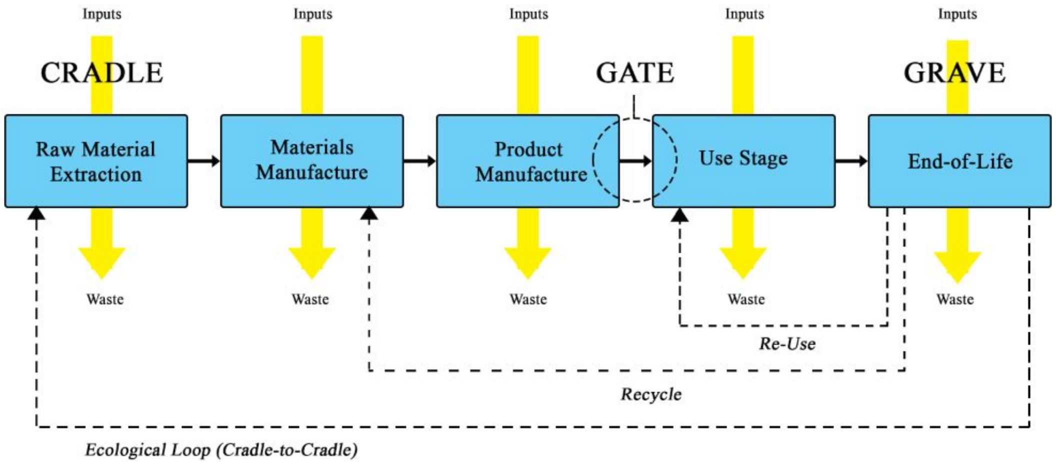


Image: Wikipedia

Life-cycle assessment (LCA)
Example of vehicle manufacturing

- Internal combustion engine (ICEV) vs. electric vehicle (BEV)
- Use phase (“operational”): Fuel production and consumption, maintenance
- Cradle & grave (“embodied”): Battery and vehicle manufacture
- Future life cycle emissions of BEVs decrease

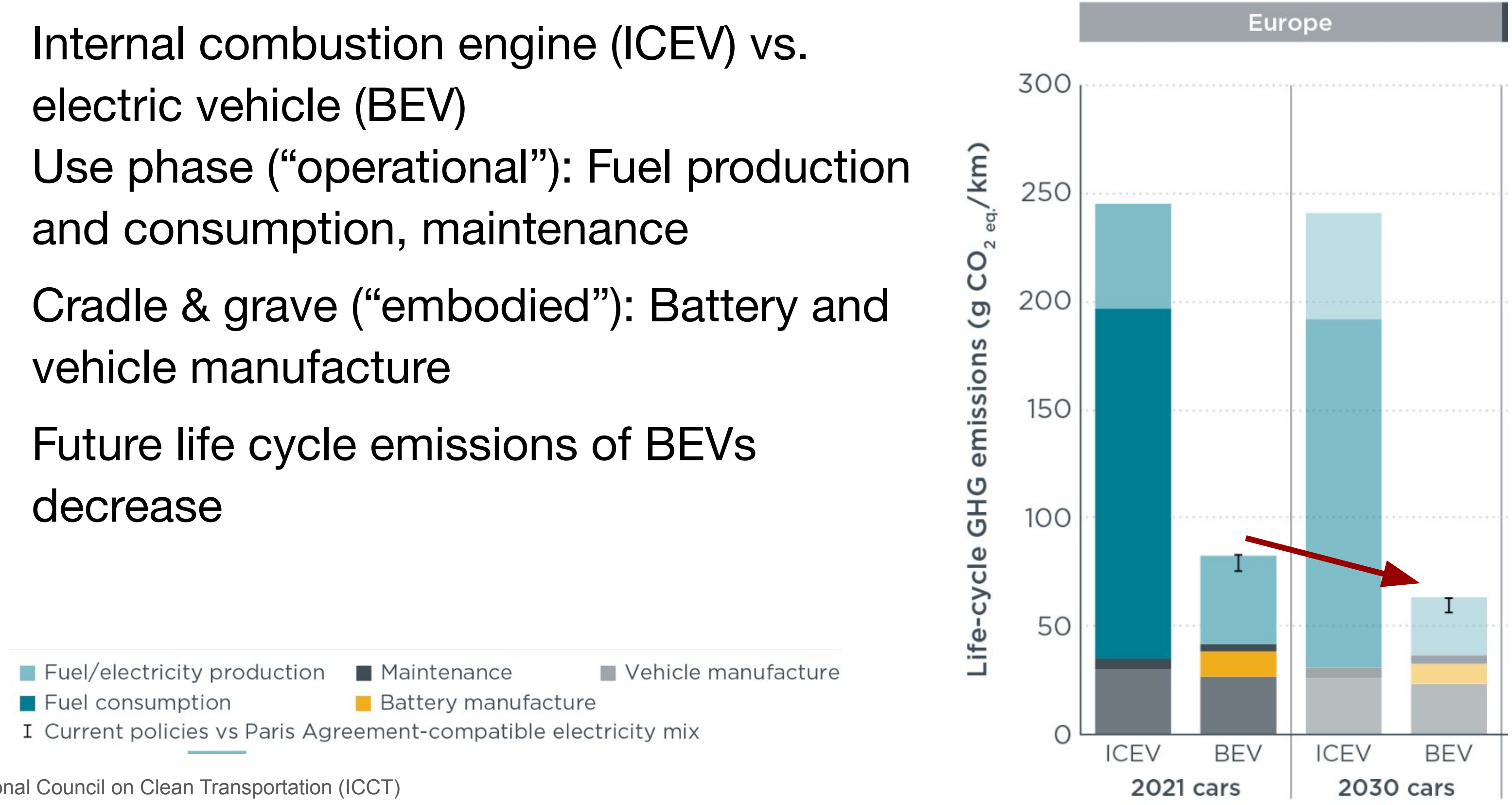


Image: International Council on Clean Transportation (ICCT)

Inventory analysis: GHG Protocol

- Global standardized frameworks to measure GHG emissions from private and public sector operations, value chains and mitigation actions
- Scope 1: emissions from direct use of fossil fuels, e.g. for heat, fleet vehicles
- Scope 2: emissions from purchased electricity and heat
- Scope 3: emissions from value chains including customers

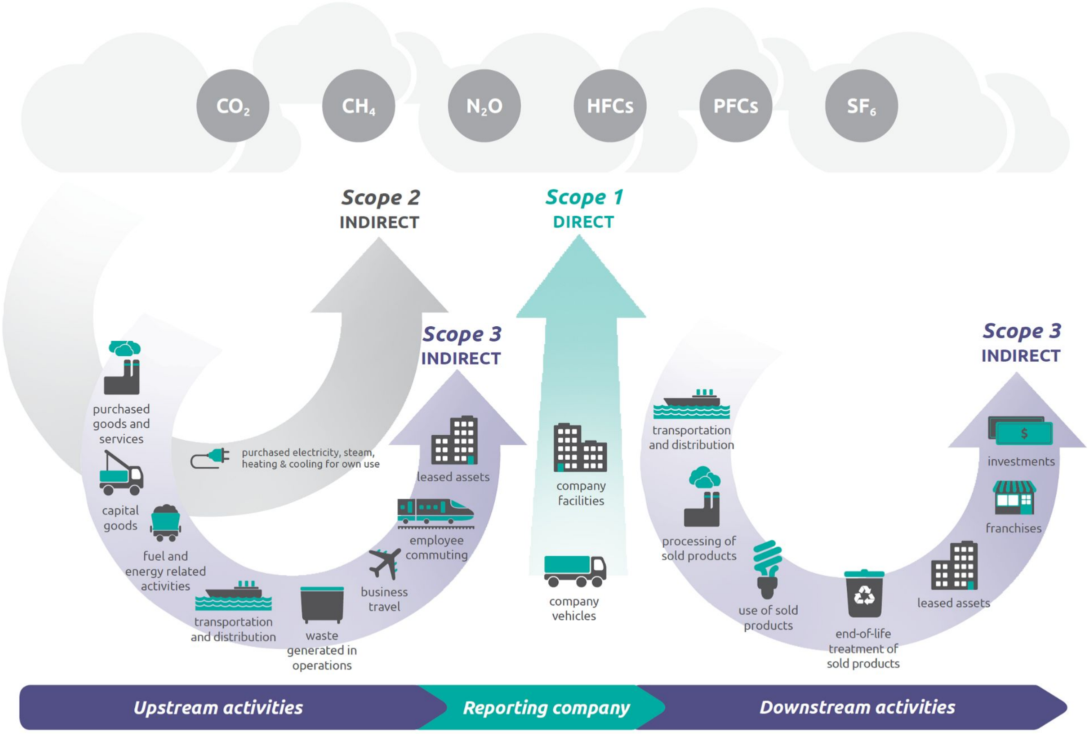
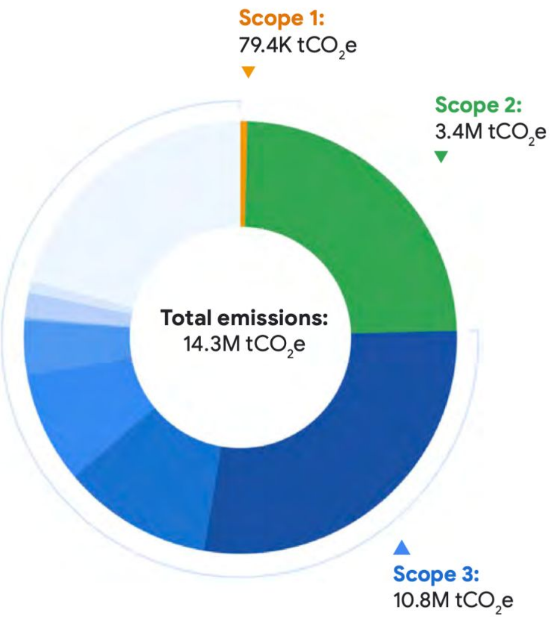


Image: US EPA

GHG Protocol
Example of Google

FIGURE 16 **Our 2023 carbon footprint**



Scope	tCO_2e	%
Scope 1	79,400	1%
Scope 2 (market-based)	3,423,400	24%
Scope 3: (1) Purchased goods and services	4,038,000	28%
Scope 3: (2) Capital goods	1,605,000	11%
Scope 3: (3) Fuel- and energy-related activities (not included in Scope 1 or 2)	1,186,000	8%
Scope 3: (4) Upstream transportation	584,000	4%
Scope 3: (5) Waste generated in operations	10,000	<1%
Scope 3: (6) Business travel	283,000	2%
Scope 3: (7) Employee commuting (including teleworking)	113,000	<1%
Scope 3: Other categories	2,993,000	21%
Scope 3 (total)	10,812,000	75%
Total emissions	**14,314,800**	**100%**

Google's inventory analysis (Environmental Report 2024)

ML's GHG emissions impact

ML applications in climate change mitigation

Emissions from ML computation & hardware

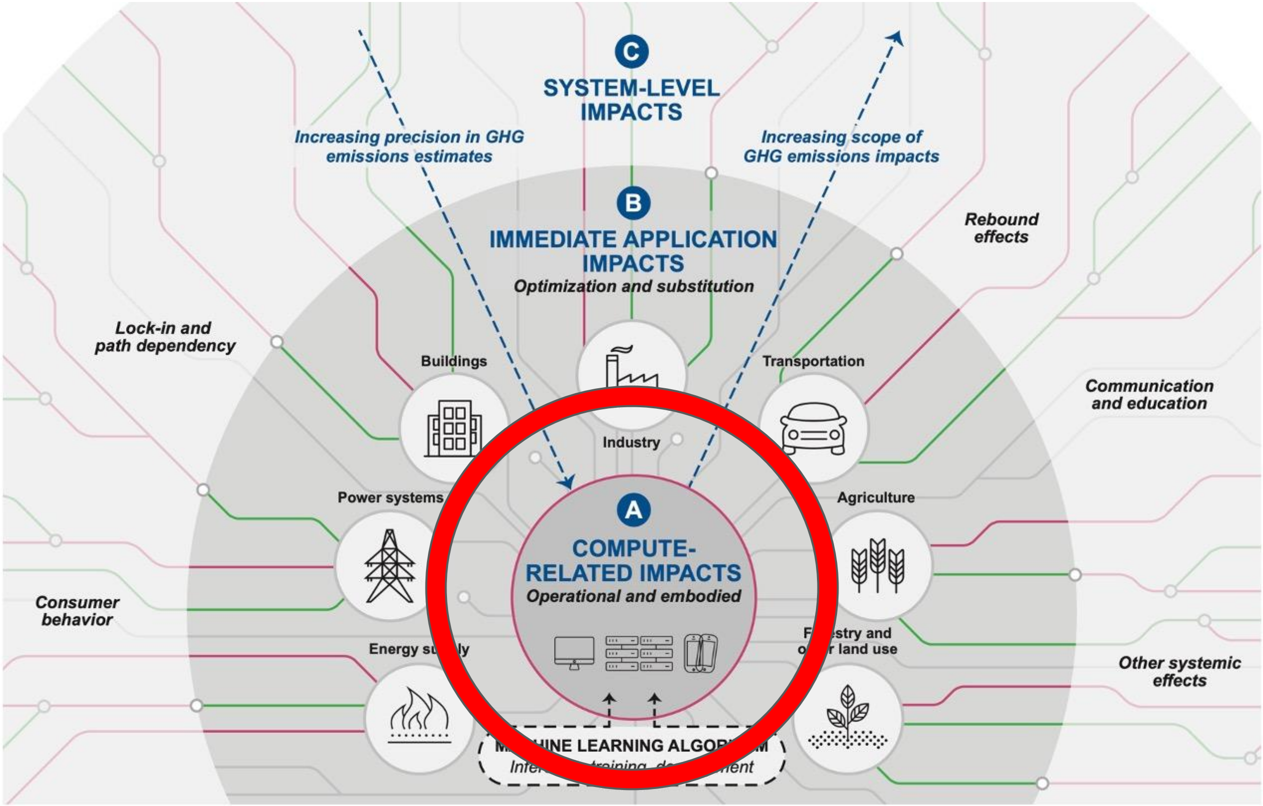
ML's system-level impacts

ML applications that increase emissions

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1-10.

Overall framework for ML

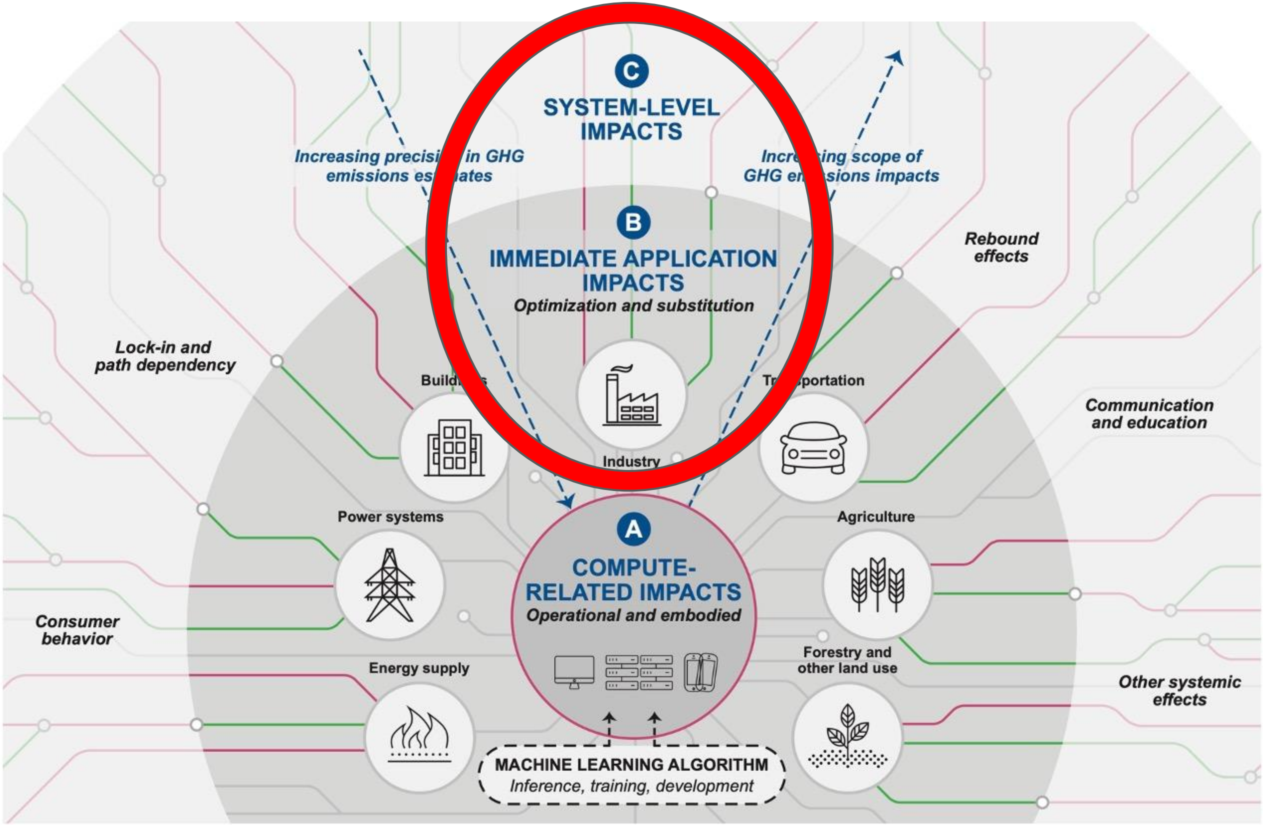
Compute-related or direct impacts



Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1-10.

Overall framework for ML

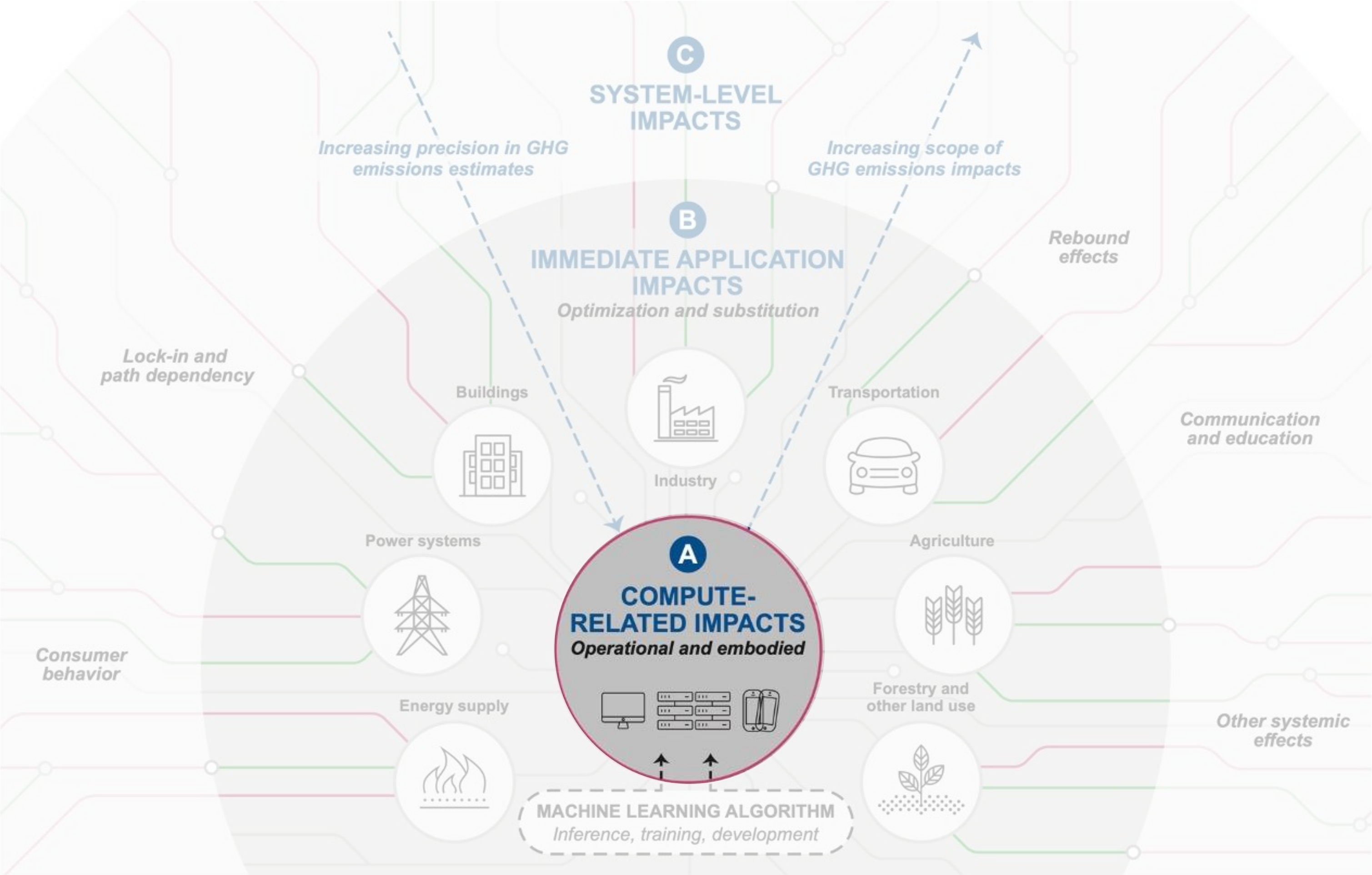
Application-related or indirect impacts



Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1-10.

Compute-related impacts

George Kamiya



Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1-10.

Computing-related impacts

- Life cycle emissions from digital infrastructure and devices used by ML, including:
  - **“Operational” (use-phase) emissions** from energy to power data centres, networks, and devices
  - **“Embodied” emissions** from raw materials extraction & processing, manufacturing, transport, and disposal of hardware

MIT Technology Review

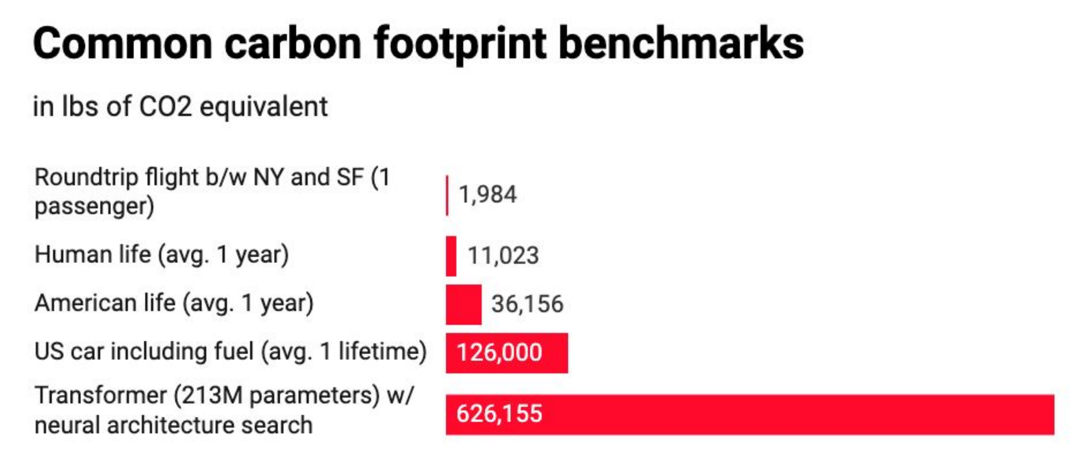
Topics Magazine Newsletters Events

Artificial intelligence / Machine learning

Training a single AI model can emit as much carbon as five cars in their lifetimes

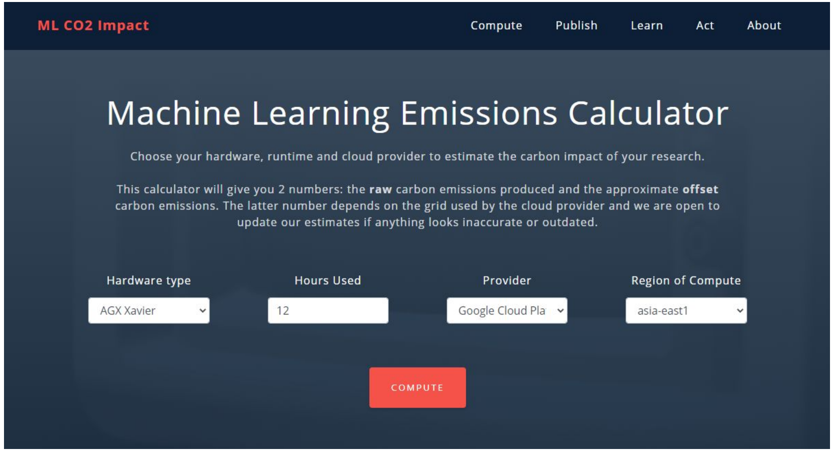
Deep learning has a terrible carbon footprint.

by **Karen Hao**



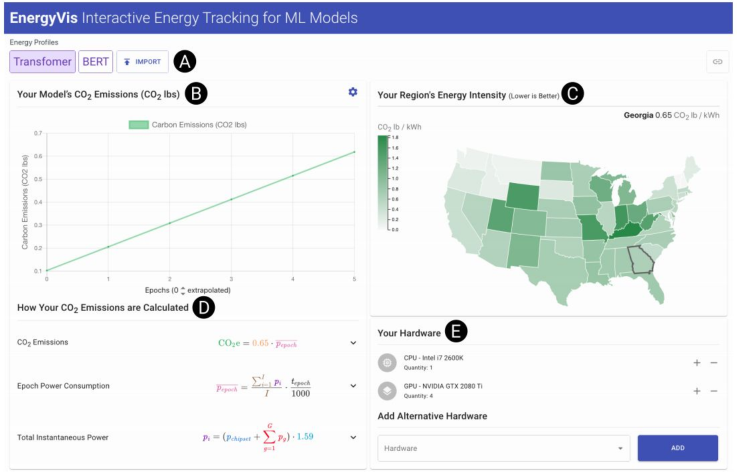
Hao (2019), Training a single AI model can emit as much carbon as five cars in their lifetimes, MIT Technology Review.
Strubell et al. (2019), Energy and Policy Considerations for Deep Learning in NLP.

ML energy and emissions calculators

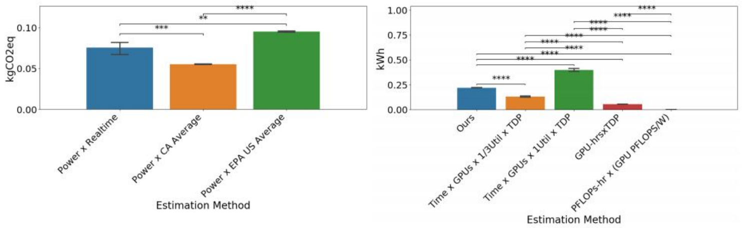


https://mlco2.github.io/impact/

Lacoste et al. (2019), Quantifying the Carbon Emissions of Machine Learning.



Shaikh et al. (2021), EnergyVis: Interactively Tracking and Exploring Energy Consumption for ML Models.

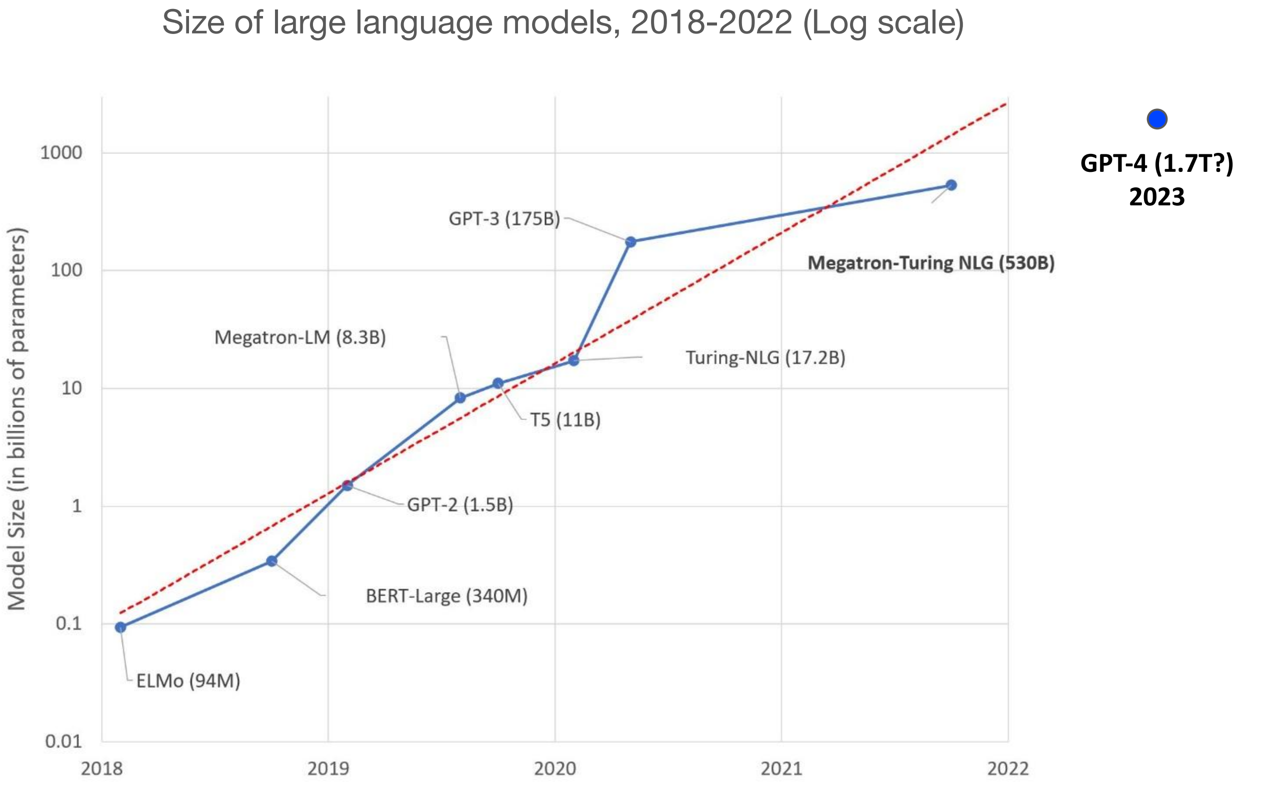


Estimate and track carbon emissions from your compute, quantify and analyze their impact.

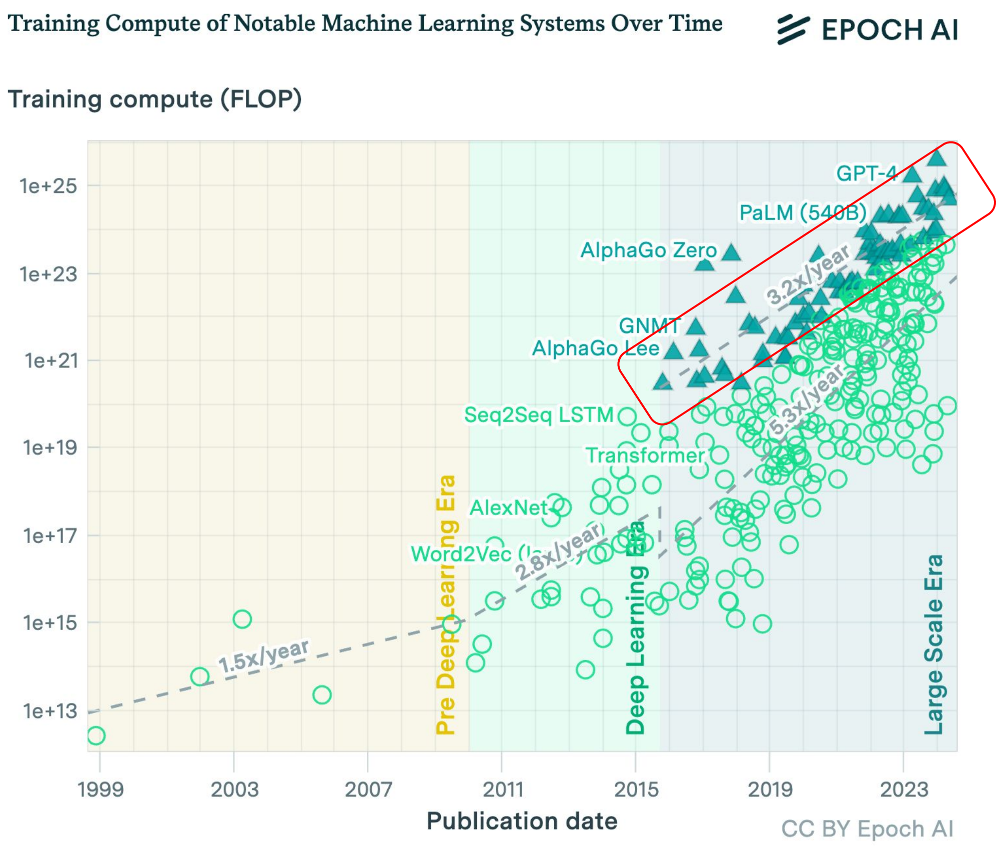
https://codecarbon.io/

Henderson et al. (2020), Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning.

The largest ML models today are >5000x larger than in 2018



Training compute for the largest models is tripling each year



Interactive poll

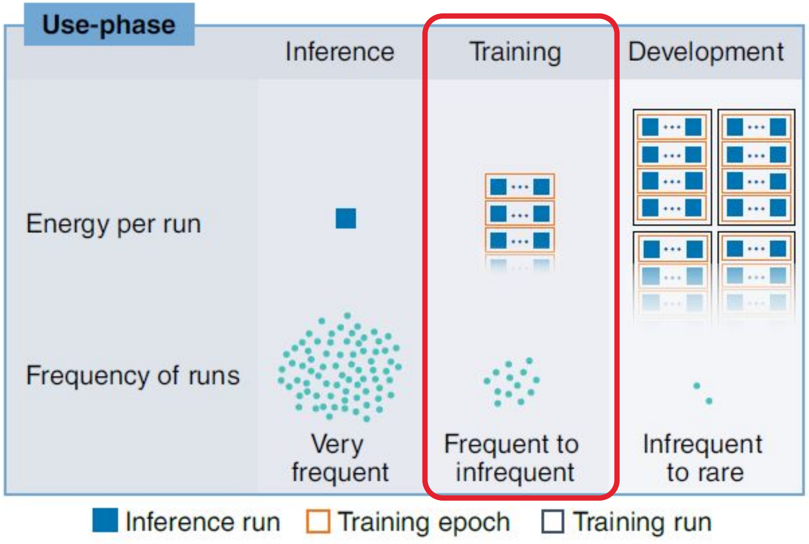
Do you think that the GHG footprint of <u>training</u> GPT-3 and GPT-4 is less, equal, or more than the cumulative GHG footprint of <u>deploying</u> ChatGPT since its public release?

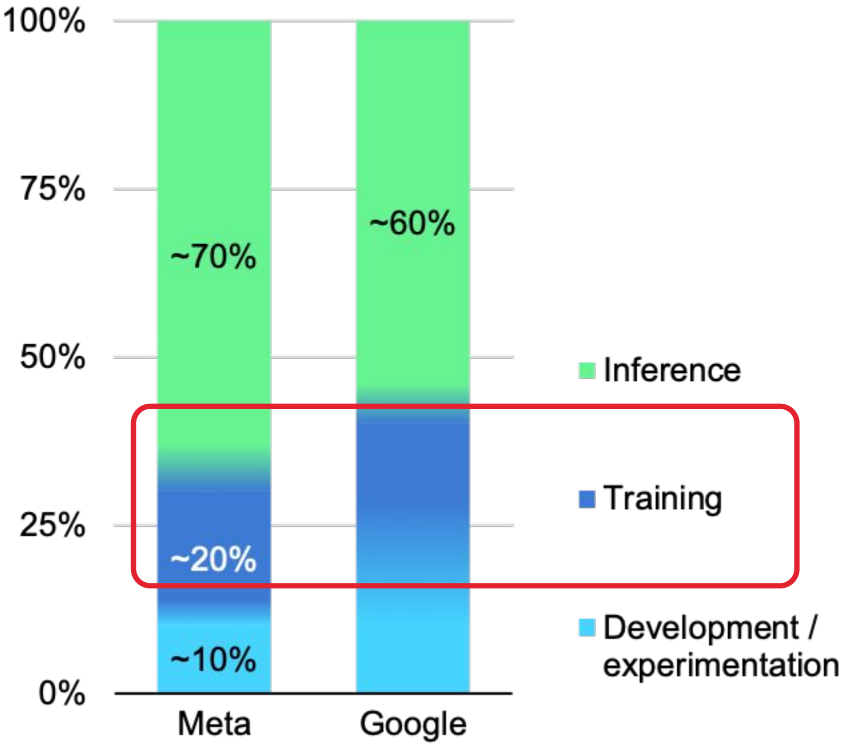
a. Training is less than the deployment

b. About the same

c. Training is more than deployment

Training is just one part of total ML-related energy use





Kaack et al. (2022). Aligning artificial intelligence with climate change mitigation; Wu et al. (2022). Sustainable AI: Environmental Implications, Challenges and Opportunities; Patterson et al. (2022). The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink.

Energy use of inference

	BLOOMz-7B	BLOOMz-3B	BLOOMz-1B	BLOOMz-560M
Training energy (kWh)	51,686	25,634	17,052	10,505
Finetuning energy (kWh)	7,571	3,242	1,081	543
Inference energy (kWh)	1.0×10^{-4}	7.3×10^{-5}	6.2×10^{-5}	5.4×10^{-5}
Cost parity (# inferences)	592,570,000	395,602,740	292,467,741	204,592,592

Table 5: The BLOOMz models from our study with their training energy cost (from [31]), finetuning energy cost (from [34]), inference cost (from the present study), and cost parity, as the number of inferences required to sum to the training cost.

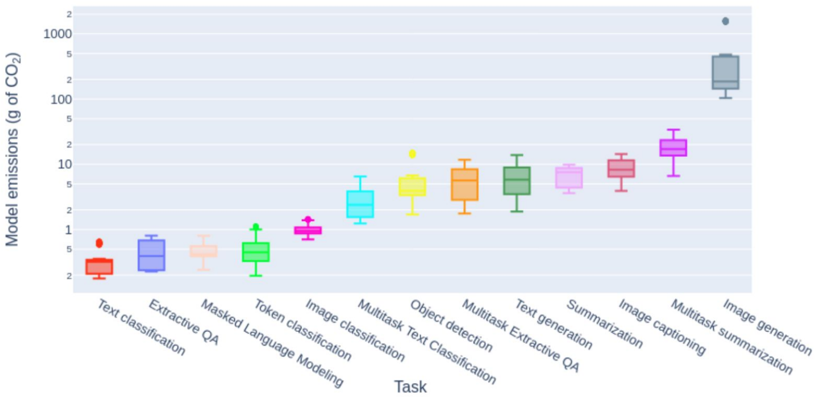


Figure 1: The tasks examined in our study and the average quantity of carbon emissions they produced (in g of CO_2eq) for 1,000 queries. N.B. The y axis is in logarithmic scale.

Luccioni et al. (2024), Power Hungry Processing: Watts Driving the Cost of AI Deployment?

Beware of outdated numbers and references

CLIMATE

AI already uses as much energy as a small country. It's only the beginning.

The energy needed to support data storage is expected to double by 20…
can do something to stop it.

by **Brian Calvert**
Mar 28, 2024, 12:00 PM GMT

US homes. According to the IEA, a single Google search takes 0.3 watt-hours of electricity, while a ChatGPT request takes 2.9 watt-hours. (An

The New York Times

Google Details, and Defends, Its Use of Electricity

By James Glanz
Sept. 8, 2011

Google also released an estimate that an average search uses 0.3 watt-hours of electricity, a figure that may be difficult to understand intuitively.

Recent ‘projections’ of AI / data centre energy use

B B C Sign in Home News Sport Weather i

NEWS

Home | Election 2024 | InDepth | Israel-Gaza war | Cost of Living | War in Ukraine | Climate | UK | W

Technology

Warning AI industry could use as much energy as the Netherlands

10 October 2023

B B C Sign in Home News Sport Weather iPlayer Soun

NEWS

Home | Election 2024 | InDepth | Israel-Gaza war | Cost of Living | War in Ukraine | Climate | UK | World | Business

Technology

Data centre power use 'to surge six-fold in 10 years'

26 March · Comments

Bloomberg UK

Live TV Markets Economics Industries Tech Politics Businessweek Opinion More

The AI Race: Why AI Is So Expensive | DOJ Scrutiny | Chip Arms Race | Startups to Watch

Technology | AI

AI Computing Is on Pace to Consume More Energy Than India, Arm Says

By Ian King

April 17, 2024 at 4:00 PM GMT+1

Goldman Sachs Intelligence

AI is poised to drive 160% increase in data center power demand

Published on 14 MAY 2024

Forbes

May 30, 1999

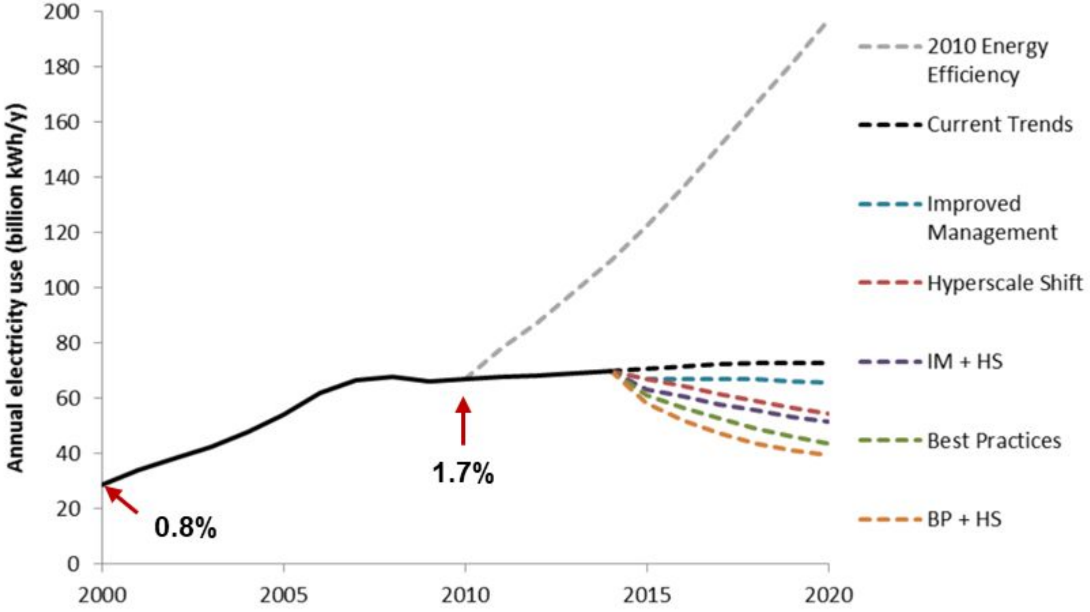
Dig more coal -- the PCs are coming

This article is more than 10 years old.

"It's now reasonable to project that half of the [US] electric grid will be powering the digital-Internet economy within the next decade."

Forbes (1999). https://www.forbes.com/forbes/1999/0531/6311070a.html#21a128aa2580.

US data centre energy use trends

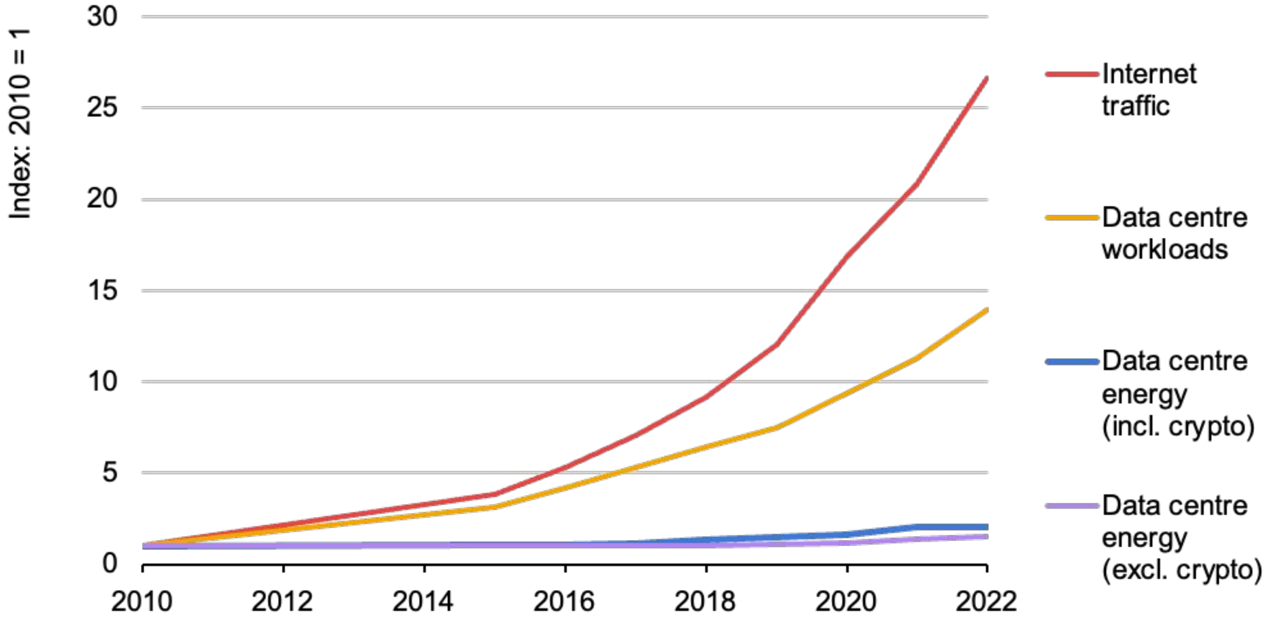


Source: LBNL / Shehabi et al. (2016). United States Data Center Energy Usage Report.

Data centre energy use in the US more than doubled between 2000 and 2010, but remained less than 2% of total electricity use

Global data centre energy use trends

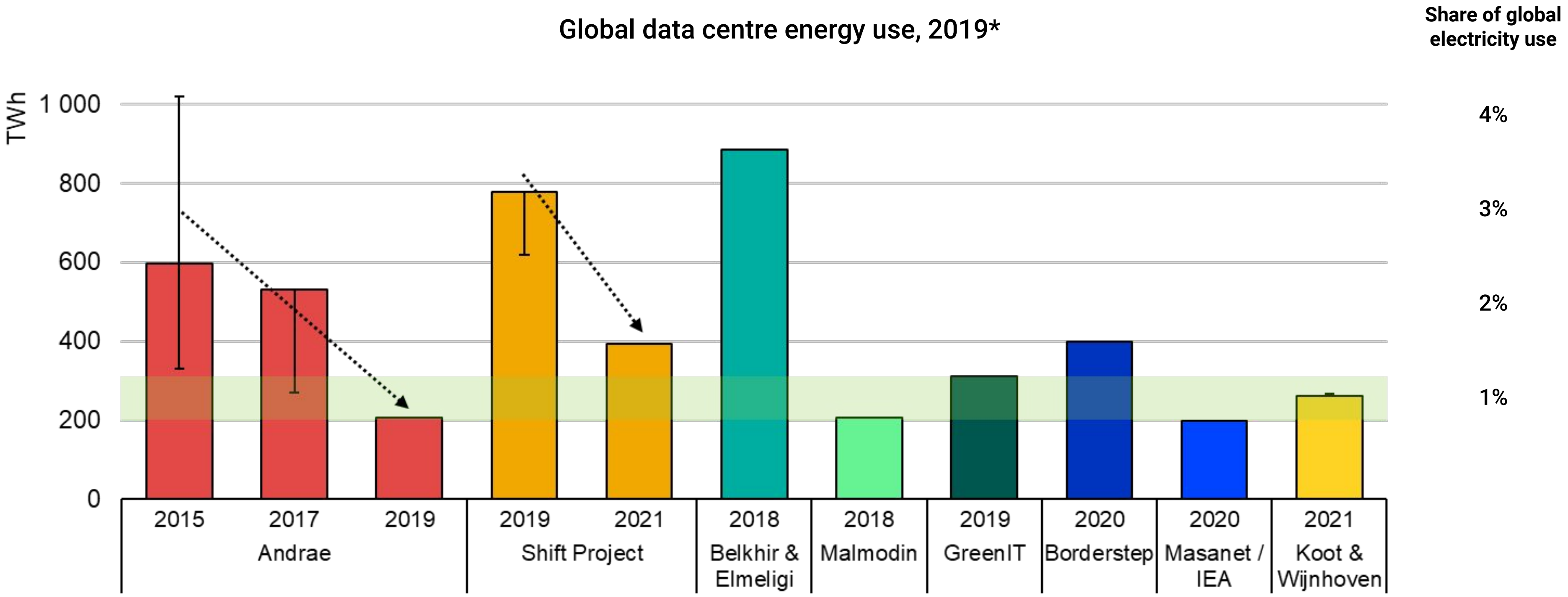
Internet traffic, data centre workloads and data centre energy use, 2010-2022



Sources: Masanet et al. (2020). Recalibrating global data center energy-use estimates. IEA (2023). Data centres and data transmission networks; Cisco (2018). Global Cloud Index: Forecast and Methodology, 2016-2021; Cisco (2019). Visual Networking Index: Forecast and Trends, 2017-2022, Telegeography (2023), The State of the Network.

Note: Figures exclude cryptocurrency mining

Comparing global estimates



*2019 values except for Borderstep and Malmodin which are 2018. Shift Project (2019) values are extrapolations of stated 2017 and 2020 values.
Values typically exclude cryptocurrency mining, which was likely around 60 TWh in 2019. Shift Project (2021) value in this chart excludes bitcoin.

Sources: Andrae & Edler (2015); Andrae (2017); Andrae (2019); Andrae (2019); Andrae (2020); The Shift Project (2019); The Shift Project (2021); Belkhir & Elmeligi (2018); Malmodin & Lunden (2018); Bordage / GreenIT.fr (2019); Hintemann / Borderstep (2020); IEA (2020); Masanet et al. (2020); Koot & Wijnhoven (2021).

The Guardian

News | Opinion | Sport | Culture | Lifestyle

Environment ▶ Climate change Wildlife Energy Pollution

Guardian Environment Network Environment

‘Tsunami of data’ could consume one fifth of global electricity by 2025

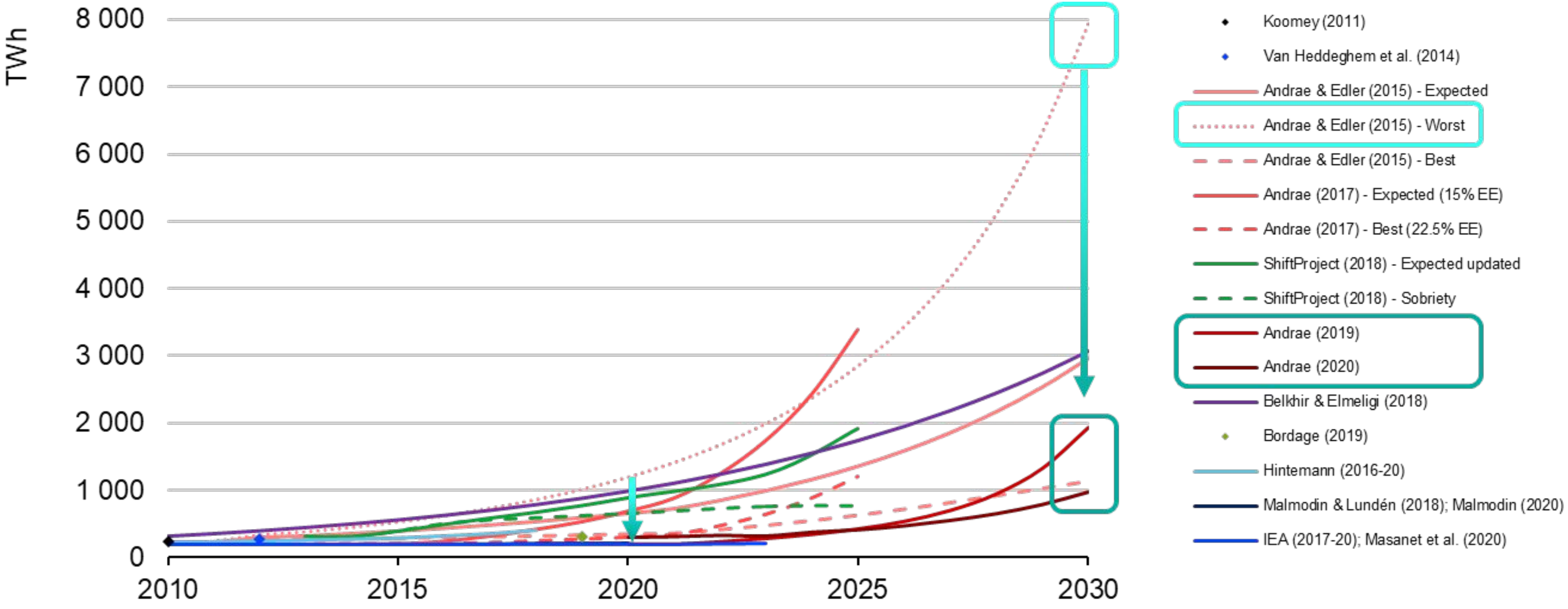
Billions of internet-connected devices could produce 3.5% of global emissions within 10 years and 14% by 2040, according to new research, reports Climate Home News

Mon 11 Dec 2017 13.27 GMT

1,454 73

The Guardian (2017). https://www.theguardian.com/environment/2017/dec/11/tsunami-of-data-could-consume-fifth-global-electricity-by-2025.

Comparing data centre energy projections



Sources: Koomey (2011), Growth in Data Center Electricity Use 2005 to 2010; Van Heddeghem et al. (2014), Trends in worldwide ICT electricity consumption from 2007 to 2012; Andrae & Edler (2015), On Global Electricity Usage of Communication Technology: Trends to 2030; Andrae (2017), Total Power Consumption Forecast; The Shift Project (2018), Lean ICT: Towards Digital Sobriety; Andrae (2019), Projecting the chiaroscuro of the electricity use of communication and computing from 2018 to 2030; Andrae (2019), Comparison of Several Simplistic High-Level Approaches for Estimating the Global Energy and Electricity Use of ICT Networks and Data Centers; Andrae (2020), New perspectives on internet electricity use in 2030; Belkhir & Elmeligi (2018), Assessing ICT global emissions footprint: Trends to 2040 & recommendations; Bordage / GreenIT.fr (2019), Environmental footprint of the digital world; Hintemann & Claus (2016), Green Cloud? The current and future development of energy consumption by data centers, networks and end-user devices; Hintemann / Borderstep (2020), Efficiency gains are not enough: Data center energy consumption continues to rise significantly; Malmodin & Lunden (2018), The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015; Malmodin (2020), Energy consumption and carbon emissions in the ICT sector (presentation to TechUK); IEA (2017), Digitalization & Energy; IEA (2018-20), Tracking Clean Energy Progress: Data centres and data transmission networks; Masanet et al. (2020), Recalibrating global data center energy-use estimates.

Recent headlines on data centre energy projections

S&P Global
https://www.spglobal.com › latest-news-headlines › datac...

Datacenter power demand to double in three years – IEA

24 Jan 2024 — Of the 8,000 **datacenters** in operation globally, 33% are based in the US, where **datacenter electricity consumption** is set to grow from 200 TWh in ...

EL PAÍS English
https://english.elpais.com › Economy And Business

Data centers, AI and cryptocurrencies will double electricity

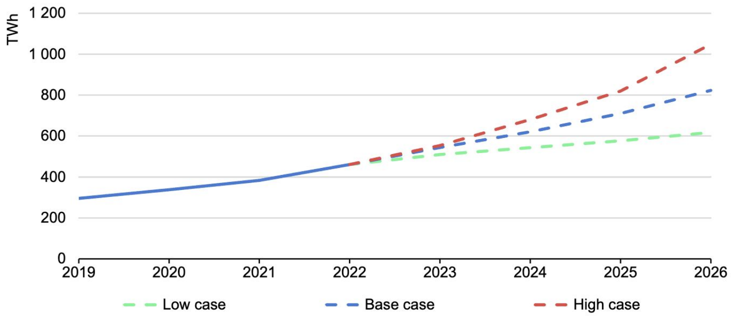
25 Jan 2024 — **Data** centers, artificial intelligence (AI) and cryptocurrencies will **dou** **electricity** consumption by **2026**, according to the latest ...

Yale E360
https://e360.yale.edu › features › artificial-intelligence-c...

As Use of A.I. Soars, So Does the Energy and Water It ...

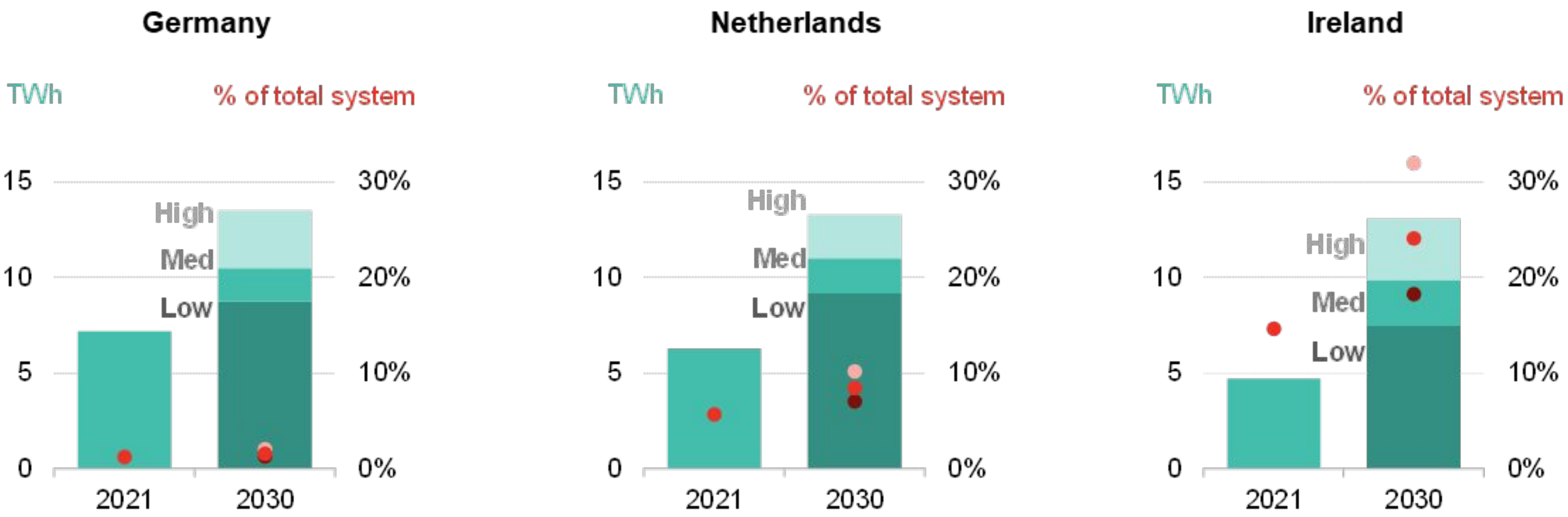
6 Feb 2024 — The **International Energy Agency** (**IEA**) projects that **data** centers' **electricity consumption in 2026** will be **double** that of 2022 — 1,000 ...

Global electricity demand from data centres, AI, and cryptocurrencies, 2019-2026



Notes: Includes traditional data centres, dedicated AI data centres, and cryptocurrency consumption; excludes demand from data transmission networks. The base case scenario has been used in the overall forecast in this report. Low and high case scenarios reflect the uncertainties in the pace of deployment and efficiency gains amid future technological developments.

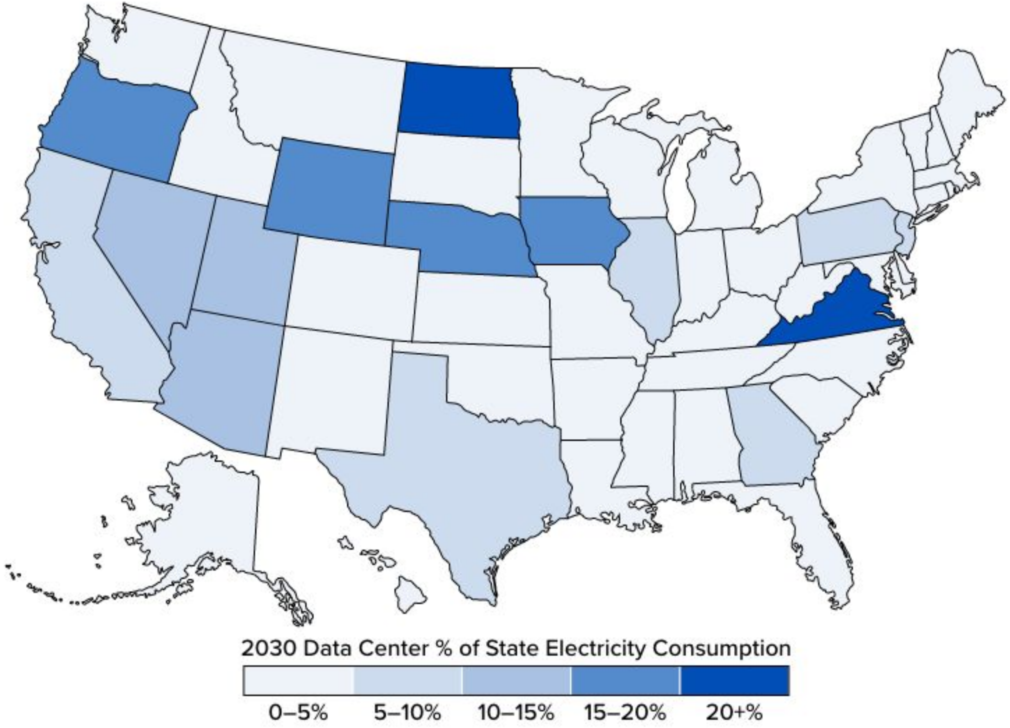
Uneven country (and local) grid impacts



Source: BloombergNEF, Statkraft, Eaton (2021). Data Centers and Decarbonization: Unlocking Flexibility in Europe's Data Centers.

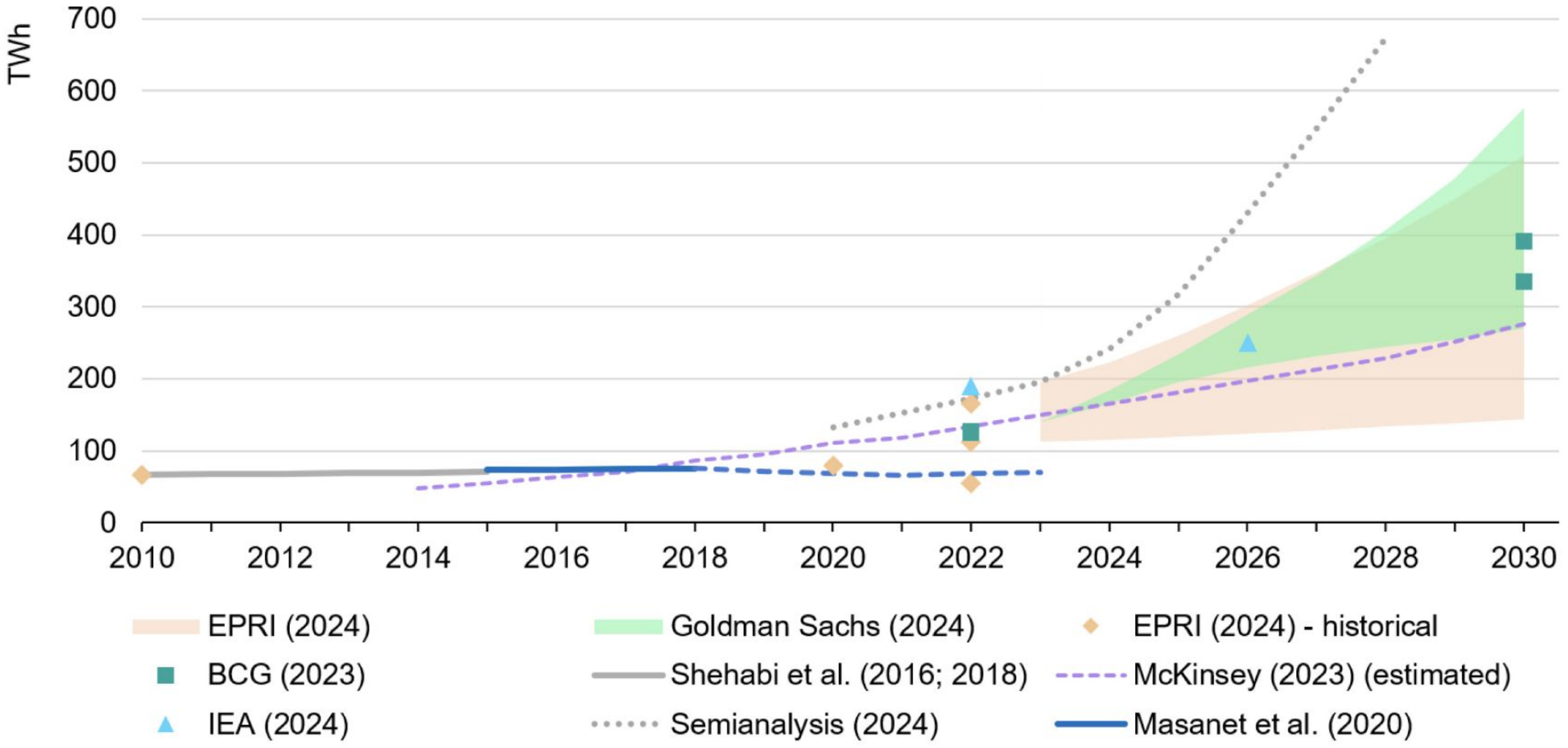
Impacts on the grid vary significantly by country, now and in the future

Uneven grid impacts in the US



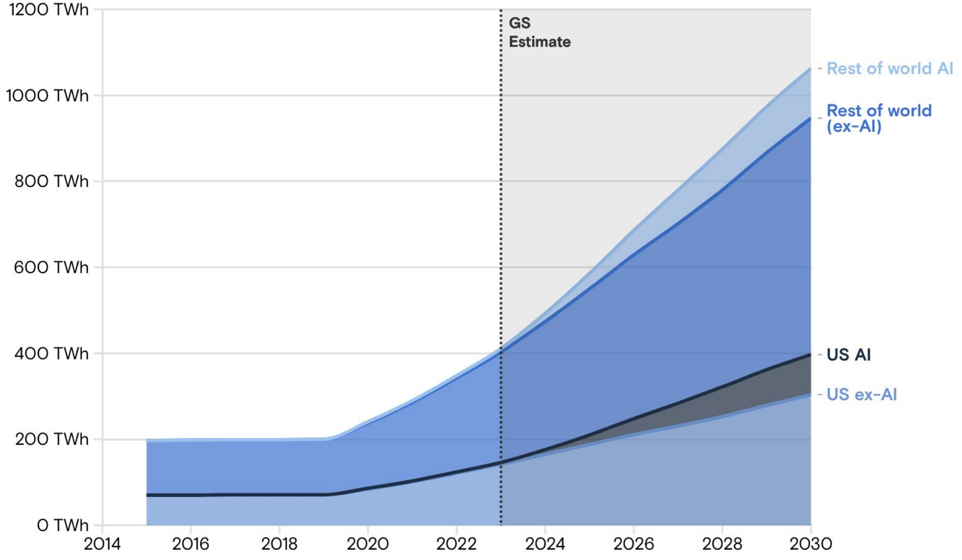
Source: EPRI (2024), Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption

Projections of US data centre energy use



Recent projections of AI energy use

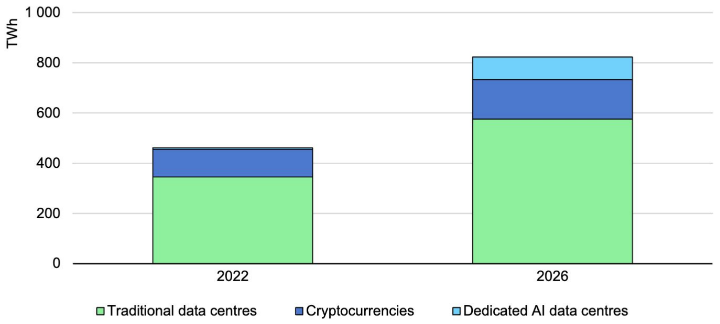
Data center power demand



Source: Masanet et al. (2020), Cisco, IEA, Goldman Sachs Research

Goldman Sachs

Estimated electricity demand from traditional data centres, dedicated AI data centres and cryptocurrencies, 2022 and 2026, base case

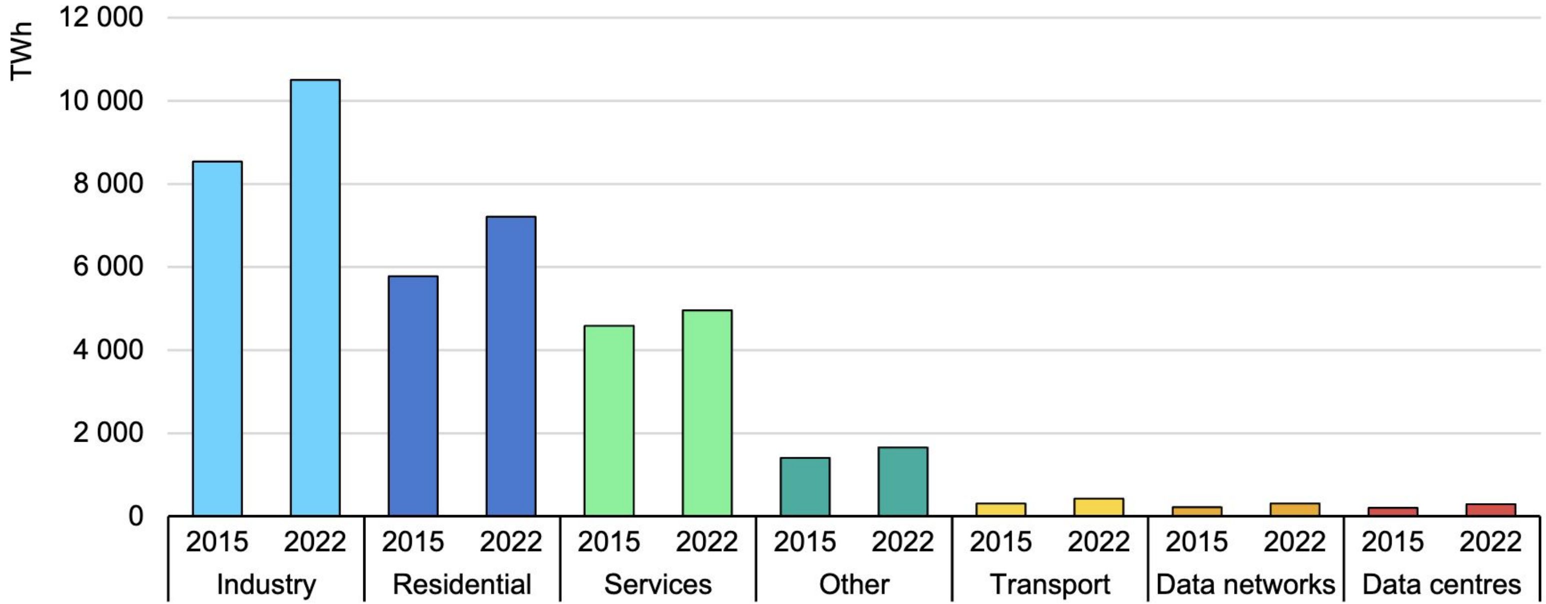


Note: Data centre electricity demand excludes consumption from data network centres.

Sources: IEA forecast based on data and projections from Data Centres and Data Transmission Networks; Joule (2023), Alex de Vries, The growing energy footprint of artificial intelligence; Crypto Carbon Ratings Institute, Indices; Ireland Central Statistics Office, Data Centres Metered Electricity Consumption 2022; and Danish Energy Agency, Denmark's Energy and Climate Outlook 2018.

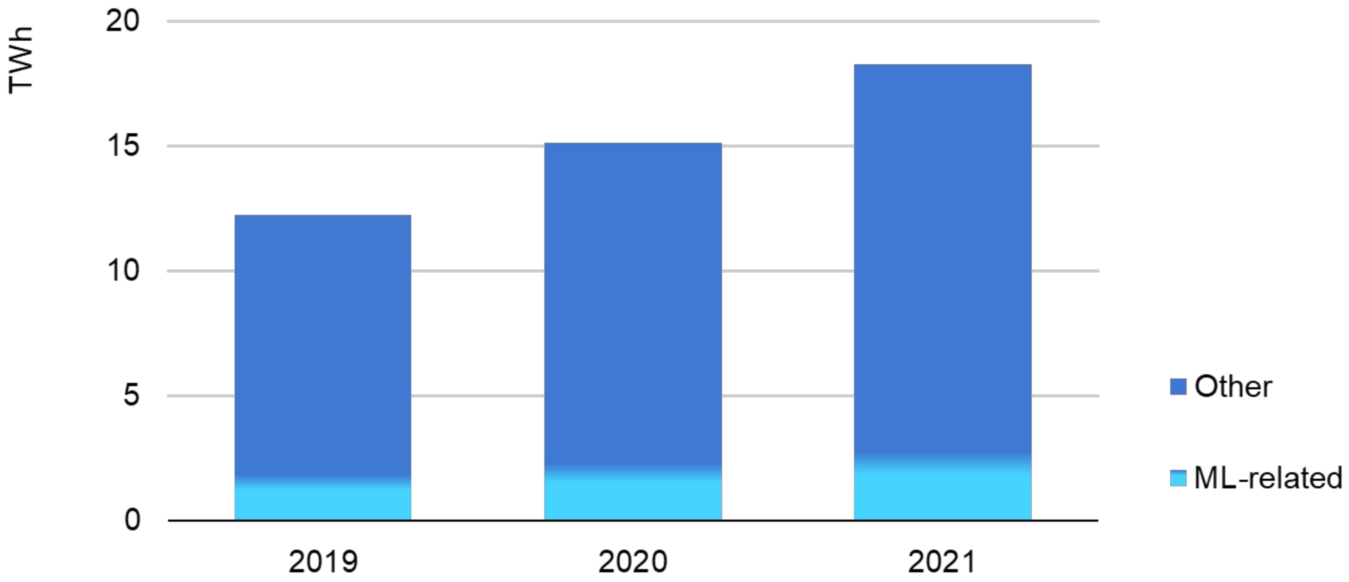
Electricity use by sector

Global final electricity demand by sector, 2015-2022



IEA (2024), Electricity Market Report 2024 Mid-Year Update.

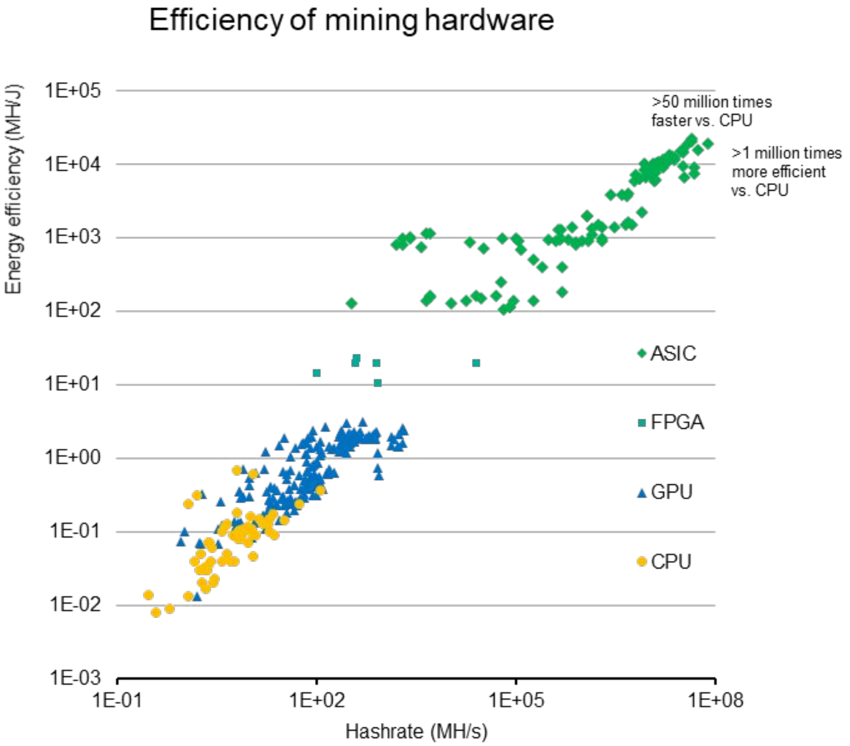
ML as a share of total data centre energy

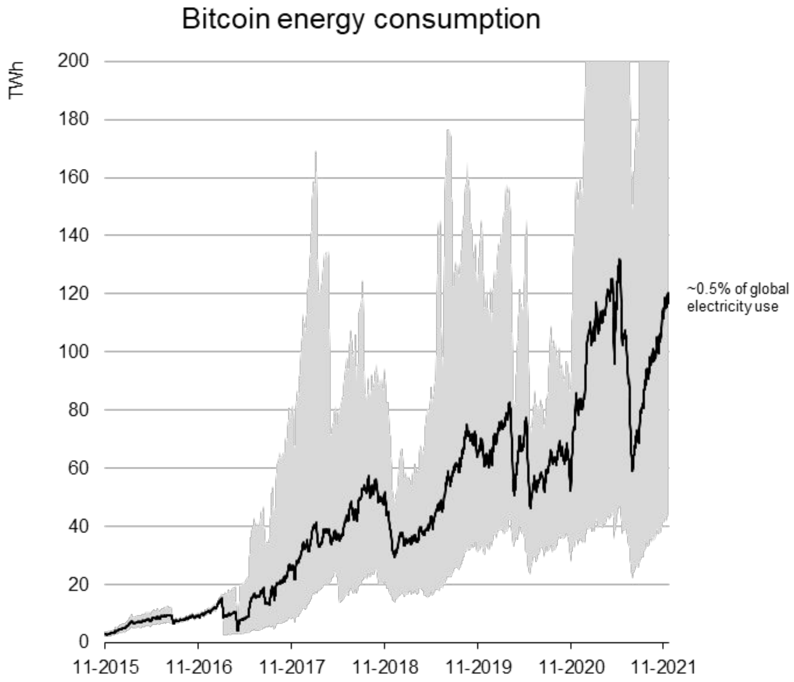


ML accounted for 10-15% of Google's total energy use
based on three sample weeks in 2019, 2020 and 2021

Google (2022), Google Environmental Report 2022. Patterson et al. (2022). The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink.

More efficient hardware has not moderated crypto energy use

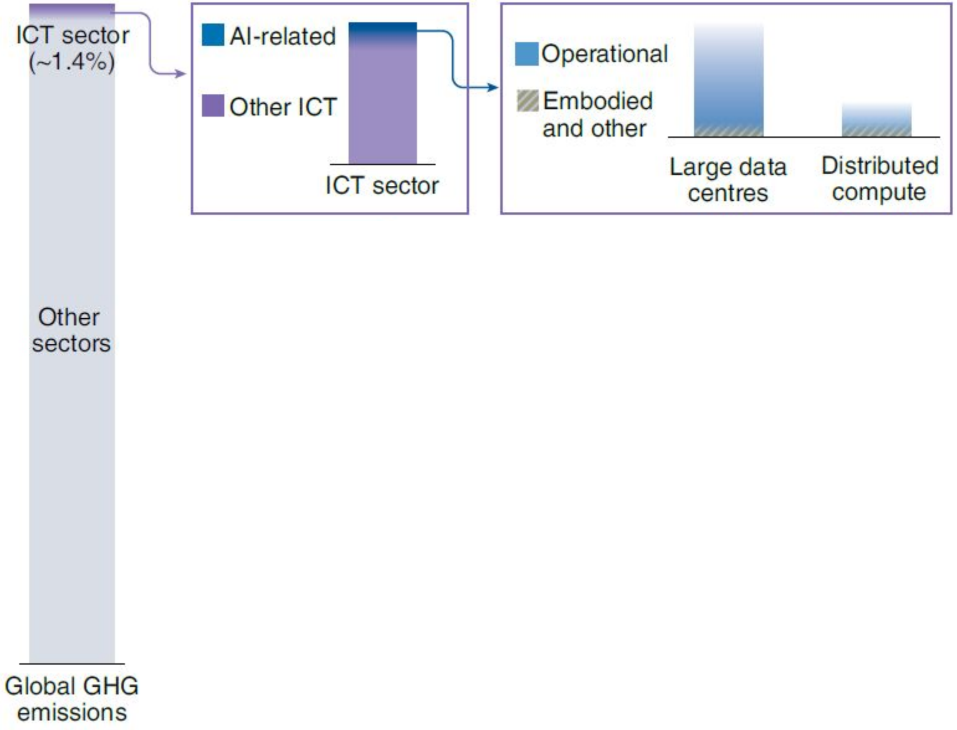




IEA (2019). Bitcoin energy use – mined the gap. https://www.iea.org/commentaries/bitcoin-energy-use-mined-the-gap.

Cambridge Centre for Alternative Finance (2020), Cambridge Bitcoin Electricity Consumption Index. https://www.cbeci.org/.

AI/ML-related energy use and GHG are relatively small…

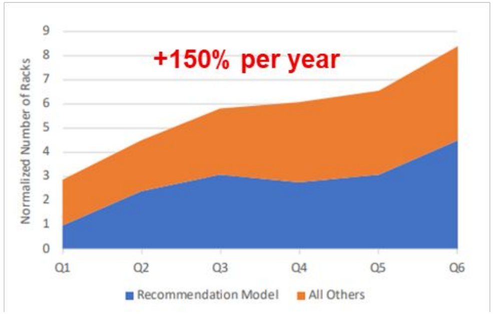


Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1-10.

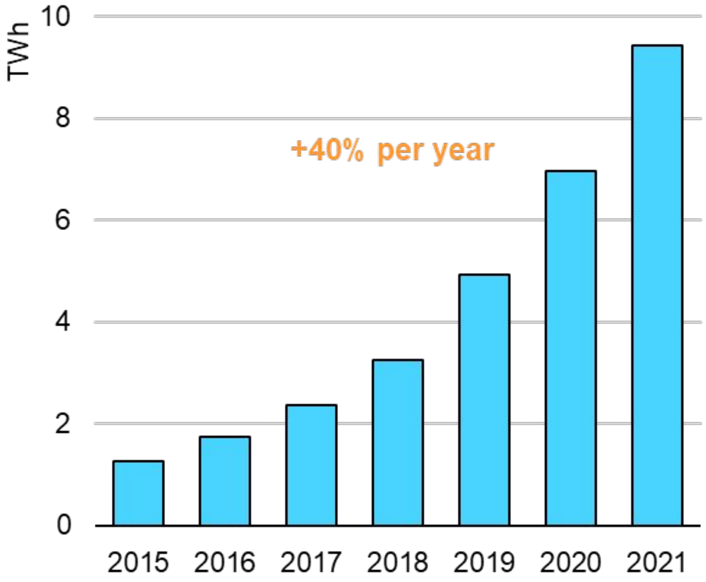
…but ML demand (and DC energy use) is growing quickly

Server demand for DL inference

Server compute demand for training

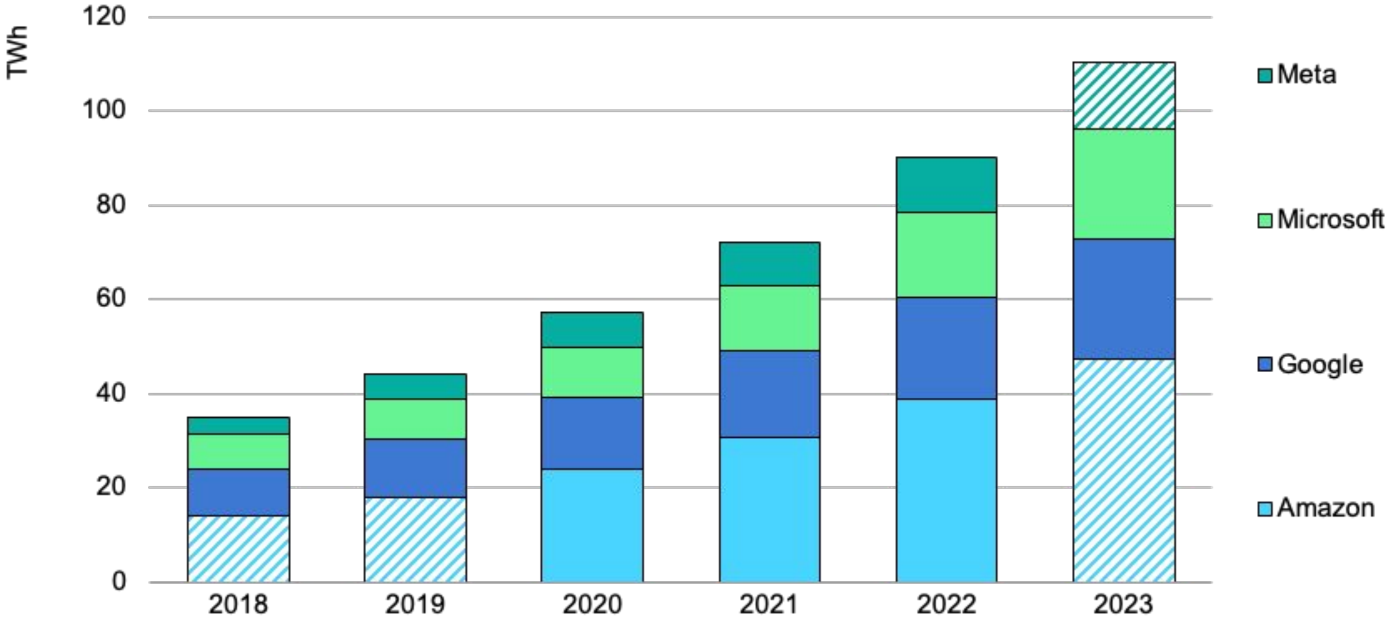


Facebook data centre energy use, 2015-2021



Park et al. (2018), Deep Learning Inference in Facebook Data Centers: Characterization, Performance Optimizations and Hardware Implications. Naumov et al. (2020), Deep Learning Training in Facebook Data Centers: Design of Scale-up and Scale-out Systems. Meta (2022), Sustainability Report 2021.

Electricity use by hyperscalers



*Amazon, Google, and Microsoft electricity use values are across their entire operations, e.g. including offices, stores, warehouses, etc. Only Meta reports data centre electricity use separately. Amazon did not publicly report electricity use for 2018-19 and 2023; values shown above are approximated based on reported Scope 2 emissions, reported renewable energy consumption, and other available energy indicators.

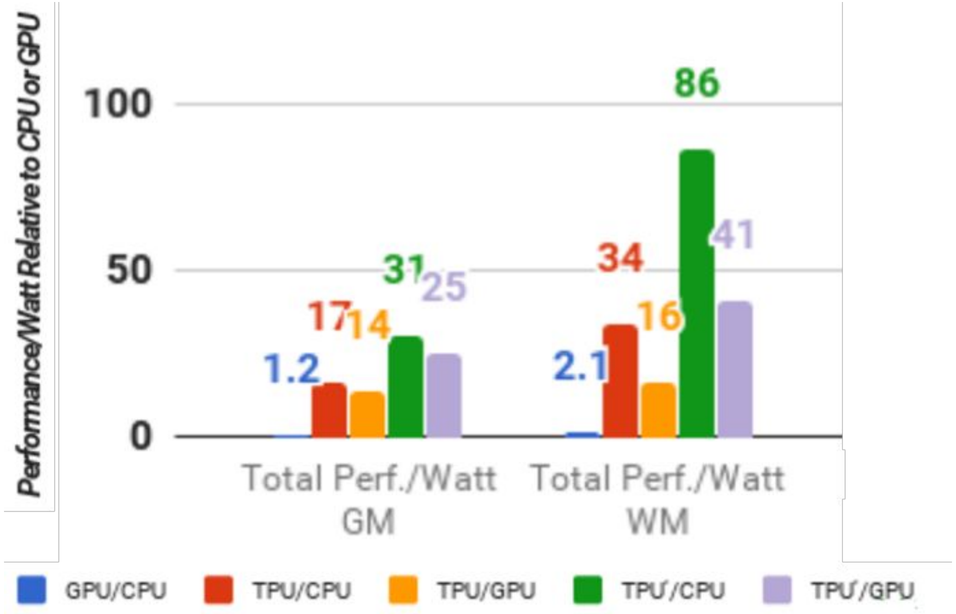
Sources: Meta (2022), Meta 2021 Sustainability Report. Microsoft (2022), 2021 Environmental Sustainability Report. Google (2022), Google Environmental Report 2022. Amazon (2022), Reporting and Downloads; CDP (2023).

Strategies to reduce compute-related impacts

- Model-related opportunities
- More efficient computing hardware
- Decarbonise electricity
- Decarbonise supply chains and increase circularity

Energy efficiency gains from specialised hardware

Relative performance/Watt of TPU, GPU and CPU servers

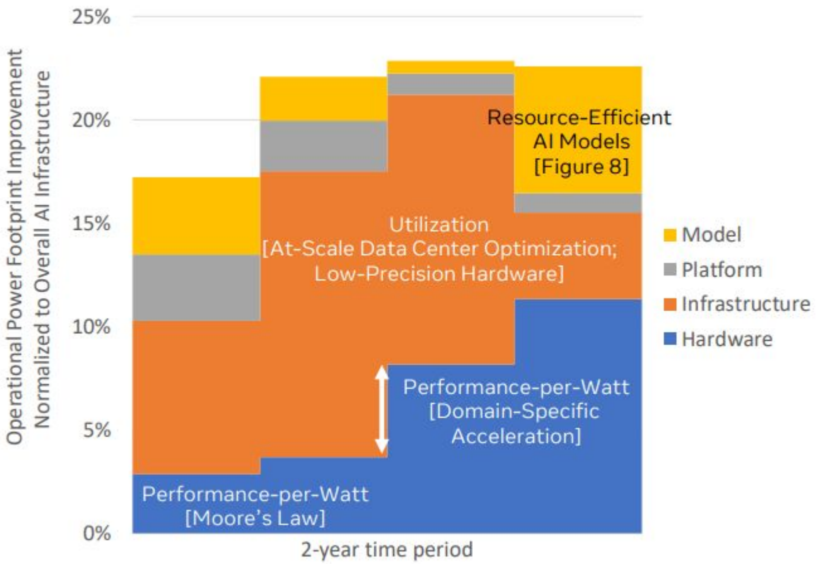


Application-specific integrated circuits (ASICs) for machine learning are 15-30x faster and 30-80x more energy efficient compared to a contemporary CPU or GPU

Jouppi et al. (2017). In-Datacenter Performance Analysis of a Tensor Processing Unit. https://dl.acm.org/doi/pdf/10.1145/3140659.3080246

Reducing energy use through hardware-software co-design

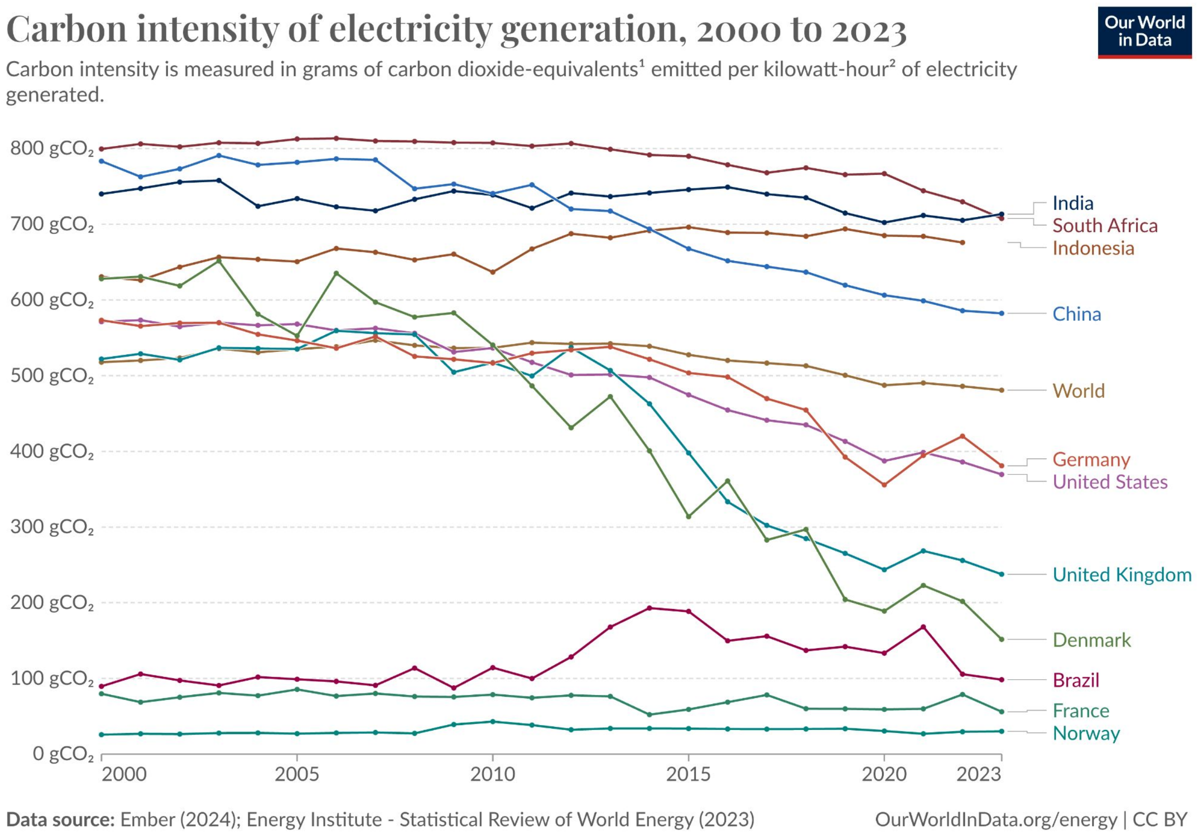
Operational power footprint reduction across Facebook's AI fleet over two years



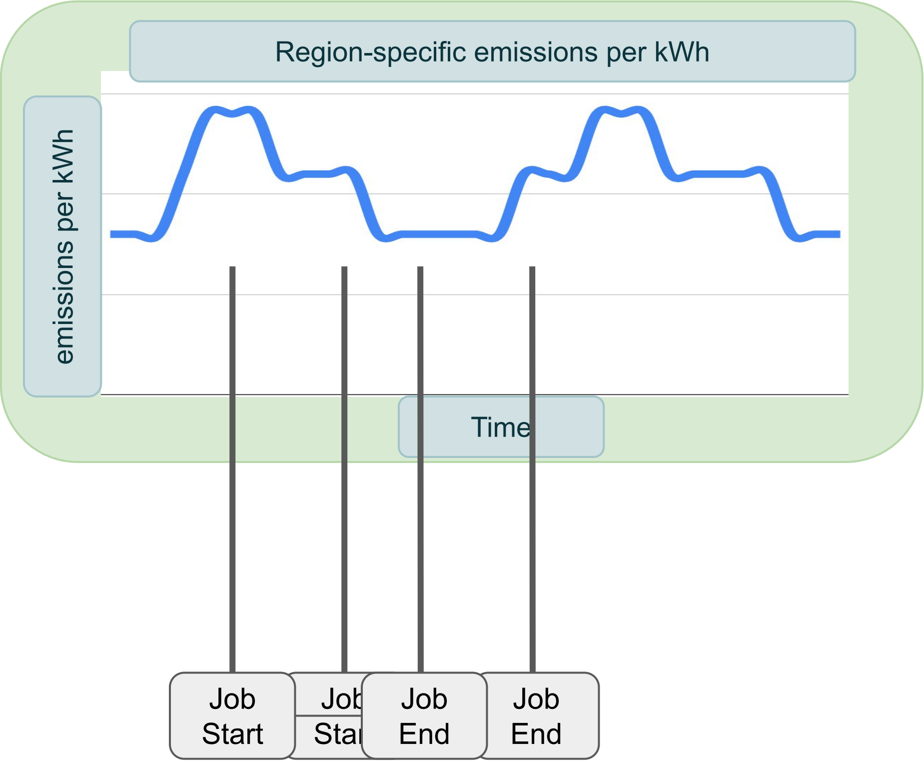
ML power consumption fell by ~20% every 6 months, thanks to iterative optimisation in software and hardware

Wu et al. (2021). Sustainable AI: Environmental Implications, Challenges and Opportunities. https://arxiv.org/abs/2111.00364.

GHG intensity of grids varies significantly across the world



Strategies: Flexible start



Flexible Start: Start the workload at the time in the next N hours s.t. the emissions after the run are minimized.

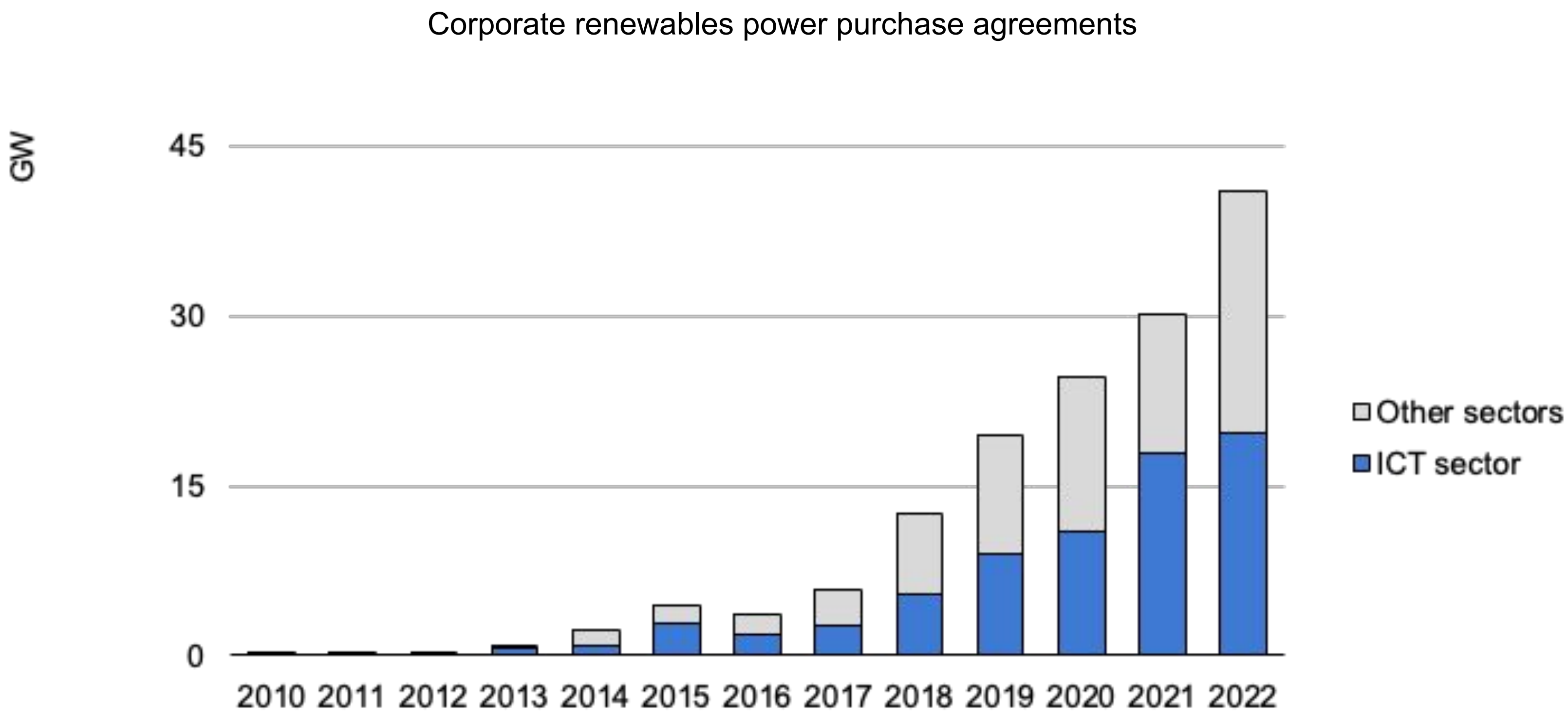
Flexible start over 24 hour period

~25% fewer emissions for DenseNet

~2% fewer emissions for 6B Transformer

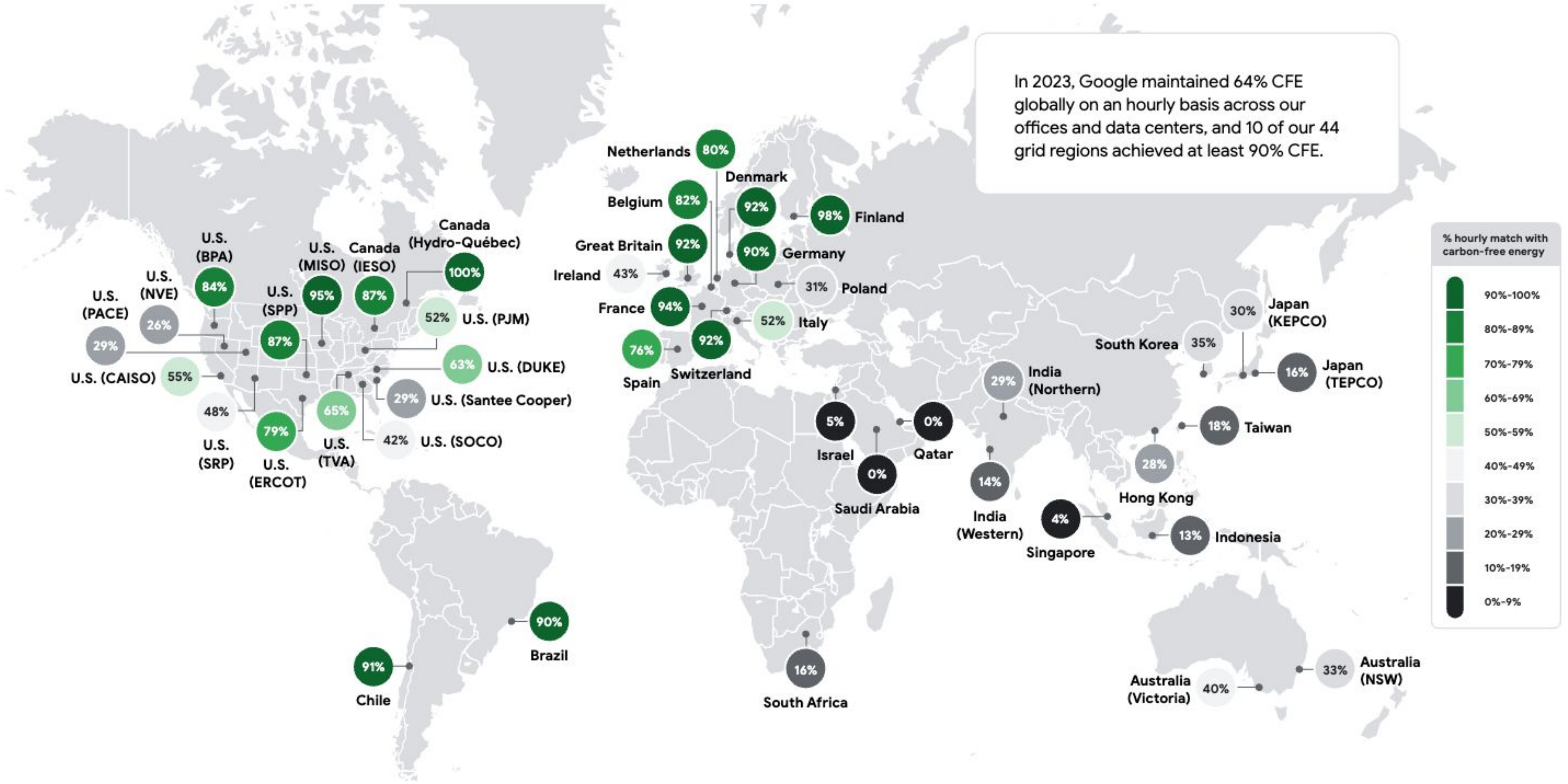
Luccioni (2022), Climate Change AI Summer School. Dodge et al. (2022). Measuring the carbon intensity of AI in cloud instances.

Renewable energy procurement



Source: BloombergNEF (2023), 1H 2023 Corporate Energy Market Outlook.
ICT sectors include Technology and Telecommunications sectors.

Google CFE percentage in every grid region in which we have data center operations, including third-party-operated facilities



Source: Google (2024). Environmental Report.

Coal, gas, and nuclear?

Bloomberg UK Subscribe

The AI Race: Why It's So Expensive | Chip Arms Race | Global Energy Strain | DOJ Scrutiny | How Cha

Businessweek | Economics

AI Needs So Much Power That Old Coal Plants Are Sticking Around

Power companies are scrambling to satisfy the needs of data centers and new factories in a country where the grid is already strained.

By Saijel Kishan and Josh Saul
January 25, 2024 at 12:00 PM GMT+1
Updated on January 25, 2024 at 8:28 PM GMT+1

Bloomberg UK Subscribe

Nuclear Energy Revival: What's Boosting Nuclear | Why Fusion Is So Hard | Uranium's Moment | Labor S

Wealth | Billionaires

Altman's $3.7 Billion Fusion Startup Leaves Scientists Puzzled

With significant backing from OpenAI's billionaire CEO, Helion Energy promises to deliver a fusion power plant by 2028. Its results so far remain mysterious.

By Annie Massa
July 18, 2024 at 2:00 PM GMT+2

CNBC

ENERGY

AI could drive a natural gas boom as power companies face surging electricity demand

PUBLISHED SUN, MAY 5 2024·6:53 AM EDT UPDATED SUN, MAY 5 2024·12:00 PM EDT

FORTUNE

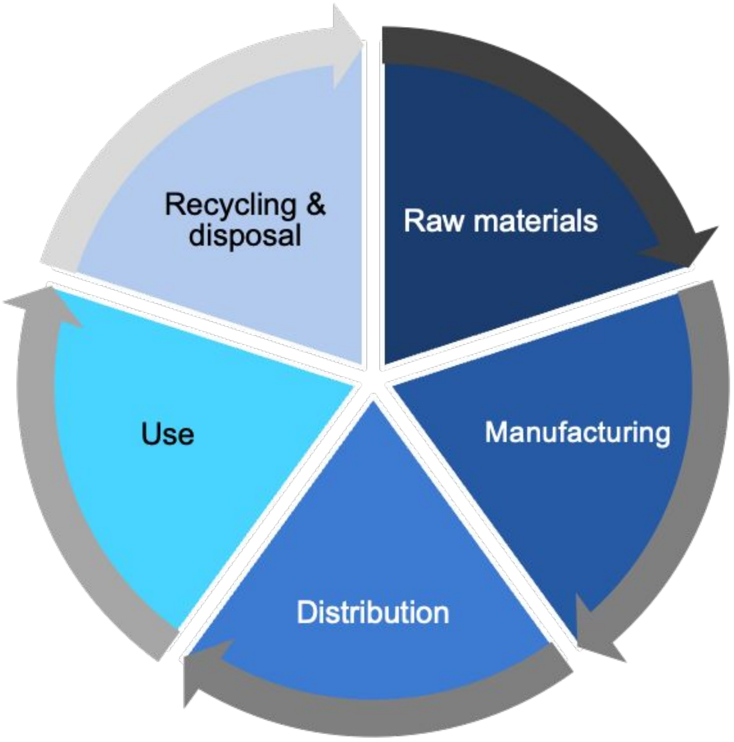
Home News Tech Finance Leadership Well Recommends Fortune 500

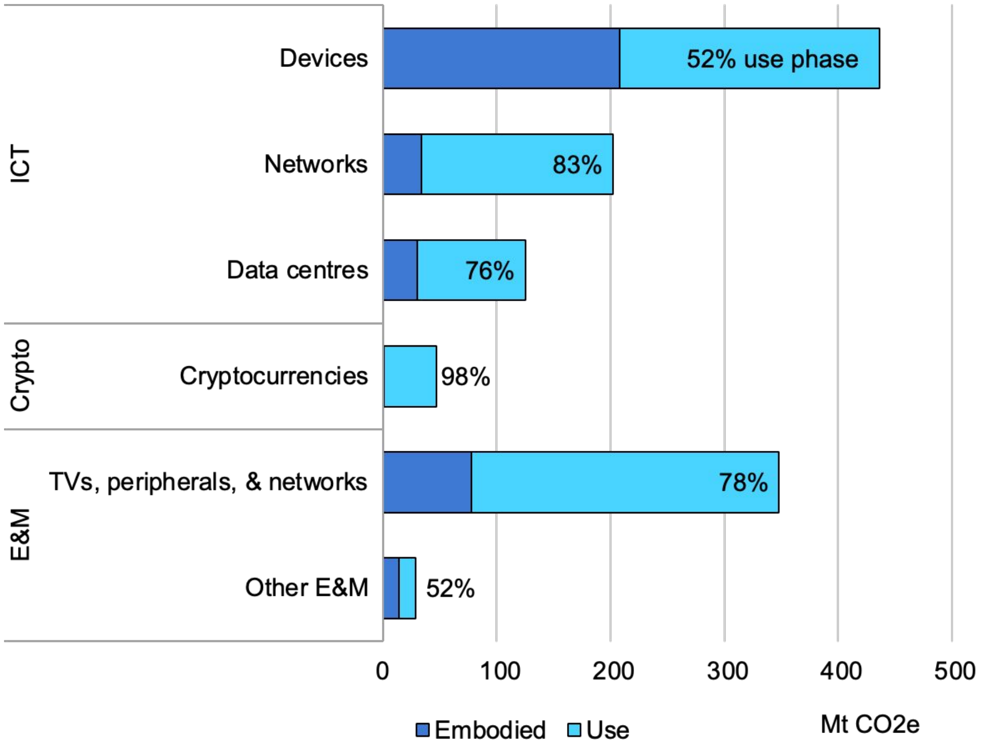
TECH· NUCLEAR POWER

Tech companies are turning to nuclear plants as AI increases demand for power

BY CHRIS MORRIS
July 1, 2024 at 5:40 PM GMT+2

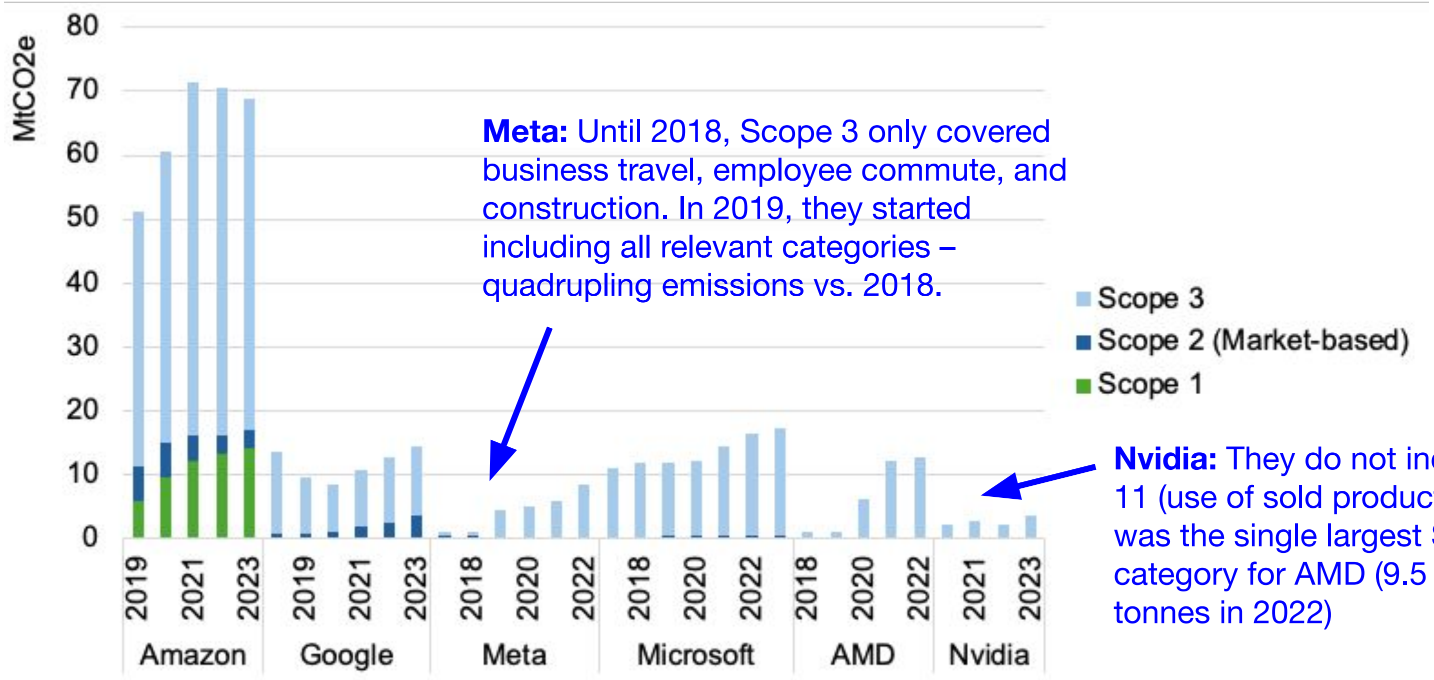
Environmental impacts throughout the hardware lifecycle





Malmodin & Lundén (2018), The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015.

Scope 3 emissions



Meta: Until 2018, Scope 3 only covered business travel, employee commute, and construction. In 2019, they started including all relevant categories – quadrupling emissions vs. 2018.

Nvidia: They do not include Cat. 11 (use of sold products), which was the single largest Scope 3 category for AMD (9.5 million tonnes in 2022)

Exercise caution when comparing Scope 3 emissions, since coverage (i.e. exclusions), data sources, and methodologies can all change

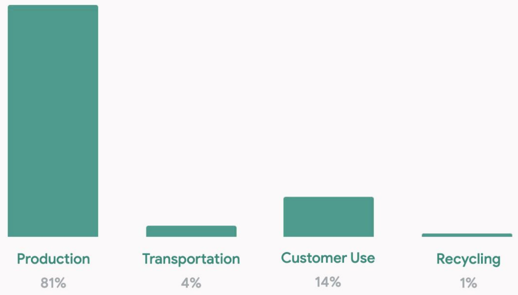
Source: Company environment / sustainability reports and CDP disclosures.

AI-related emissions from devices

Apple Intelligence is designed to protect your privacy at every step. It's integrated into the core of your iPhone, iPad, and Mac through on-device processing. So it's aware of your personal information without collecting your personal information. And with groundbreaking Private Cloud Compute, Apple Intelligence can draw on larger server-based models, running on Apple silicon, to handle more complex requests for you while protecting your privacy.

Estimated GHG emissions for Pixel 8 Pro

assuming three years of use:[6] 79 kg CO_2e



Tensor SoC (System-on-Chip) is the processor developed by the Google Research Team. It was conceptualised in 2016 and first integrated on Pixel 6 in 2021. Now Pixel phones use an AI processor and machine learning powered by Tensor to perform increasingly advanced tasks, like instantly translating messages and videos[3] without internet.[2]

Energy efficiency

The Pixel 8 Pro incorporates power-management software to maximize battery-charging efficiency and extend battery life during use.

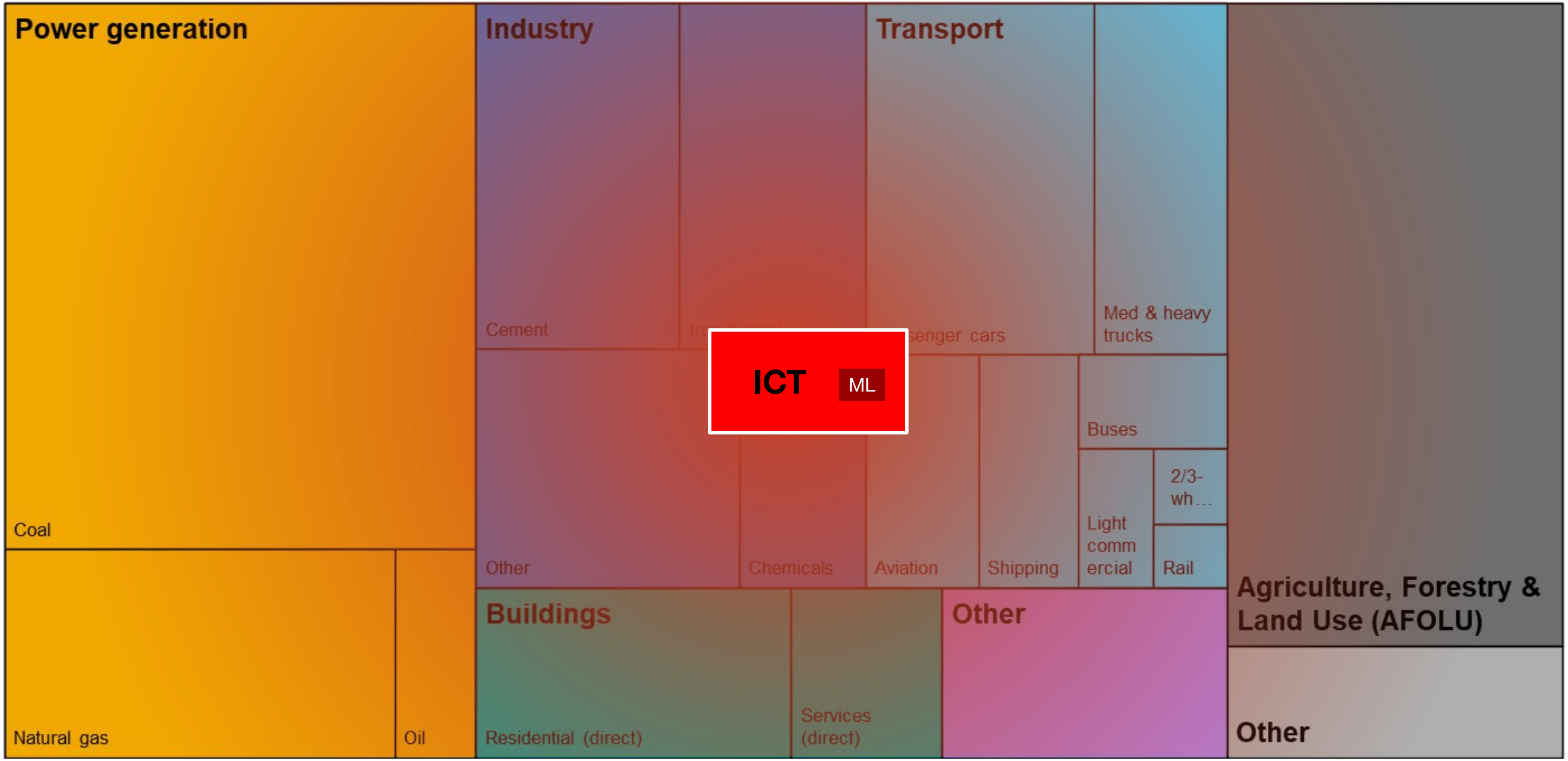
Energy efficiency of Pixel 8 Pro

	115 V, 60 Hz	230 V, 50 Hz
Standby (battery maintenance mode) power[7]	0.19 W	0.20 W
Annual energy use estimate[8]	9 kWh	9 kWh
Annual cost of energy estimate	US$1.45[9]	€2.56[10]

Key takeaways

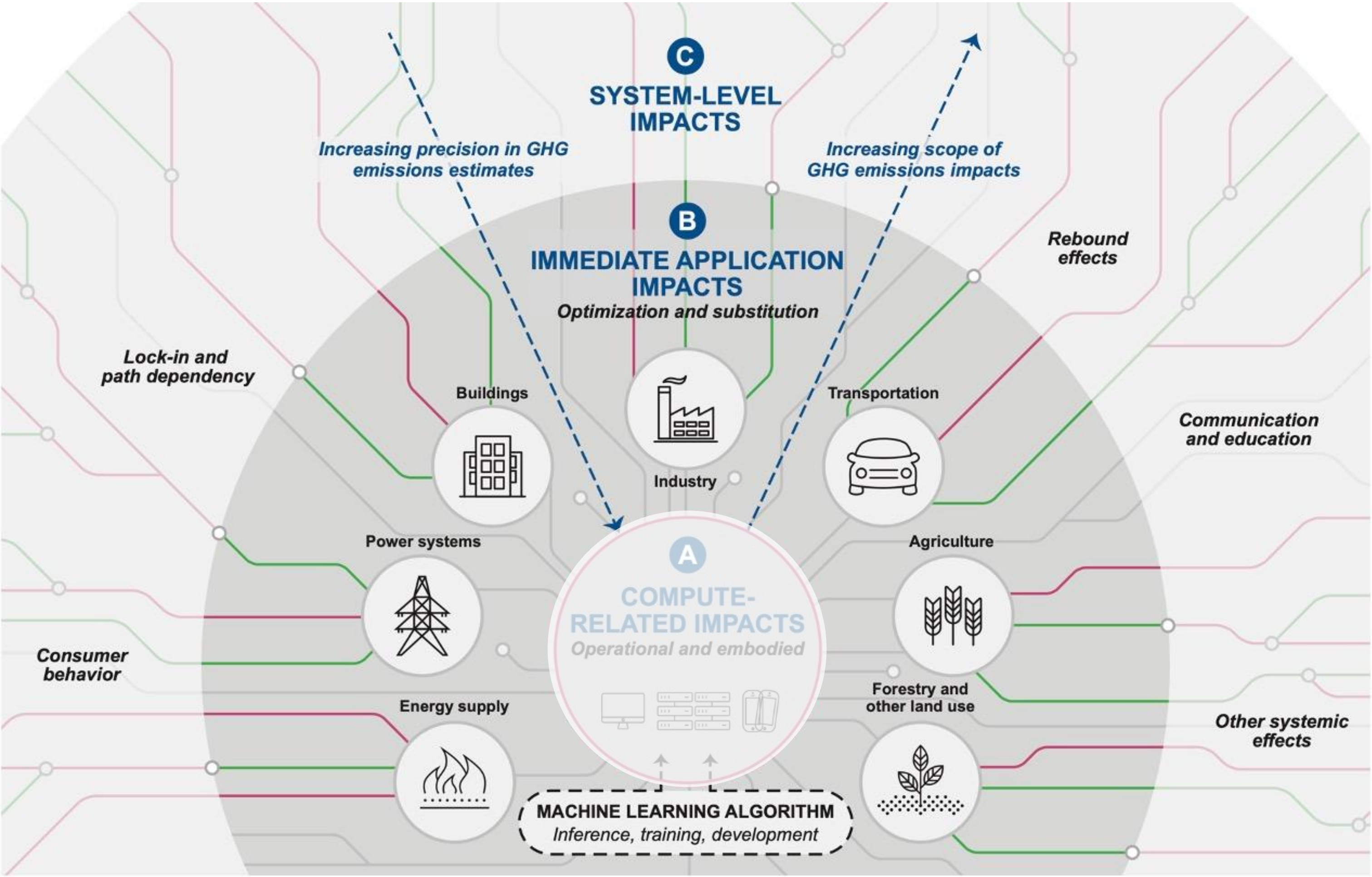
- Overall, we still have a poor understanding of total ML-related energy use due to a lack of data
- ML-related energy use is probably 20-40 TWh in 2023 (mostly in data centres + some distributed compute), with inference accounting for more than half
- Demand for ML workloads is increasingly rapidly, and is a factor in driving growth in energy demand for hyperscalers (+20-30% per year)
- Curbing energy demand growth and reducing emissions from ML requires further policy and technological progress on:
    - energy efficiency in software, hardware, and infrastructure (incl. new tech / innovation)
    - purchasing renewable electricity, decarbonising electricity grids, and adopting more flexible operations (e.g. time, location) (Scope 2)
    - reducing supply chain emissions (Scope 3) and other environmental impacts

Greenhouse gas emissions come from many sectors and sources



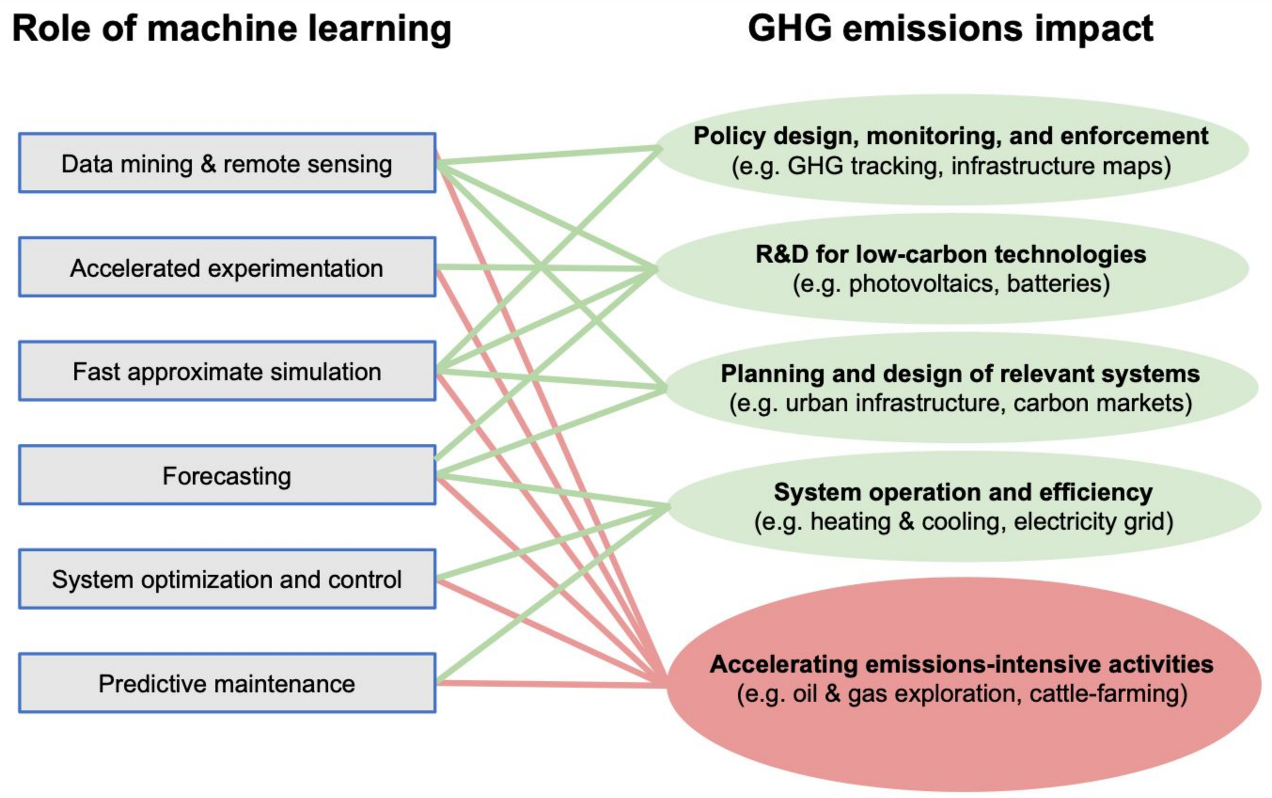
Emission impacts from applying ML

Lynn Kaack



Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1-10.

Immediate application impacts



Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1-10.

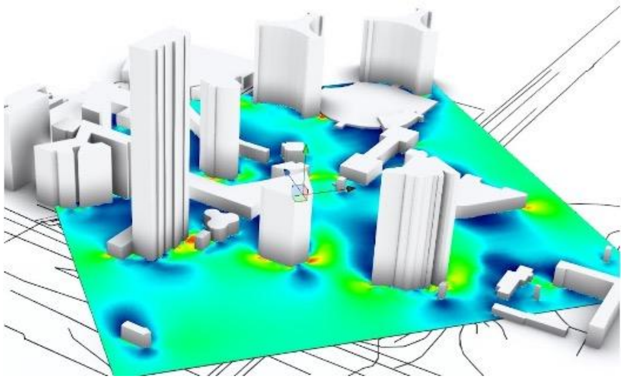
Examples of immediate application impacts

Data center cooling efficiency with ML-based control (Google DeepMind)

Industry metric for cooling energy efficiency: kW/ton (or energy input per ton of cooling achieved)

Improving urban design by speeding up models of the urban microclimate with ML (InFraReD of the AIT)

Qualitative impact on design and sustainability that is more difficult to measure



Broader scope of application impacts

Example: Data center cooling efficiency with ML-based control

- ML-based efficient cooling **reduces energy consumption and GHG emissions**
- The application is also **reducing energy costs**
- Computing may be offered at **lower cost**
- Lower costs may increase demand → **higher energy consumption**
- “**Rebound effect**” with system-level impact

ML's carbon footprint

ML applications in climate change mitigation

Emissions from ML computation & hardware

ML's system-level impacts

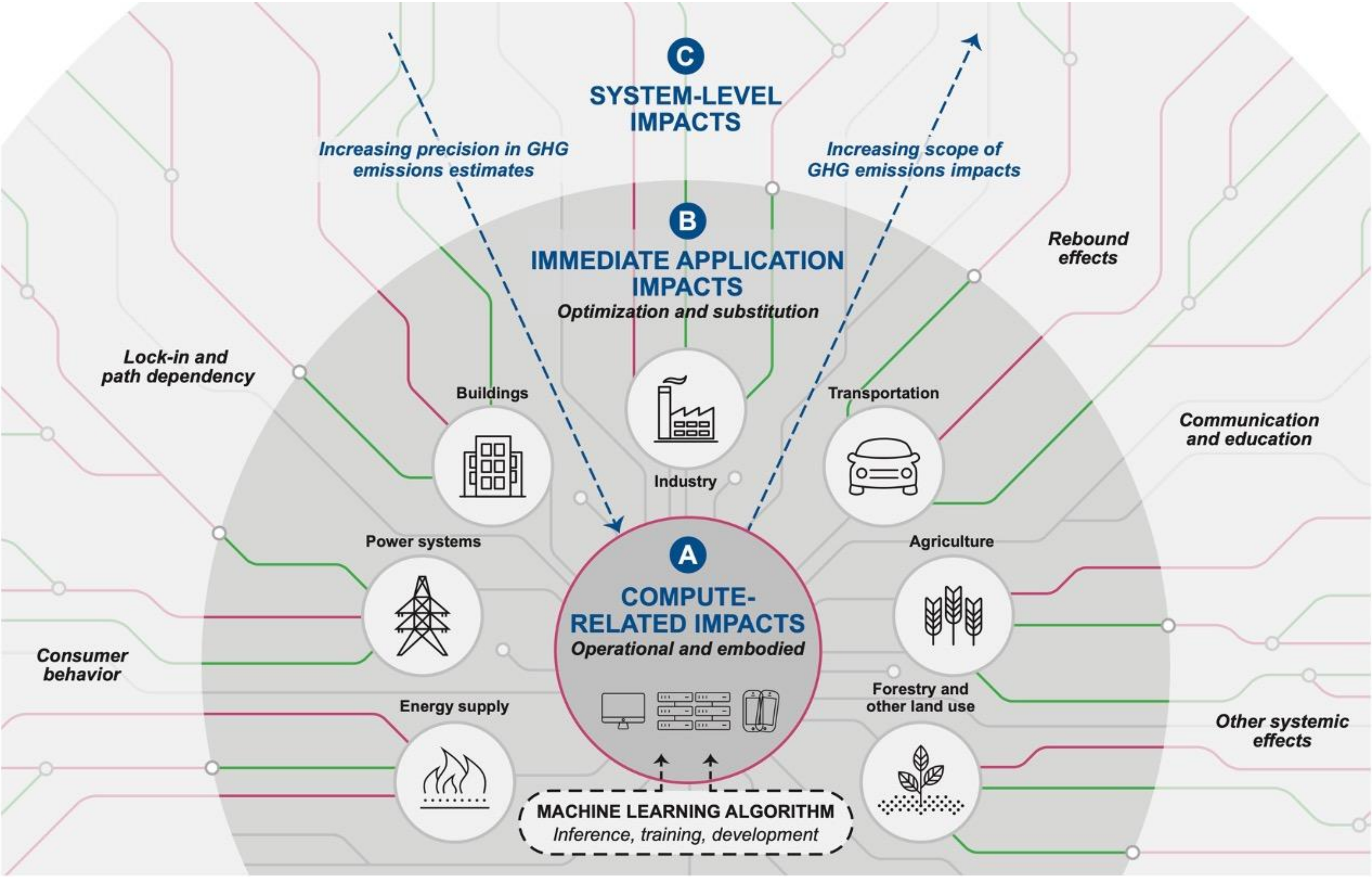
ML applications that increase emissions

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1-10.

60

System-level impacts of ML

Rebound effects	Reducing energy consumption reduces costs → money saved may be used and cause more emissions *Example: ML for optimizing systems*
Lock-in and path dependency	Technologies compete and dominate → lock-in to suboptimal technologies hampering decarbonization *Example: Autonomous driving and car use*
Consumer behavior	Trends and advertising may change consumption patterns → embodied emissions in those products *Example: ML in advertising and social media*
Communication and education	Societal support for climate action essential *Example: ML on social media*



Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1-10.

Why do impact assessment?

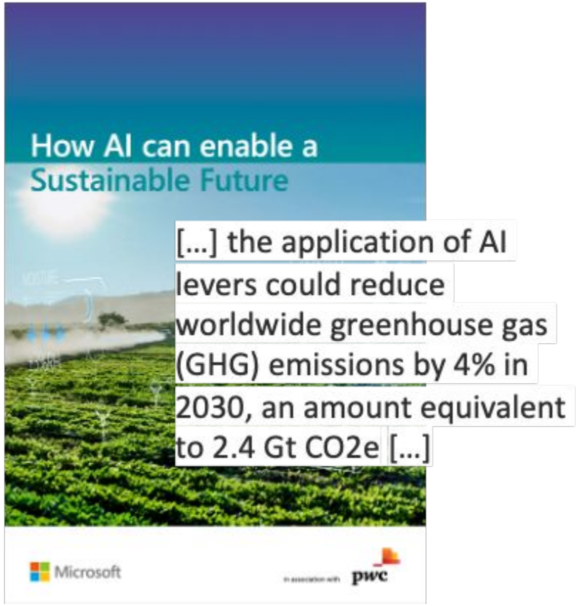
Goal 1: Assess costs and benefits of new ML-driven projects a priori

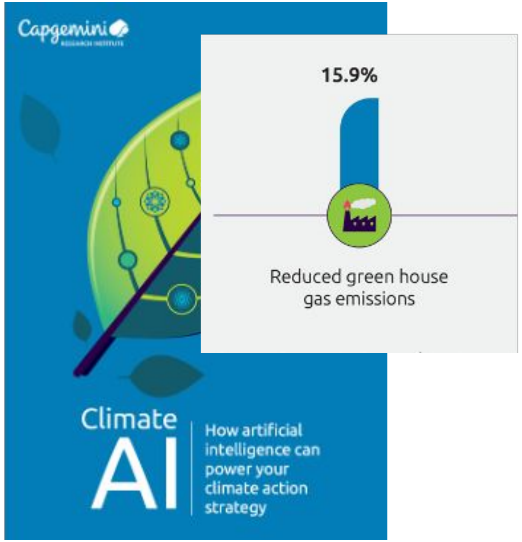
- ML-based approaches are potentially **risky and expensive**
- Problem: Lack of established methodology
- Need for case studies and generalizing beyond

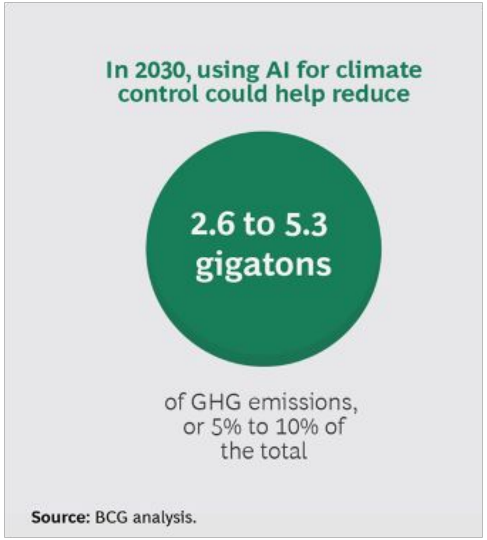
Goal 2: Evaluate impacts and dynamics of ML as a field

- ML is a **fast-growing enabling technology** that has the potential to affect present and future societal and technological trajectories
- Problem: Energy and climate models do not even account for digitalization
- Need to understand possible large-scale effects of the technology

Total application-related emissions of ML







- Estimates from consulting and tech firms
- Based on undisclosed modeling or extrapolation of stakeholder interviews
- Some sector-specific modeling by EIA (autonomous vehicles) and IEA (buildings)

Approaches to shape impacts

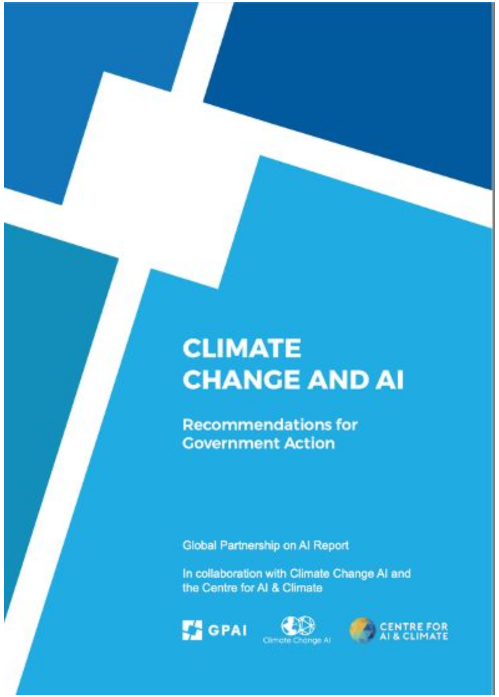
Check out the recording of Day 8 (Prof. Philipp Hacker)!

Strategies for application-related/indirect impacts

- **Quantify and evaluate** the impacts where possible
- **Be transparent** about impacts in publications and with stakeholders (quantitatively and qualitatively!)
- **Choose what** you work on, bring climate considerations into how you build an application
- **All ML applications** may potentially have an effect on the climate (e.g. recommender systems)
- Initiate **company-wide policies** such as internal carbon pricing or conditions on the projects or products realized
- Don’t forget about **other social and other environmental impacts**

Regulatory approaches

- "General" climate policies and actions are important, e.g. carbon pricing
- Specific approaches to align ML with climate change goals:
    - incentivizing applications that help address climate change
    - requiring transparency and accountability in cases where ML could increase emissions (computing and applications)
    - conducting climate-cognizant technology assessment within AI strategies and when regulating ML-driven emerging technologies



Report by GPAI, Climate Change AI and Center for AI & Climate

Existing initiatives to foster applications

- Dedicated governmental research and development funding (EU, Germany, Austria, Sweden, US, etc.)
- Large research and monitoring programs (Copernicus, Nasa Harvest, etc.)
- Some AI strategies and EU AI Act (2024) emphasize climate change as application area
- US Executive Order on AI (2023)

Administration Priori

OCTOBER 30, 2023

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

BRIEFING ROOM › PRESIDENTIAL ACTIONS

(g) Within 180 days of the date of this order, to support the goal of strengthening our Nation's resilience against climate change impacts and building an equitable clean energy economy for the future, the Secretary of Energy […] shall:

(i) issue a public report describing the potential for AI to improve planning, permitting, investment, and operations for electric grid infrastructure […]

(ii) develop tools that facilitate building foundation models useful for basic and applied science, including models that streamline permitting and environmental reviews while improving environmental and social outcomes; […]

(iv) take steps to […] utilize the Department of Energy's computing capabilities and AI testbeds to build foundation models that support new applications in science and energy […]

Existing initiatives on requiring transparency

- The EU AI Act (2024)
  - Reporting on the compute resources and energy use of high-risk systems and large models (>10^25 FLOPs for training)
- US Executive Order on AI (2023)
  - Reporting requirements for large models (>10^26 FLOPs for training) (only security!)
- Proposed US Artificial Intelligence Environmental Impacts Act (2024)
  - Require environmental impacts assessments (by the EPA and NIST)
  - Application- and computing-related impacts

Official Journal
of the European Union

EN
L series

2024/1689

12.7.2024

REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 13 June 2024

laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

(Text with EEA relevance)

Annex IV: Technical Documentation
[for high-risk AI systems]
(c) […] the computational resources used to develop, train, test and validate the AI system;

Annex XI: Section 1 – Information to be provided by all providers of general-purpose AI models
(d) the computational resources used to train the model (e.g. number of floating point operations), training time, and other relevant details related to the training;
(e) known or estimated energy consumption of the model. With regard to point (e), where the energy consumption of the model is unknown, the energy consumption may be based on information about computational resources used.