

Between Hype and Hazard - wie KI den Klimawandel beeinflusst

INNOVATE. INTEGRATE. EXCEED.



inovex



 mara.hermann@inovex.de

 Mara Hermann

Hallo,

ich bin Mara Hermann

Data & Machine Learning Engineer, Köln

- Mathestudium in Ulm (Bachelor & Master)
- 2014: Praktikum bei inovex -> 2018: Festanstellung in der DMA
- 2019: Mitgründung des Nachhaltigkeits-Teams
- 2023/24: Mitgründung des inoCircles Sustainable Computing

Auswirkungen von KI auf die 3 Säulen der Nachhaltigkeit



Was erwartet euch heute (und was nicht)

Was dieser Vortrag will...

- ... über die Auswirkungen von AI auf die Umwelt (und teilweise auf die Gesellschaft & Wirtschaft) aufklären
- ... die Rolle von Big Tech Firmen beleuchten
- ... zum Nachdenken anregen
- ... möglichst konkrete Ansätze für mehr Nachhaltigkeit im (Projekt-)Alltag geben

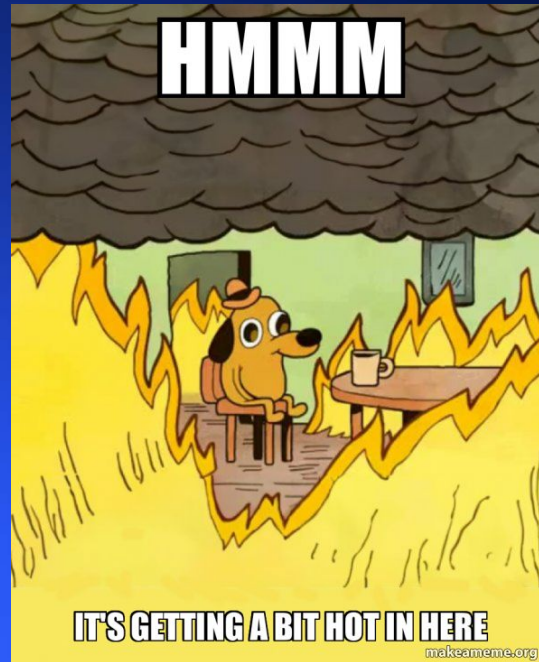
Was er nicht will...

- ... AI “verteufeln”
- ... Patentrezept für “nachhaltige KI” verkaufen

Disclaimer



Warum sollte uns das beschäftigen?



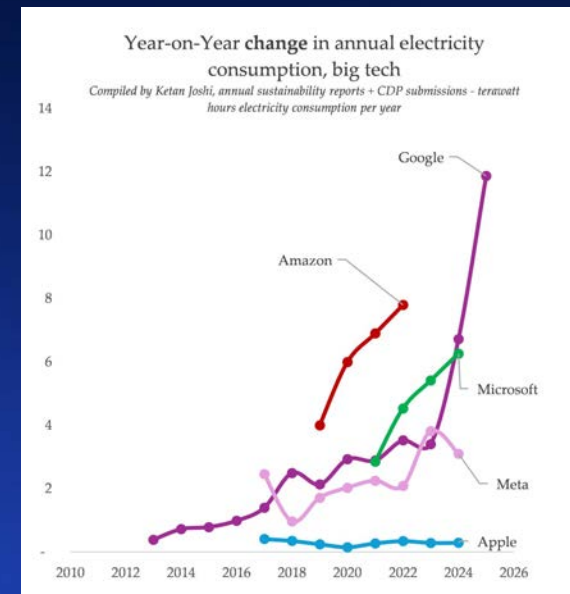
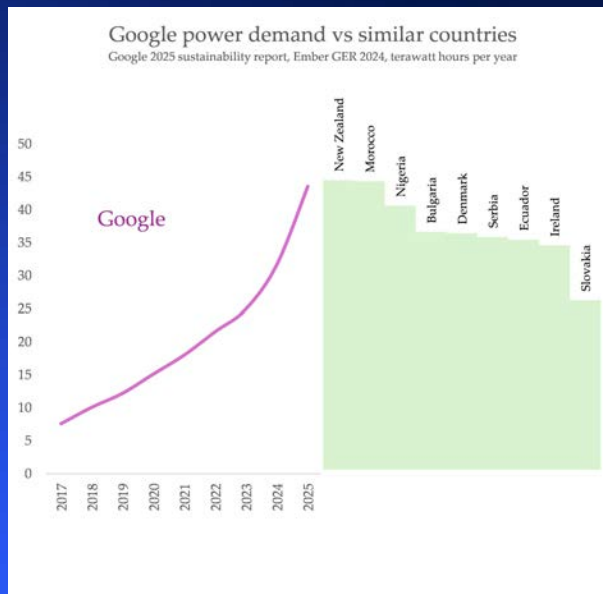
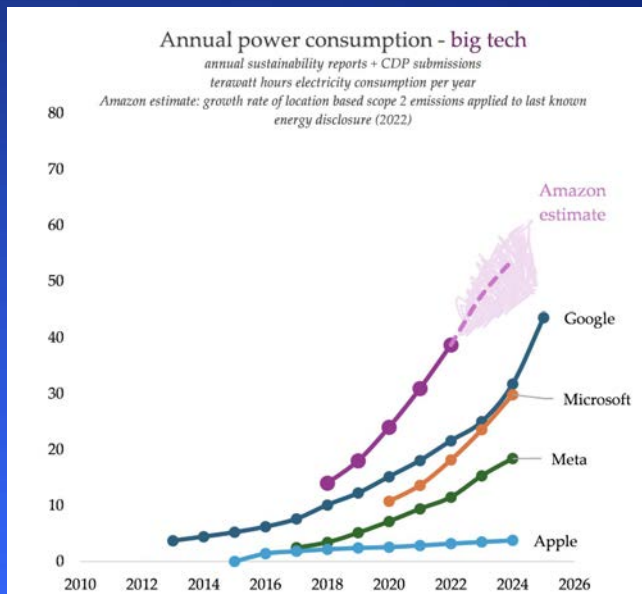


AI Could Soon Offset Its Own Environmental Impact By Improving Energy Efficiency, Report Finds*

Quelle 2025

KI ist “immateriell”

Energieverbräuche der Hyperscaler gehen durch die Decke



Data Center sprießen aus dem Boden



Musk Data Center in Memphis

© *steveheap/Depositphotos*



Meta's 5 GW Hyperion Data Center in Louisiana

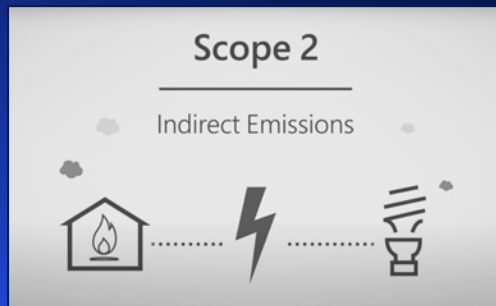
Quelle 1

Quelle 2

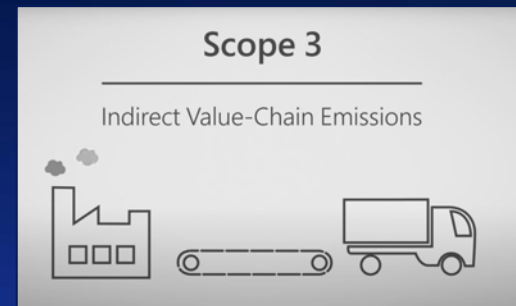
Data Center Emissionen nach GHG-Protokoll-Scopes



Verbrennung von Diesel
zum Betreiben von
Notfallgeneratoren



Energieverbrauch aus dem
Stromnetz



Herstellung und Recycling der
IT Hardware sowie das
Gebäude

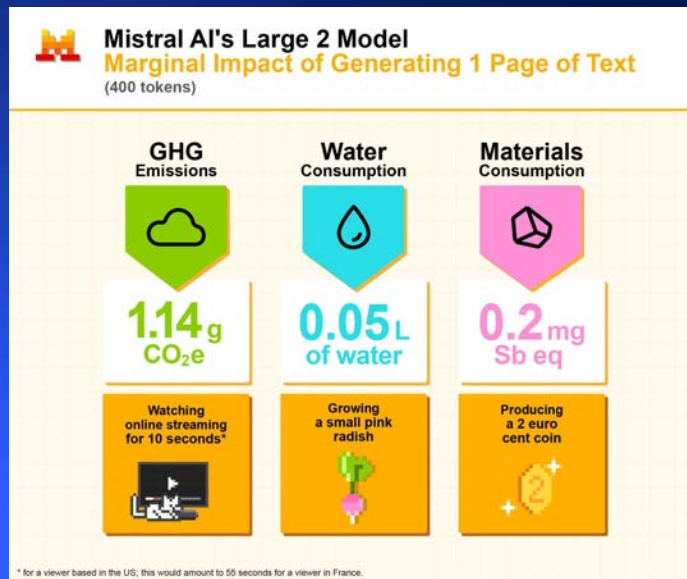
Planetary Justice Impacts along the AI Supply Chain



Wie könnten Data Center nachhaltiger werden?

- Lokalität:
 - globale Verfügbarkeit, aber lokalen Impact -> mehr Verteilung statt geografische Cluster
 - lokale Gegebenheiten mehr beachten, um erneuerbare Energien zu fördern
 - möglicherweise gestützt durch Regulatorik (z.B. Irland: mind. 80% EE/DC)
 -
- Konstruktion
 - Nachhaltigere Materialien wie Holz und low-carbon Beton
 - statt Neubauten bestehende vakante Gebäude mit bestehender Infrastruktur wiederverwenden
 - Abwärme weiterverwenden, um Häuser zu heizen
- KI Nachhaltigkeit
 - als Treiber des Data Center Wachstums
 - mehr Nachhaltigkeit durch
 - Transparenz
 - Konzeption
 - Umsetzung

KI Nachhaltigkeit - Schritt 1: Transparenz



[Quelle Mistral](#)

KI Nachhaltigkeit - Schritt 2: Konzeption

Suffizienz

nur das Notwendige
nutzen &
Verschwendung
vermeiden

Knowledge Distillation

Effizienz

das Bestehende mit
möglichst wenig Einsatz
von Ressourcen, Energie
& Zeit nutzen (mit
möglichst guten
Ergebnissen)

sparsamere Modelle

Konsistenz

das Bestehende nutzen
ohne dabei natürliche
Rohstoffe oder
Kreisläufe zu zerstören

*Carbon-Aware
Computing*

KI Nachhaltigkeit – Schritt 3: Umsetzung beim Modelltraining

- Hyper-Parameter-Optimierung
 - Größere Batch Size
Reduktion der Trainingsdauer bei vernachlässigbarer Erhöhung der GPU-Auslastung
 - Anzahl der Trainingsdaten & Anzahl der Epochen reduzieren
unter Abwägung der Modellqualität
 - Random od. Bayesian Search statt Grid Search
- Early Stopping
 - Reduktion der Trainingsdauer
- Mixed Precision (unterschiedliche Genauigkeit der Gewichte)
 - Reduktion der Trainingsdauer & Verarbeitung größerer Batches
- Kompakte Modelle
 - durch Quantisierung oder Pruning
 - unter Abwägung der Modellqualität

KI Nachhaltigkeit – Schritt 3: Umsetzung bei der Nutzung von Coding Agents

- Fragen mit gezielten Anweisungen stellen
 - (z. B. „Optimiere für geringe Laufzeit“ oder „Minimiere den Speicherverbrauch“)
- Coding Assistants nicht dauerhaft im Hintergrund laufen lassen
 - nur ein Bruchteil der generierten der automatisierten Requests schaffen es in den finalen Code
- kurze Antworten fordern
- dauerhafte Anweisungen zur Nachhaltigkeit konfigurieren
 - z.B. instructions Datei bei GitHub CoPilot
- Refactoring nicht vergessen!
 - GitHub hat Rückgang von Code Wiederverwendung festgestellt

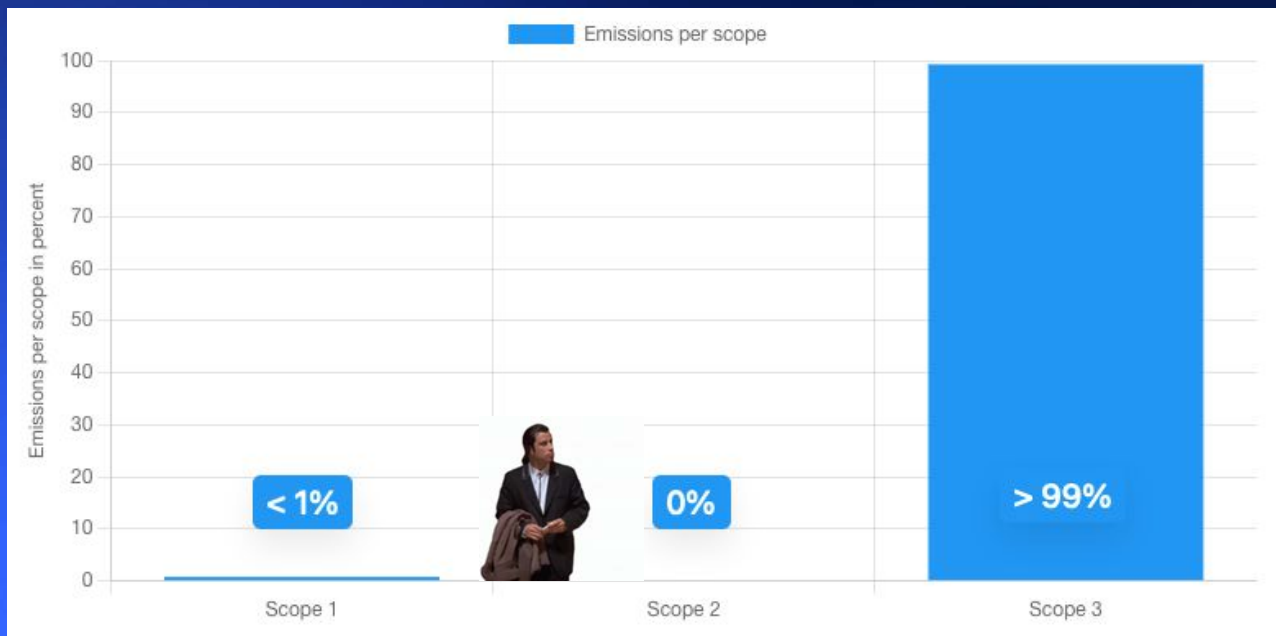
KI Nachhaltigkeit – Schritt 3: Umsetzung mittels Model Selection

- Modelle variieren stark bzgl der Emissionen in der Inferenz
- kleinere Modelle verbrauchen weniger
-> für kleinere Aufgaben geeignet
- LLM Calculators können dabei helfen, ein besseres Gefühl für die verbrauchten Ressourcen zu bekommen

(Mini) Demo
Time

Der Strom für KI kommt aus erneuerbaren
Energien

Aufteilung der Cloud Emissionen nach Scopes anhand eines Real-Beispiels



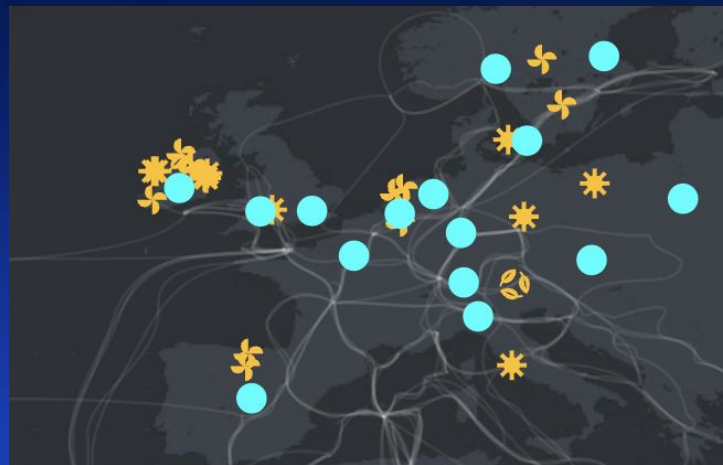
“While the path to achieving our climate ambitions will not be linear—given our AI infrastructure buildout is currently accelerating faster than the grid is decarbonizing—we remain focused on scaling abundant and affordable clean power globally and progressing technological innovations that drive down emissions across our operations and the broader industry”

Google 2025 Environmental Report

0% Scope 2 Emissionen != 100 % erneuerbare Energien

- Durch Nachhaltigkeits-Projekte bezieht u.A. Microsoft Renewable Energy Certificates” (REC) & “Power Purchase Agreements” (PPA)
- GHG Protokoll erlaubt es, Emissionen damit zu offsetten
- Probleme
 - Zu grobe Grenzen: EACs von Schweden können Energieverbrauch in Spanien kompensieren, weil gleicher Strommarkt
 - Falsche Incentivierung: EACs werden häufig dort gekauft, wo erneuerbare Energien bereits verbreitet sind, anstatt lokale Netze zu dekarbonisieren
 - Nicht-brauchbare Daten: Kunden können eigene Emissionen nicht optimieren, weil sie sie nicht kennen

Microsoft Data Center Exploration



= Data Center



= Nachhaltigkeits-Projekte

Wie berechnet man die tatsächlichen (ortsbasierten) Emissionen?

Energieverbrauch

eines Data Centers an einem bestimmten Zeitpunkt

x

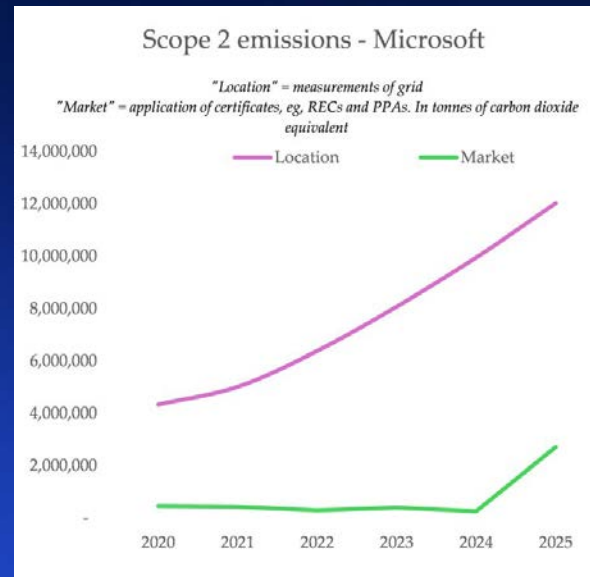
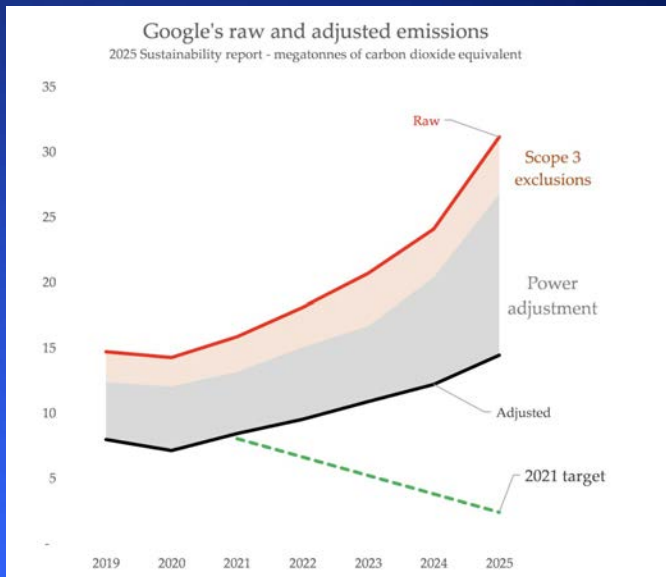
CO2-Intensität im Stromnetz

eines Data Centers zu diesem Zeitpunkt



Electricity Maps

Location based vs. marked based emissions



Data Collection & Visualization by Ketan Joshi

Quelle

KI wird durch fossile Energien betrieben und/oder begünstigt den Ausbau fossiler Energien

- Höherer Stromverbrauch als durch erneuerbare allein gedeckt werden könnte
 - Höherer Bedarf an fossilen Energien im bestehenden Netz
 - Ausbau, Re-Aktivierung und Verlängerung fossiler Kraftwerke
- “Air-booking”: Netzkapazität (fossil und erneuerbar) wird für KI reserviert, auch wenn sie ggf. nicht genutzt wird
- “Behind-the-meter”: Stromerzeugung direkt am Data Center oder in unmittelbarer Nähe (aus fossilen und erneuerbaren Energien)
- Nutzung von KI zum Ausbau von fossilen Energien durch schnellere Exploration & Extraktion

Carbon-Aware Computing

- Workloads in Zeitfenstern mit geringerer CO2-Intensität laufen lassen (Time Shifting)
- Regionen mit hohem Anteil erneuerbarer Energien wählen

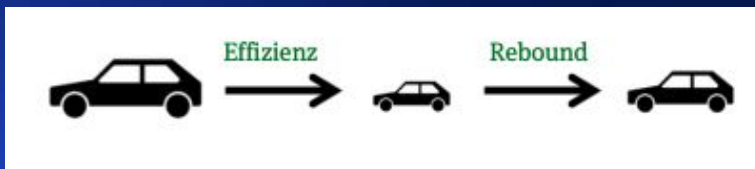
What Azure region to use to save costs and emissions?

Region	West Europe	Germany West	North Europe	France Central	Sweden Central
Costs (for B2s v2 VM, 02/25)	67.26€ / month	67.26€ / month	63.89€ / month	66.13€ / month	60.53€ / month
Scope 2 emissions (in 2024)	290g CO2e/kWh	332g CO2e/kWh	401g CO2e/kWh	31g CO2e/kWh	19g CO2e/kWh
Scope 3 emissions (in 2024)	12.12g CO2e/€	not reported*	11.64g CO2e/€	7.22g CO2e/€	2.66g CO2e/€
Capacity issues? (in our experience)	Yes	unknown	Yes	No	unknown

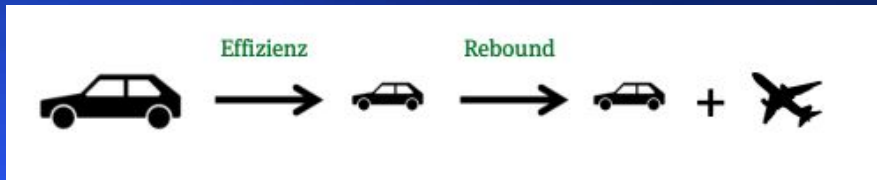
Effizientere KI ist Ressourcen schonender

Effizienzsteigerung führt nicht automatisch zu Einsparungen

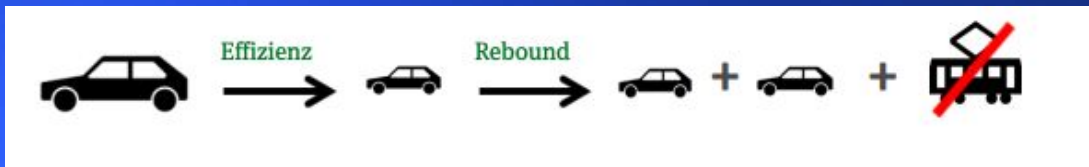
direkt



indirekt



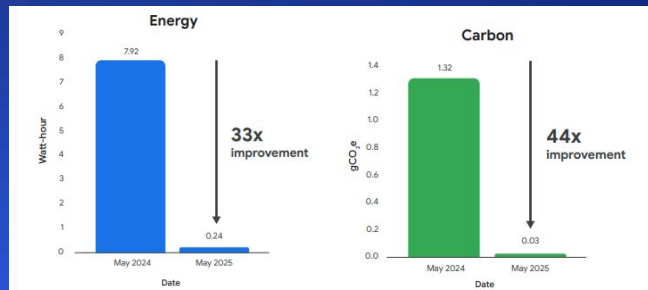
gesamt-
gesellschaftlich



Dieses Phänomen nennt man Rebound-Effekt oder (insb. bei Rohstoffen) auch Jevon's Paradox

Ablenkung & Verharmlosung durch Selektion

Google Sustainability Report 2025

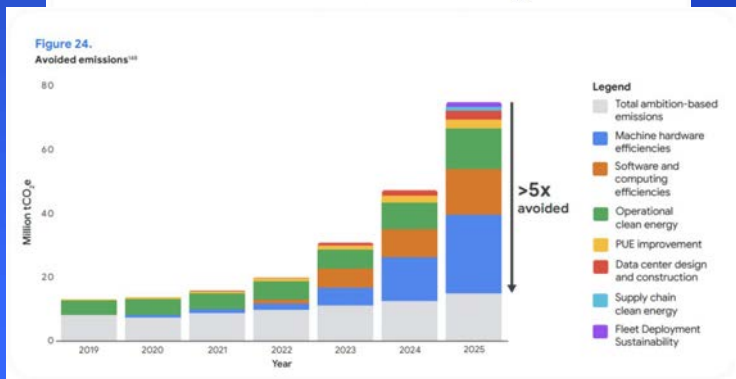


Google: 1 Chatbot Query = 9 Sekunden TV

Vergleich zu Gesamt-Energie-Verbrauch:

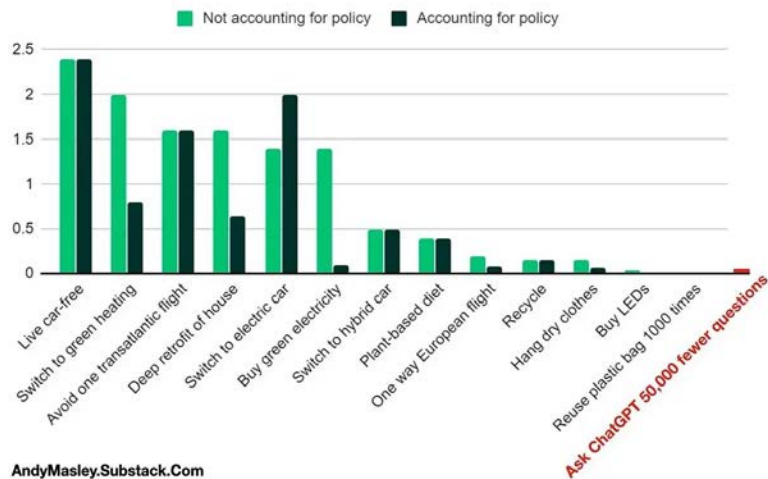
- AI 15%: 74,6 billion prompts/day (Vgl. Chat GPT: 2.5 billion prompts/day)
- AI 25% 124 billion prompts/day (JEDER Mensch 15 Gemini prompts/day)

-> Wo kommen all die Emissionen her?



Ablenkung & Verharmlosung durch Selektion

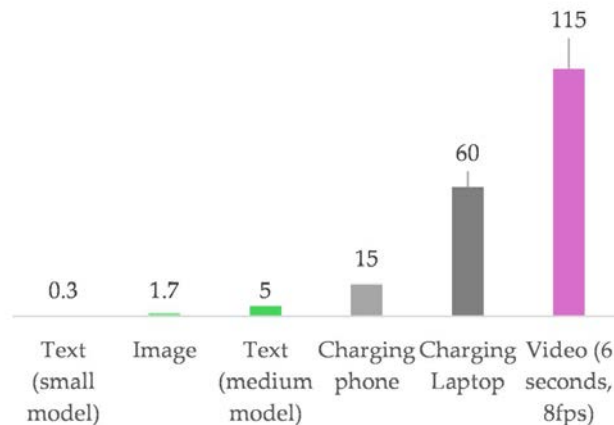
Figure 4. Tonnes of CO₂ avoided by selected personal lifestyle decisions accounting for government policy



AndyMasley.Substack.Com
Original climate graph taken from Founders Pledge

Quelle Andy Masley

IEA Inference electricity consumption estimates - various media generation tasks
Watt hours,

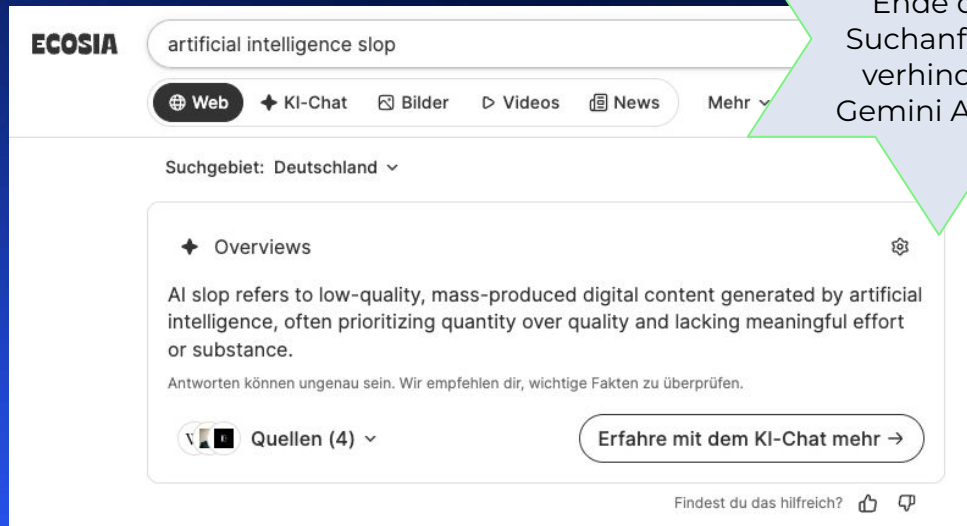


Quelle Ketan Joshi

The struggle is real!



Quelle



Tipp: **-ai** am
Ende der
Suchanfrage
verhindert
Gemini Aufruf

Auch im Projekt-Alltag macht uns der durch KI produzierte Content (zumindest aktuell) nicht unbedingt effizienter.

Die unbequeme Antwort



Durch KI können Ressourcen eingespart werden

“As we further integrate AI into our products, reducing emissions may be challenging due to increasing energy demands from the greater intensity of AI compute.”

Google 2024 Environmental Report

“While we remain deeply committed to sustainability, reaching our climate moonshot is getting harder. Growing our data center footprint to build out the infrastructure needed to make AI as helpful as possible to everyone requires energy and resources”

Google 2025 Environmental Report

AI != AI

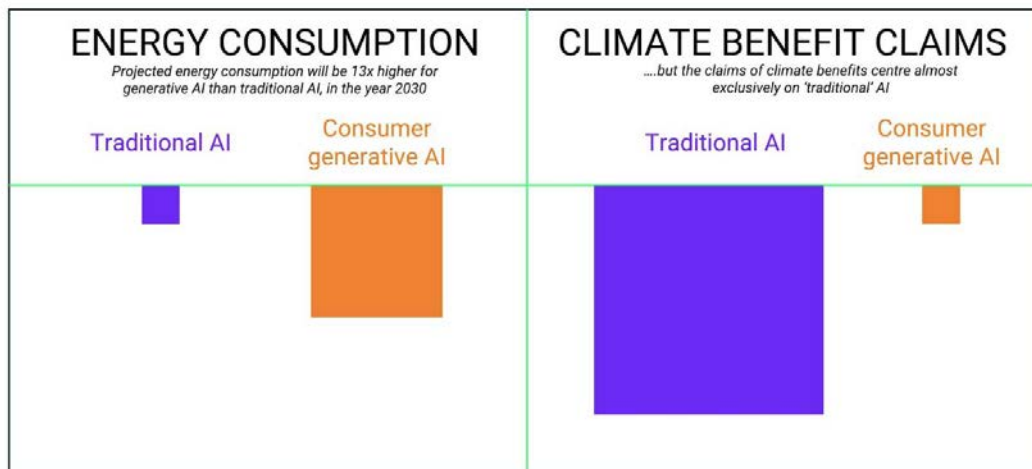


FIGURE 1 - CLIMATE BENEFIT versus DAMAGE

This graphic illustrates how most climate benefit claims relate to traditional AI, whereas most harm will relate to consumer generative AI. The first compares the projected electricity consumption in a central 2030 scenario, also represented by the area of each square. The second comparison shows the number of claims by AI type, represented by the area of each square.

Predatory Delay

- Negativer Impact von AI in der Gegenwart wird in Kauf genommen für mögliche Benefits in der Zukunft
- Notwendige Systemänderungen werden aufgeschoben
- Technischer Optimismus statt Umdenken, Regulatorik & Veränderung
- Sehr interessanter Talk von Dr. Anne Mollen auf der EcoCompute 2025
[\[Link\]](#)

Rebound Effekt



Es kommt darauf an, was man tut & wie man es tut...

Tackling Climate Change with Machine Learning

David Rolnick^{1*}, Priya L. Donti², Lynn H. Kaack³, Kelly Kochanski⁴, Alexandre Lacoste⁵, Kris Sankaran^{6,7}, Andrew Slavov⁸, Nikola Milojevic-Dupont^{10,11}, Natasha Jaques¹², Anna Waldman-Brown¹², Alexandra Luccioni^{6,7}, Tegan Maharaj^{6,8}, Evan D. Sherwin², S. Karthik Mukkavilli^{6,7}, Konrad P. Körding¹, Carla Gomes¹³, Andrew Y. Ng¹⁴, Demis Hassabis¹⁵, John C. Platt¹⁶, Felix Creutzig^{10,11}, Jennifer Chayes¹⁷, Yoshua Bengio^{6,7}

¹University of Pennsylvania, ²Carnegie Mellon University, ³ETH Zürich, ⁴University of Colorado Boulder,

⁵Element AI, ⁶Mila, ⁷Université de Montréal, ⁸École Polytechnique de Montréal, ⁹Harvard University,

¹⁰Mercator Research Institute on Global Commons and Climate Change, ¹¹Technische Universität Berlin,

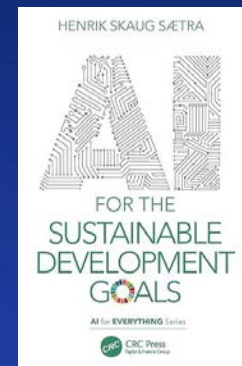
¹²Massachusetts Institute of Technology, ¹³Cornell University, ¹⁴Stanford University,

¹⁵DeepMind, ¹⁶Google AI, ¹⁷Microsoft Research

“The SDGs allow us to consider the impact of AI on our lives and environments, this requires us to understand the context of AI - the infrastructure it is built on, who develops it, who owns it, who has access to it, who uses it, and what it is used for - rather than relying on an isolationist theory of technology.”

Climate Change AI

AI For The Sustainable Development Goals



Key Take Aways

- In jedem Schritt der AI Supply Chain gibt es (negative) Auswirkungen auf die Umwelt, die Gesellschaft und Wirtschaft.
- Nicht alle dieser Schritte sind von uns direkt beeinflussbar, aber wir sollten dennoch darüber Bescheid wissen.
- Direkte Hebel sind Carbon-Aware-Computing, Suffizienz, bewusste Modellwahl und andere Techniken, um mehr KI-Nachhaltigkeit zu erzielen.
- Nur die Effizienz eines Modells zu verbessern, reicht jedoch nicht unbedingt aus. Rebound Effekte können Effizienzgewinne in & durch KI zunichte machen.
- AI ist nicht gleich AI. Generative KI mit traditionellen ML/DL Modellen gleich zu stellen ist wie von Fahrrädern auf Raketen zu schließen (oder umgekehrt).



Diese Folien sind nicht KI-generiert und sind ohne
bewussten Einsatz von KI entstanden

Vielen Dank!

Mara Hermann

Data & Machine Learning Engineer



Mara Hermann



mara.hermann@inovex.de

