UNIVERSITÄT
ZU KÖLN

Climate Change

Nils Reiter,
nils.reiter@uni-koeln.de

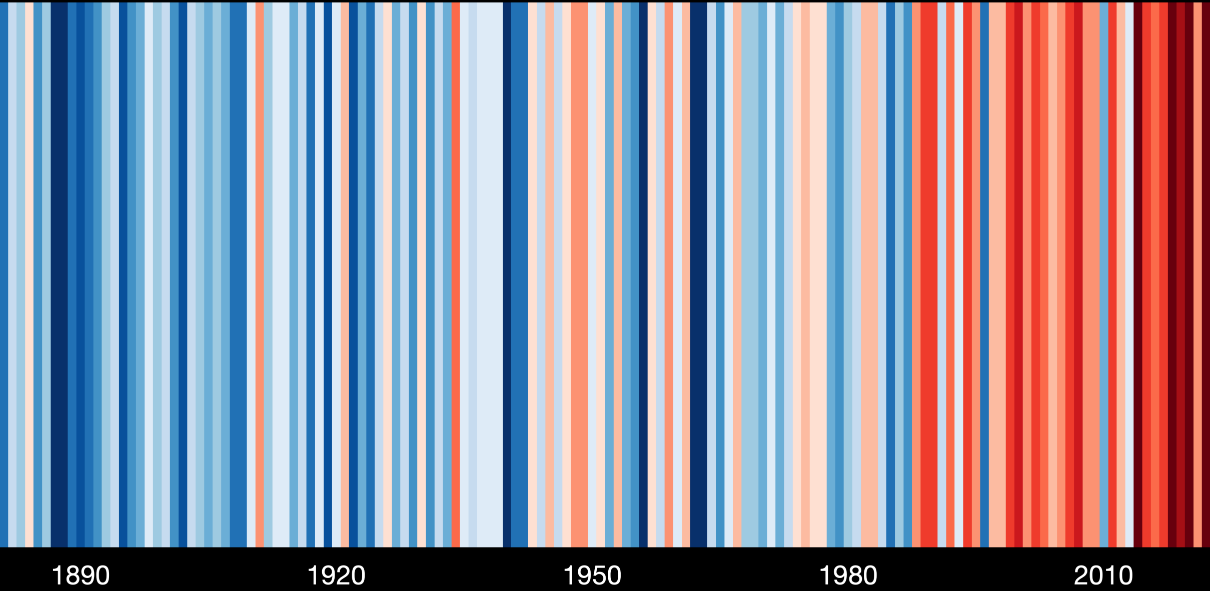
January 16, 2024

INSTITUT FÜR
DIGITAL HUMANITIES
UNIVERSITÄT ZU KÖLN

Climate Change

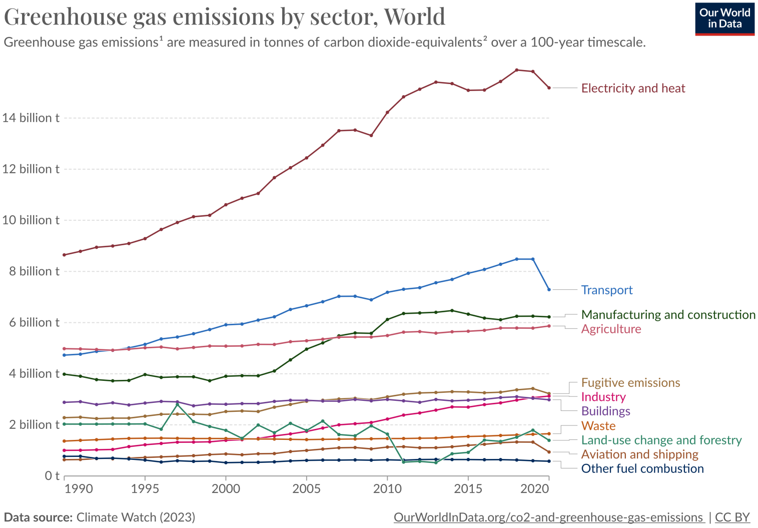
- Caused by emission of greenhouse gases
- Gather in the atmosphere, reflect heat emissions from earth back to earth
- Consequences
  - Increase in average temperature worldwide
  - Change of weather patterns, ocean currents
  - More extreme weather conditions for a longer period of time
  - Rise of the ocean level due to melting ice in Greenland / Antarctica

Temperature change in Nordrhein-Westfalen since 1881



Why is that a topic for us?

Because electricity production is among the top polluters:



Three Areas of Discussion

1. Using machine learning to fight climate change — Nicolas Kayser-Bril (2021)
2. Energy consumption by machine learning — Luccioni et al. (2023); Strubell et al. (2019)
3. Computing in the Future — Mathew Duggan (2022)

Section 1

Using Machine Learning to Fight Climate Change

Using machine learning to fight climate change

Nicolas Kayser-Bril (2021). "Falsche Versprechen". In: *netzpolitik.org*. URL: https://netzpolitik.org/2021/ki-und-klimawandel-falsche-versprechen/

Section 2

Energy Consumption by Machine Learning

Introduction

- Large language models (BERT, GPT, etc.) are becoming the dominant form of ML
- Using them requires a lot of computation, which consume power
- Estimating carbon footprint of such models is difficult
  - Ultimately depends on the primary energy source
- Use cases
  - Ignored: Hardware production, delivery, deployment, computing centre maintenance...
  - Training
  - Application (often called inference)

Training

- Generally: The more parameters, the more computation

Model	Hardware	Power (W)	Hours	kWh·PUE	CO_2e	Cloud compute cost
$T2T_{base}$	P100x8	1415.78	12	27	26	$41–$140
$T2T_{big}$	P100x8	1515.43	84	201	192	$289–$981
ELMo	P100x3	517.66	336	275	262	$433–$1472
$BERT_{base}$	V100x64	12,041.51	79	1507	1438	$3751–$12,571
$BERT_{base}$	TPUv2x16	—	96	—	—	$2074–$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	$942,973–$3,201,722
NAS	TPUv2x1	—	32,623	—	—	$44,055–$146,848
GPT-2	TPUv3x32	—	168	—	—	$12,902–$43,008

Table: CO_2 emission estimates by Strubell et al. (2019). Comparison: NY–SF by plane: 1984 CO_2e

Inference

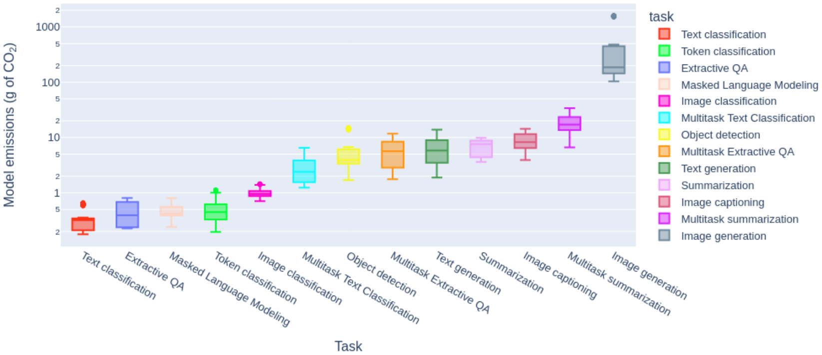


Figure: Per-task analysis by Luccioni et al. (2023)

Training vs. 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: Model training, fine-tuning and inference costs for variants of BLOOMz (Luccioni et al., 2023).
Washing machine: ca. 1kWh.

Key Take-Aways from Luccioni et al. (2023)

- Generative tasks are more energy- and carbon-intensive compared to discriminative tasks.
- Tasks involving images are more energy- and carbon-intensive compared to those involving text alone.
- Decoder-only models are slightly more energy- and carbon- intensive than sequence-to-sequence models for models of a similar size and applied to the same tasks.
- Training remains orders of magnitude more energy- and carbon- intensive than inference.
- Using multi-purpose models for discriminative tasks is more energy-intensive compared to task-specific models for these same tasks.

Section 3

Computing in the Future

Introduction

- This is (reasonable) speculation
- Based on: Mathew Duggan (2022). *Programming in the Apocalypse*. URL: https://matduggan.com/programming-in-the/

Introduction

- This is (reasonable) speculation
- Based on: Mathew Duggan (2022). *Programming in the Apocalypse*. URL: https://matduggan.com/programming-in-the/
- Scenario: It's 2050
  - Some consequences of climate change have happened
  - How does that impact information technology?

Computing Infrastructure

- Operating a datacentre is hard
- Requires constant delivery
  - of parts
    - For computers, but also cooling, cables, power supply, ...
  - of electricity
- Many sea ports become unusable due to rising sea levels

Computing Infrastructure

- Operating a datacentre is hard
- Requires constant delivery
  - of parts

> *[...] raising 221 of the world's most active seaports by 2 meters (6.5 feet) would require 436 million cubic meters of construction materials, an amount large enough to create global shortages of some commodities. The estimated amount of cement – 49 million metric tons – alone would cost $60 billion in 2022 dollars.* *(Jacques Leslie, 2022)*

Computing Infrastructure

- Operating a datacentre is hard
- Requires constant delivery
  - of parts
    - For computers, but also cooling, cables, power supply, ...
  - of electricity
- Many sea ports become unusable due to rising sea levels
- Disturbance of supply chains
- → Consequences
  - More frequent server downtimes
  - 99.999% availability are a thing of the past
  - Distributed service engines will be more popular, because cloud provider knows which computing center is currently operational

Power Infrastructure

- Power grids are old, even in 2024
- More load for heating and cooling due to changed weather
- Power grid suffers from supply chain issues as well
- ➔ Consequences
  - Regular power downtimes
  - Prioritization of who gets the remaining power

Programming

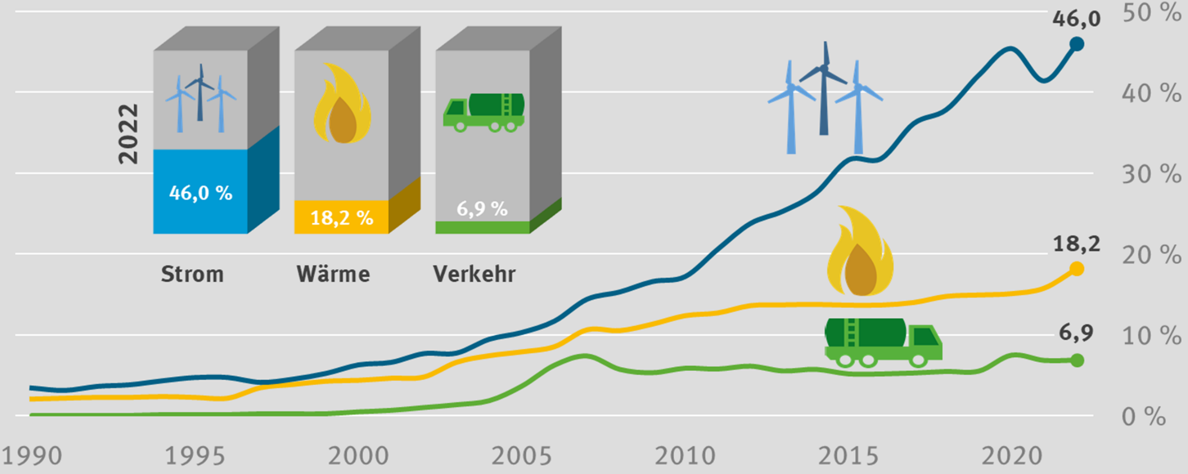
- Difficult to maintain
  - Programming environments that rely heavily on cloud services
    - E.g., dependency management by Node/NPM
  - Containers that consume a lot of bandwidth
    - E.g., docker
  - Open source projects, because
    - many projects are developed in the spare time of individuals
    - and they will be under increased economic pressure

Programming

- Difficult to maintain
  - Programming environments that rely heavily on cloud services
    - E.g., dependency management by Node/NPM
  - Containers that consume a lot of bandwidth
    - E.g., docker
  - Open source projects, because
    - many projects are developed in the spare time of individuals
    - and they will be under increased economic pressure
- More important
  - Energy-efficient programming
  - Robustness due to tests and test coverage

But there is also some good news

Erneuerbare Energien:
Anteile in den Sektoren Strom, Wärme und Verkehr bis 2022



Quelle: Umweltbundesamt auf Basis Arbeitsgruppe Erneuerbare Energien-Statistik (AGEE-Stat)
Datenstand: 10/2023

References I

- Jacques Leslie (2022). "How Climate Change Is Disrupting the Global Supply Chain". In: *Yale Environment 360*. URL: https://e360.yale.edu/features/how-climate-change-is-disrupting-the-global-supply-chain.
- Luccioni, Alexandra Sasha/Yacine Jernite/Emma Strubell (2023). *Power Hungry Processing: Watts Driving the Cost of AI Deployment?*
- Mathew Duggan (2022). *Programming in the Apocalypse*. URL: https://matduggan.com/programming-in-the/.
- Nicolas Kayser-Bril (2021). "Falsche Versprechen". In: *netzpolitik.org*. URL: https://netzpolitik.org/2021/ki-und-klimawandel-falsche-versprechen/.
- Strubell, Emma/Ananya Ganesh/Andrew McCallum (2019). "Energy and Policy Considerations for Deep Learning in NLP". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, S. 3645–3650. DOI: 10.18653/v1/P19-1355. URL: https://www.aclweb.org/anthology/P19-1355.pdf.