AI and environment

How are they connected?
What can we do about it?

Dr. Olya Kudina

TU Delft

o.kudina@tudelft.nl

@OlyaKudina

Lecture outline

1. Where is AI? Making the invisible visible
3. Data centers. A sociotechnical perspective
4. Responsible engineering potential
5. Ethics in more-than-human worlds

AI DeMoS Lab - AI as Deliberative Multimodal Systems - Developing AI for democracy

Collaboration of philosophy, design and computer science

TU Delft AI Labs Programme: https://www.tudelft.nl/ai/ai-demos-lab

Twitter: @DemosAILab

AI DEMOS LAB

Olya Kudina, PI
values and AI

Nazli Cila, PI
more-than-human design

Dmitry Muravyov,
PhD cand.
Generative AI, trust, epistemology

Jordi Viader Guerrero, PhD cand.
Social medial and political subjectivation

Karin Bogdanova,
PhD cand.
AI mental healthcare and diversity, inclusion

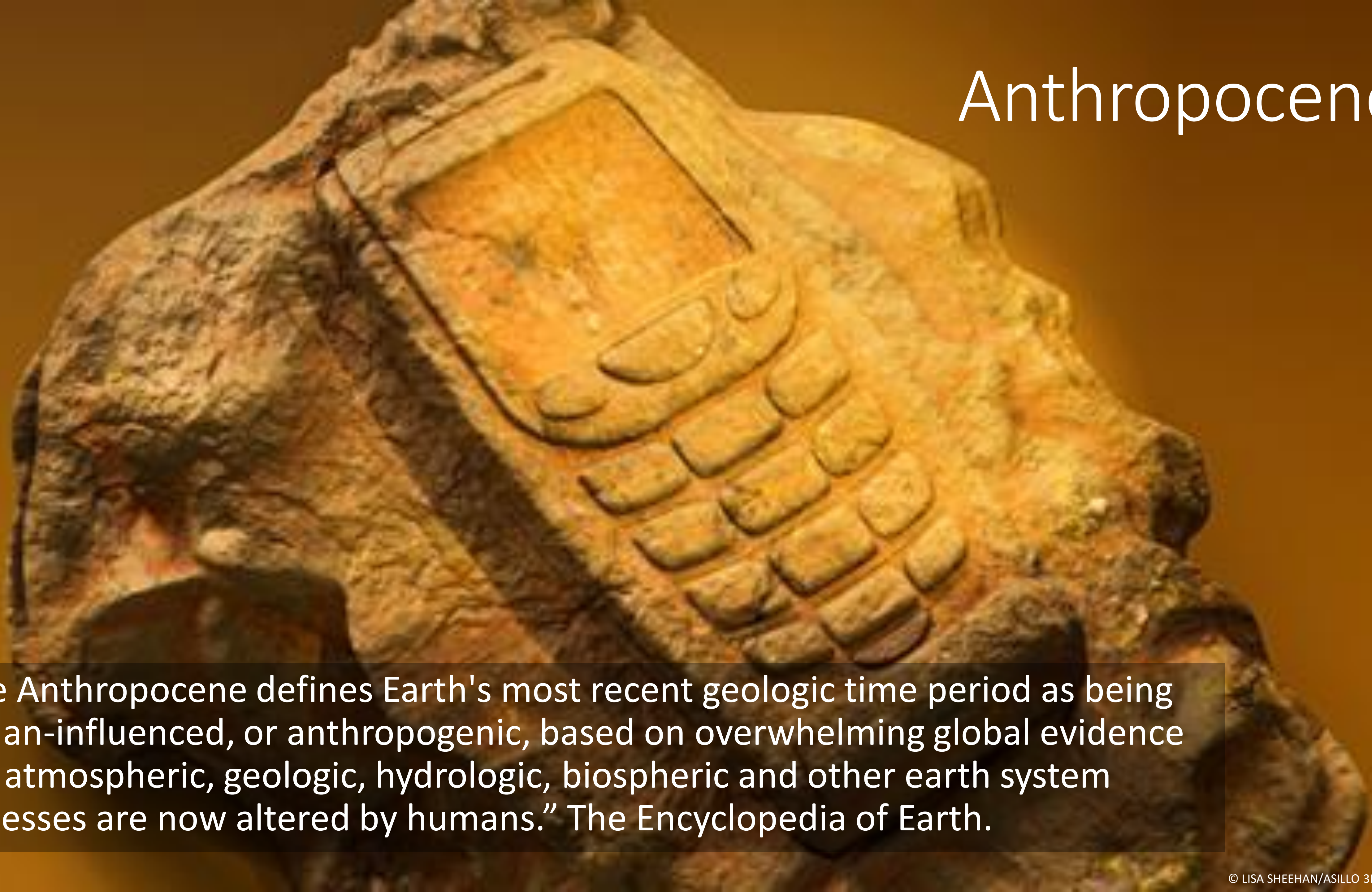
Meike Hardt,
PhD cand.
AI public infrastructures, care relations, contestations

Syafira Aulia,
PhD cand.
Adversarial ML for privacy-enhanced deliberation online

Where is AI?

AI systems:
Making the invisible visible

Anthropocene

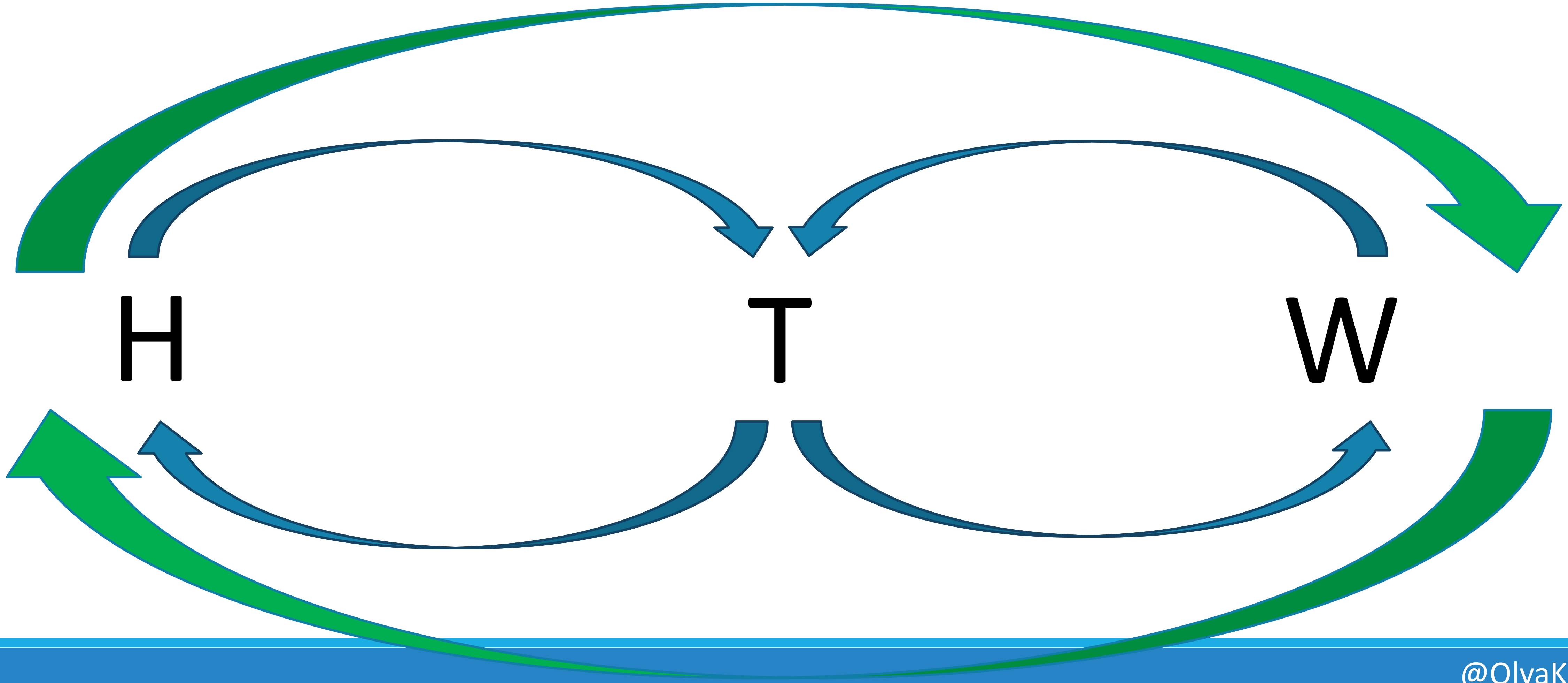


"The Anthropocene defines Earth's most recent geologic time period as being human-influenced, or anthropogenic, based on overwhelming global evidence that atmospheric, geologic, hydrologic, biospheric and other earth system processes are now altered by humans." The Encyclopedia of Earth.


Guiyu (China), “electronic graveyard of the world”
John McGarrity


Burning electrical wires to recover copper
Agbogbloshie (Ghana), 2019

AI as sociotechnical systems in the world



Making the invisible visible

AI systems are material: “relying on manufacturing, transportation, and physical work; data centers and the undersea cables that trace lines between the continents; personal devices and their raw components; transmission signals passing through the air; datasets produced by scraping the internet; and continual computational cycles. These all come at a cost.”

Crawford (2021), p. 49


411.08
536.85
667.43
292.76
64.85
53.95
@OlyaKudina

Embodying technological systems and understanding the logic of extraction

- Of the earth and to the earth

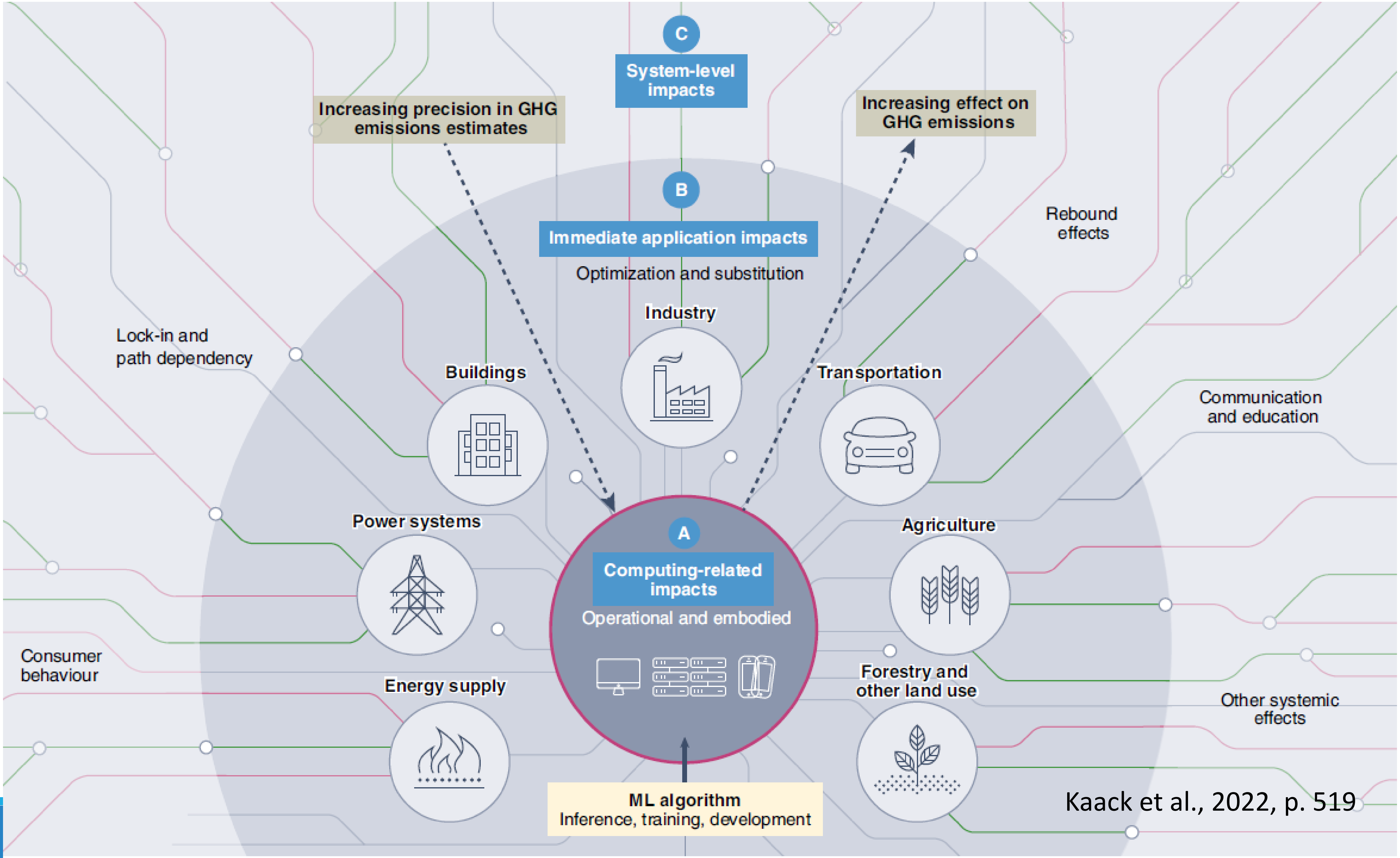
- The social factor

- The politics of data centers

- Responsible engineering: pitfalls and strategies

AI is of the earth

What exactly is AI?



Kaack et al., 2022, p. 519

What is AI made of?

What is AI made of?

The starting points:

- earth minerals

- energy/electricity

- water

The minerals

Lithium mine, Atacama Desert, Chile

OPEN COMMONS

The minerals

- Lithium = mass-market batteries but a limited lifespan (e.g. 62,5 kg of lithium in one Tesla Model S battery pack. Tesla consumes 50% of the world's lithium production, about 28000 tons/year)

- Nickel, copper, lithium – no AI without them. May not be rare, but the environmental cost to produce the usable amount is huge. Who pays the burden?

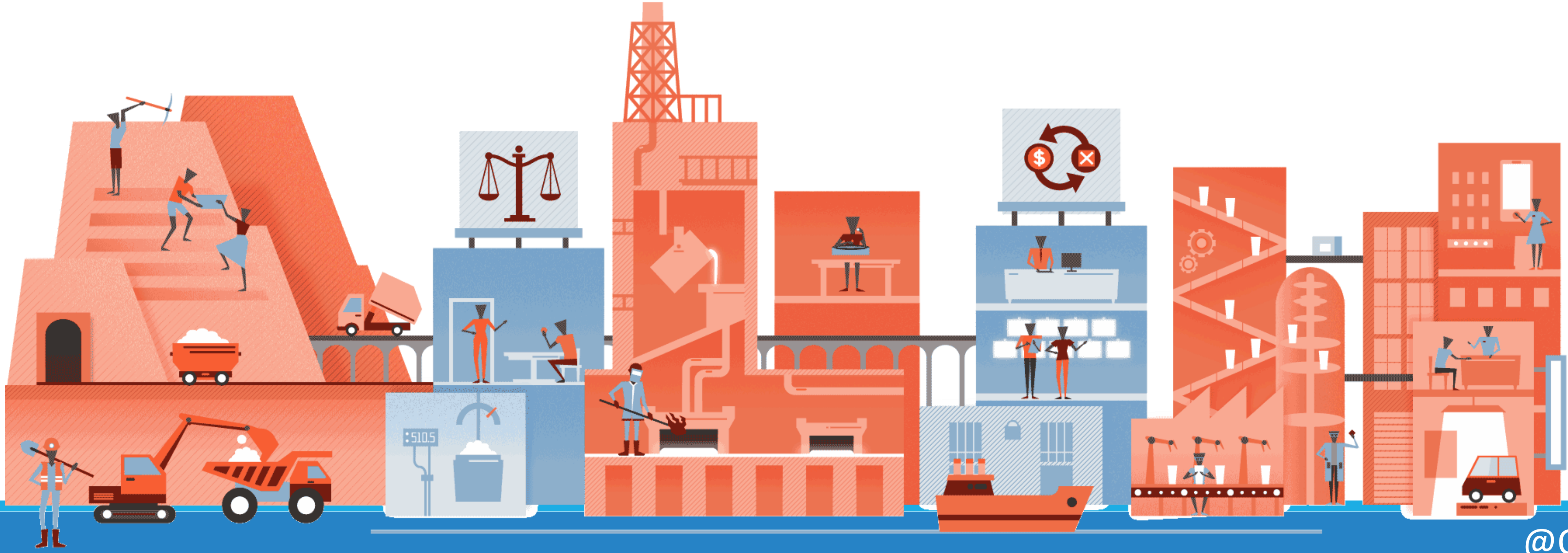
Unsustainable extraction practices

- Rare earth minerals, “conflict minerals” – e.g. Congo, Mongolia, Indonesia. Essential for fiber optic cables, GPS technology, etc. Accompanied with killing and slavery in the mining sector, and environmental degradation.

- Minerals take billions of years to develop in earth vs. 2-4 yrs average lifespan of a cellphone: “[W]e are extracting Earth’s geological history to serve a split second of contemporary technological time” (Crawford, 2021, p. 31)

Tracing the supply chain

A diffused network and a difficult task: took Intel 4 yrs to get a basic idea; Philips has thousands of component manufacturers; Dell – “an insurmountable challenge” - “Ignorance of the supply chain is baked into capitalism” (Crawford, 2021, p. 35). So should we just give up?



@OlyaKudina

The energy

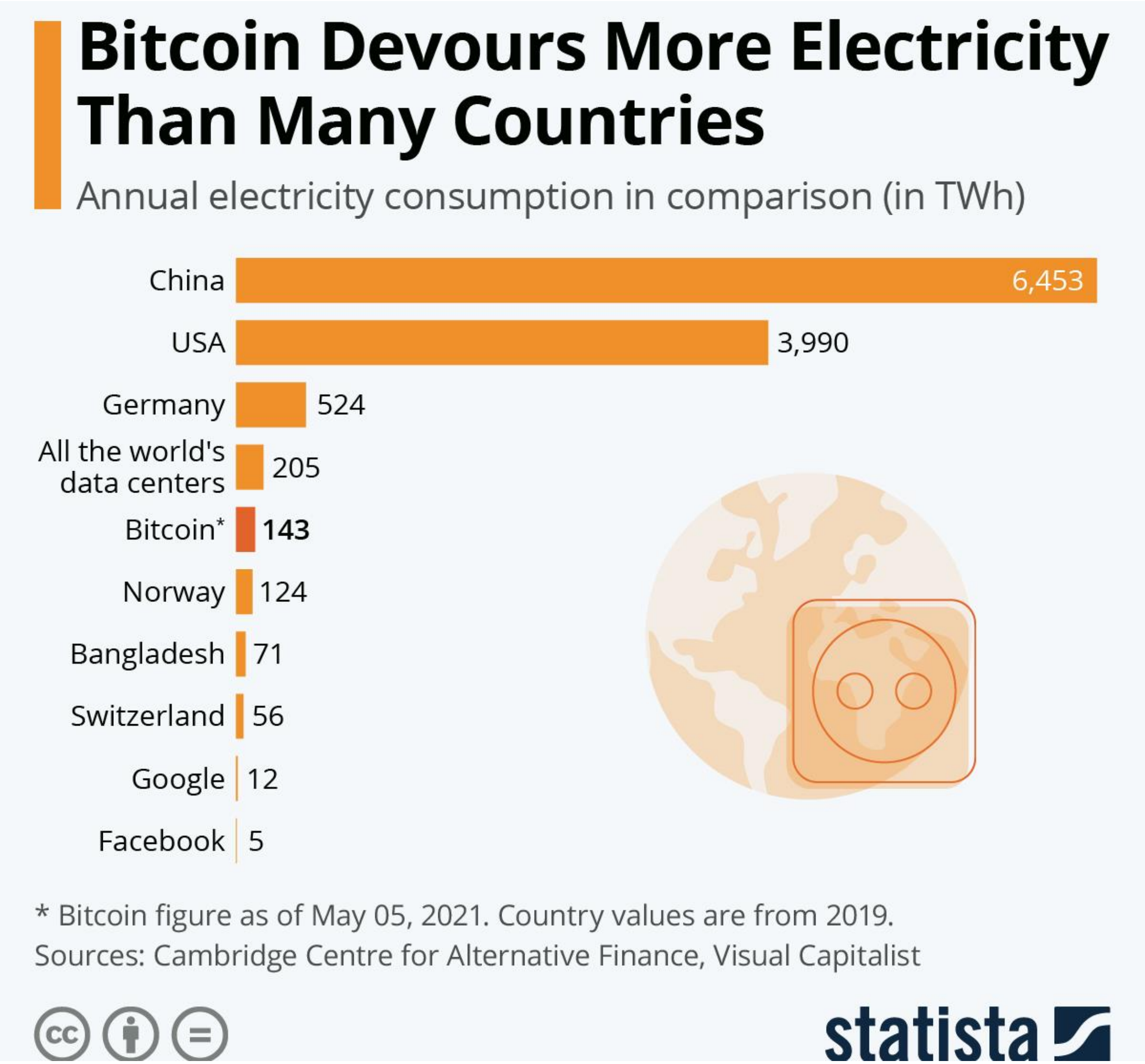
Training ML systems

1. Model inference – processing new data through a trained model, the least energy-intensive, but most frequent step: e.g. classifying toxic comments, labelling and classifying images, Google translate matching words. Adds up – at Facebook, model inference has larger carbon footprint (i.e. energy consumption) than model training.

2. Model training – to learn a function, how to achieve a result or to correct and update it. Requires many runs of the dataset, or 'epochs.' More energy-intensive than a single inference but done less frequently. But, e.g. at Facebook, ML models can be retrained from hourly to once in several months.

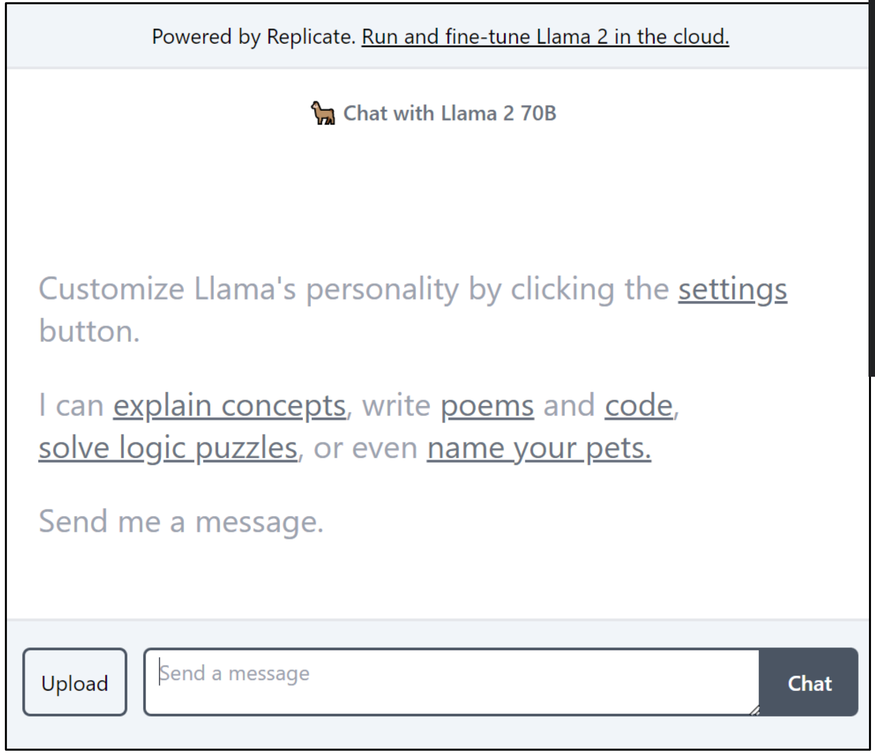
3. Model development and tuning – training different model variants on different datasets to understand what works best. Most energy-intensive, especially neural networks with many configurations possible, trial-and-error and validation.

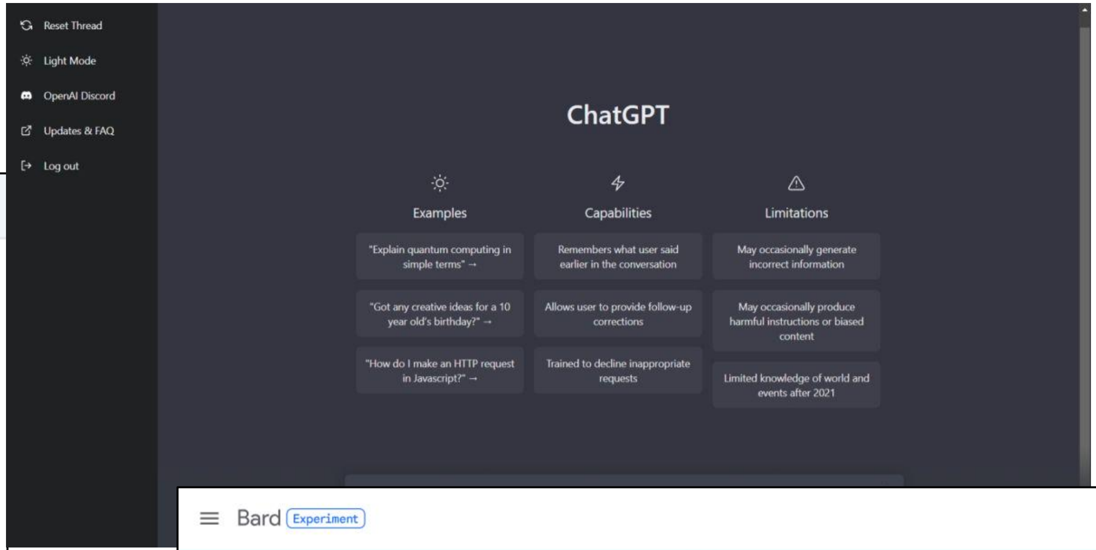
Kaack et al., 2022, p. 519

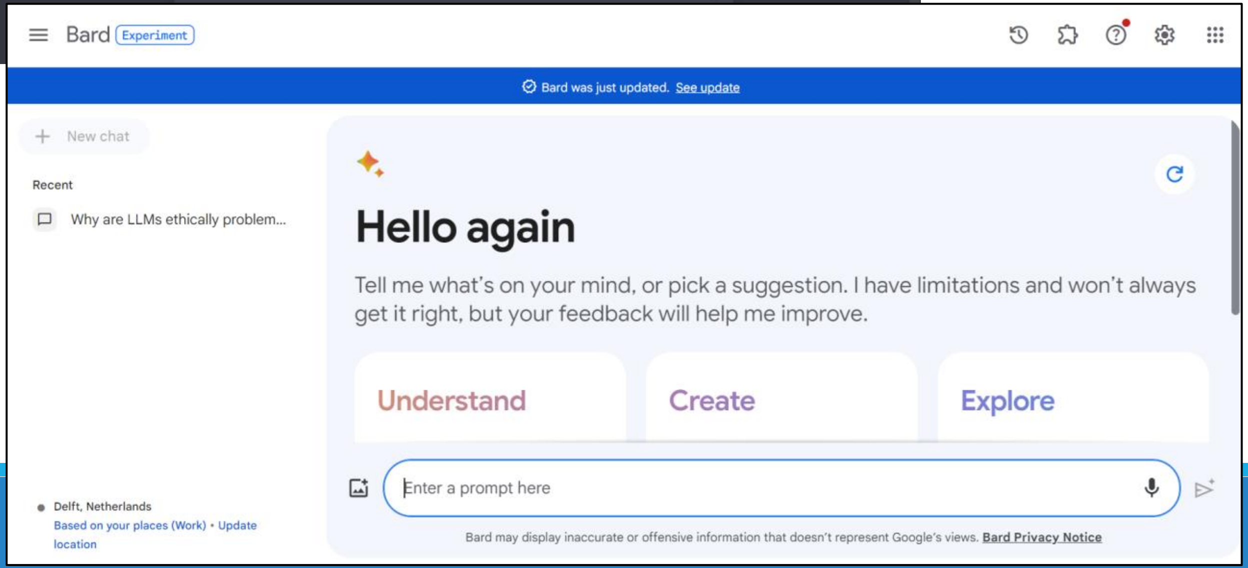


Bitcoin example

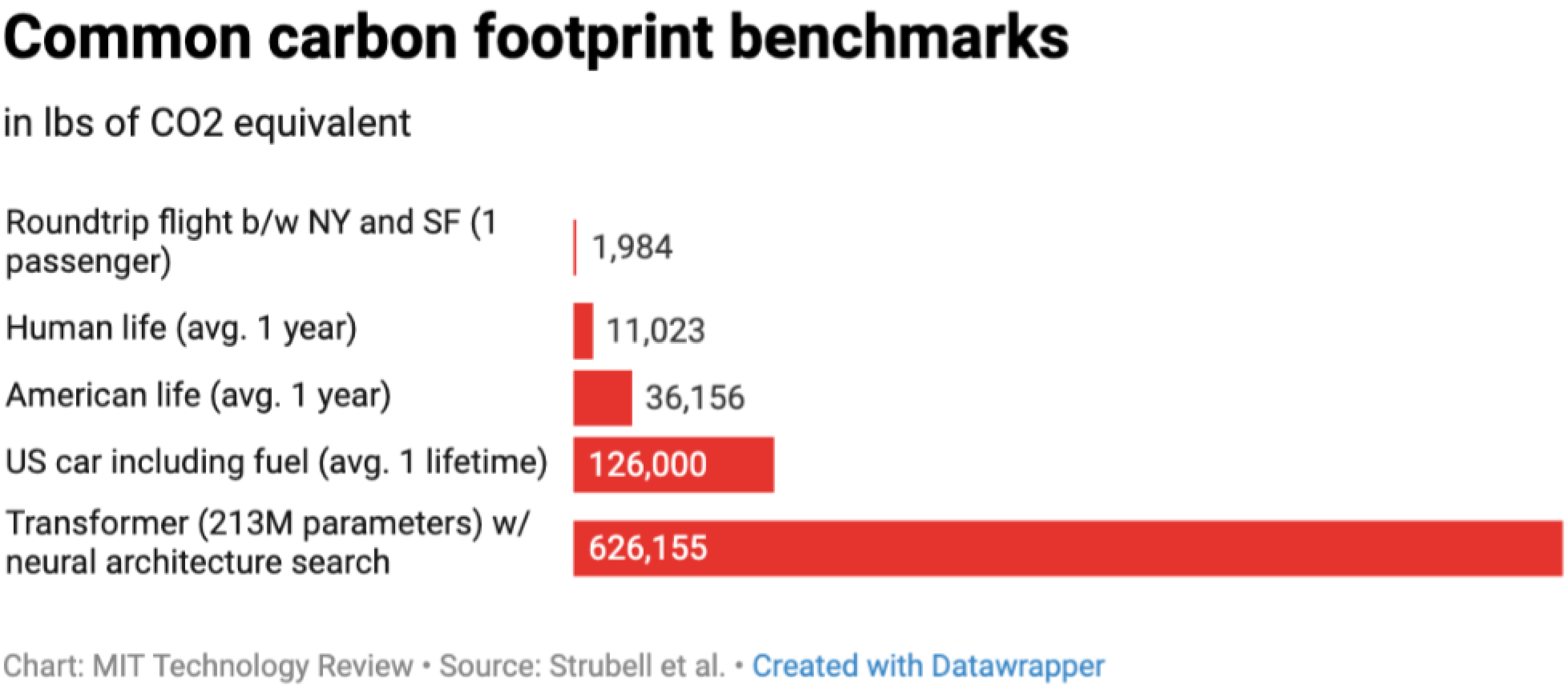
New kid on the block: GenAI & LLMs







NLP example



Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and policy considerations for deep learning in NLP." *arXiv preprint arXiv:1906.02243* (2019).

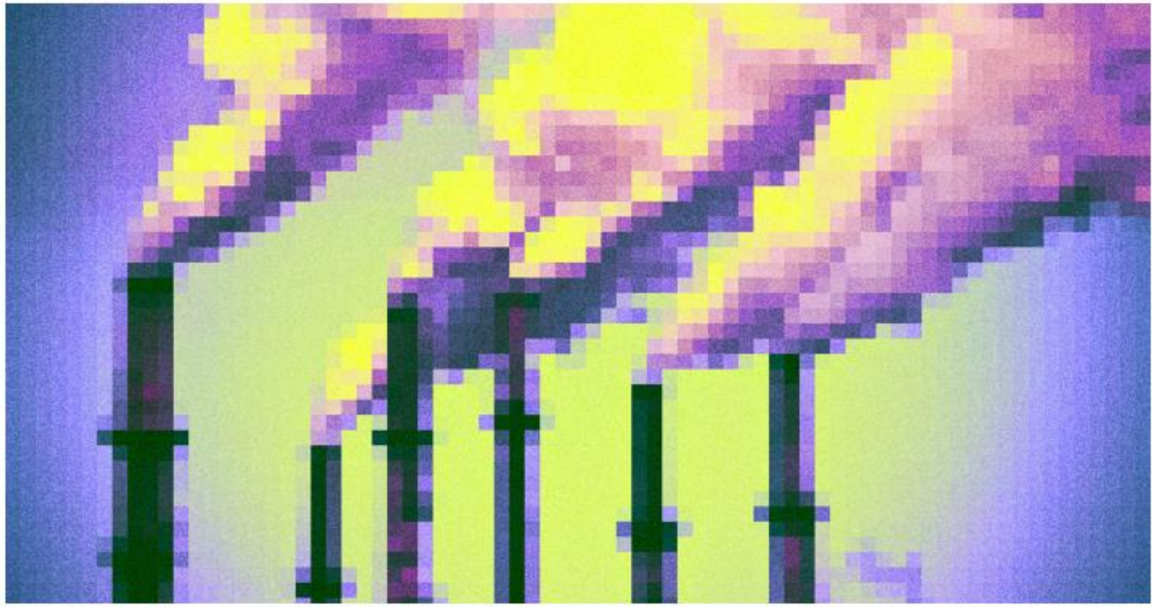
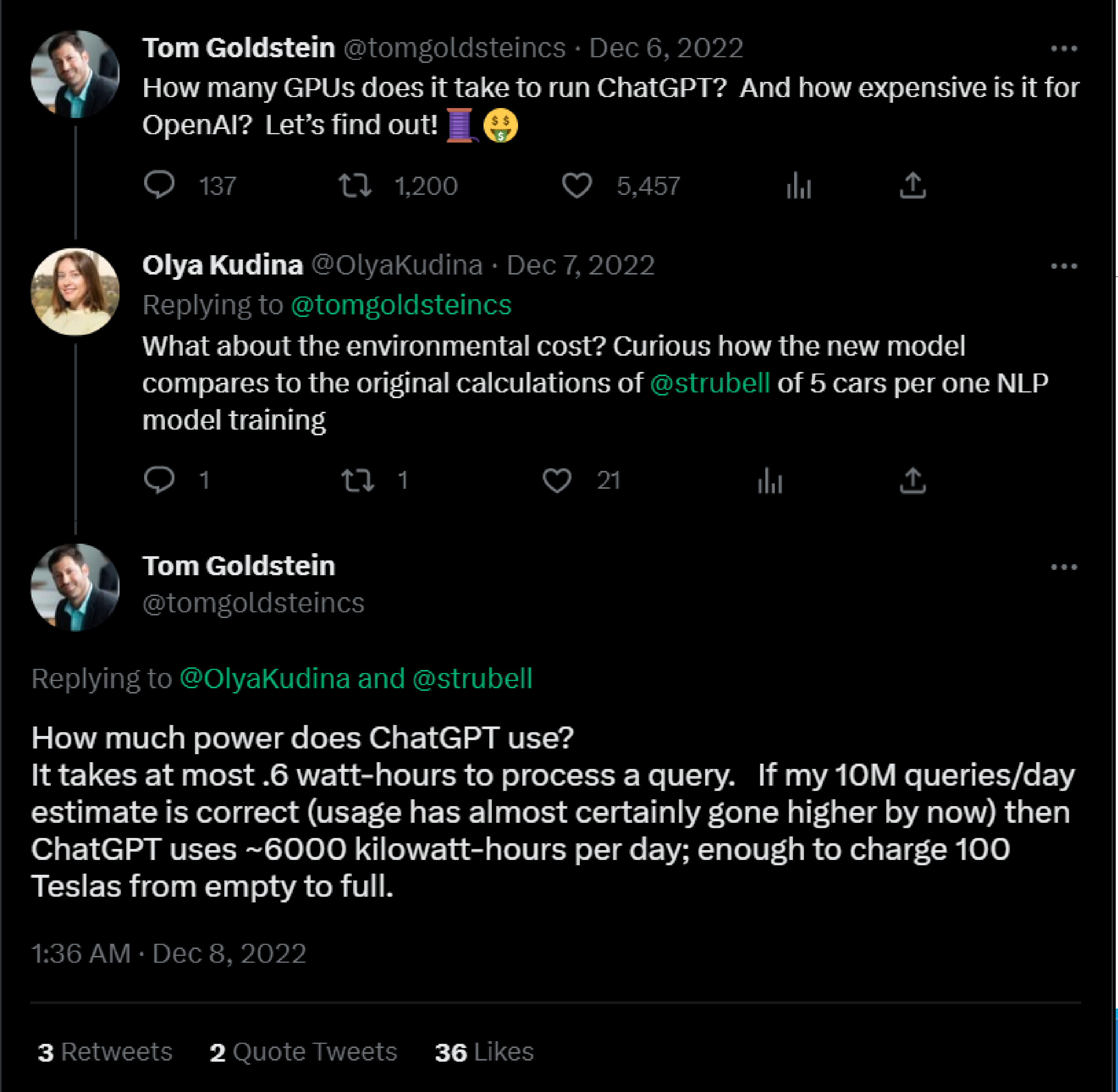


Image credit GETTY / FUTURISM

ENVIRONMENTAL COSTS

- **Lack of corporate data**, estimates based on experts' investigations.
- Training (energy): 51000-62000 MWh = 5-6 years of 1000 households
  Using (energy): e.g. Jan 2023 - 590 mln visits = 1.1M-23M KWh = monthly 175,000 Danes
  Using (water): every 20-50 prompts = 500 ml fresh water
- **ChatGPT is bad for the environment.** Consider if you REALLY need to use it at all.

Ludvigsen, 2023 - https://towardsdatascience.com/the-carbon-footprint-of-gpt-4-d6c676eb21ae
Li et al., 2023 - https://arxiv.org/pdf/2304.03271.pdf



ChatGPT is highly popular – ≈13 mln visitors per day! (Reuters). The step **of model inference** - processing new data through a trained model, is the least energy-intensive. But! It is the most frequent step: e.g. matching words or running prompts. **Adds up to the biggest cost category.**

ChatGPT is very bad for the environment. What do we do when we include it in the education setting? How does it fit with your university's carbon footprint goals?

The water

Lauren Callaway

The water demand

Water needed to cool down large data centers. NSA data center in Bluffdale, Utah – 6,4 mln liters per day. Can't be used for consumption after, polluted.

Water needed to make earth minerals usable. E.g. neodymium and cerium, etc. require dissolving in large quantities of sulfuric and nitric acid, dissolved then in water. Produces acidic baths, "dead lakes" (e.g. Baotou in Mongolia), acid-bleached rivers, dehydrated landscapes. The land and water will take generations to renew.


Patrick Famisaran


Liam Young/Unknown Fields

AI is of the earth: summary

AI systems are driving environmental extraction.

The costs of computational infrastructure already surpass aviation industry per year. Expected to increase to 14% of global greenhouse emissions by 2040.

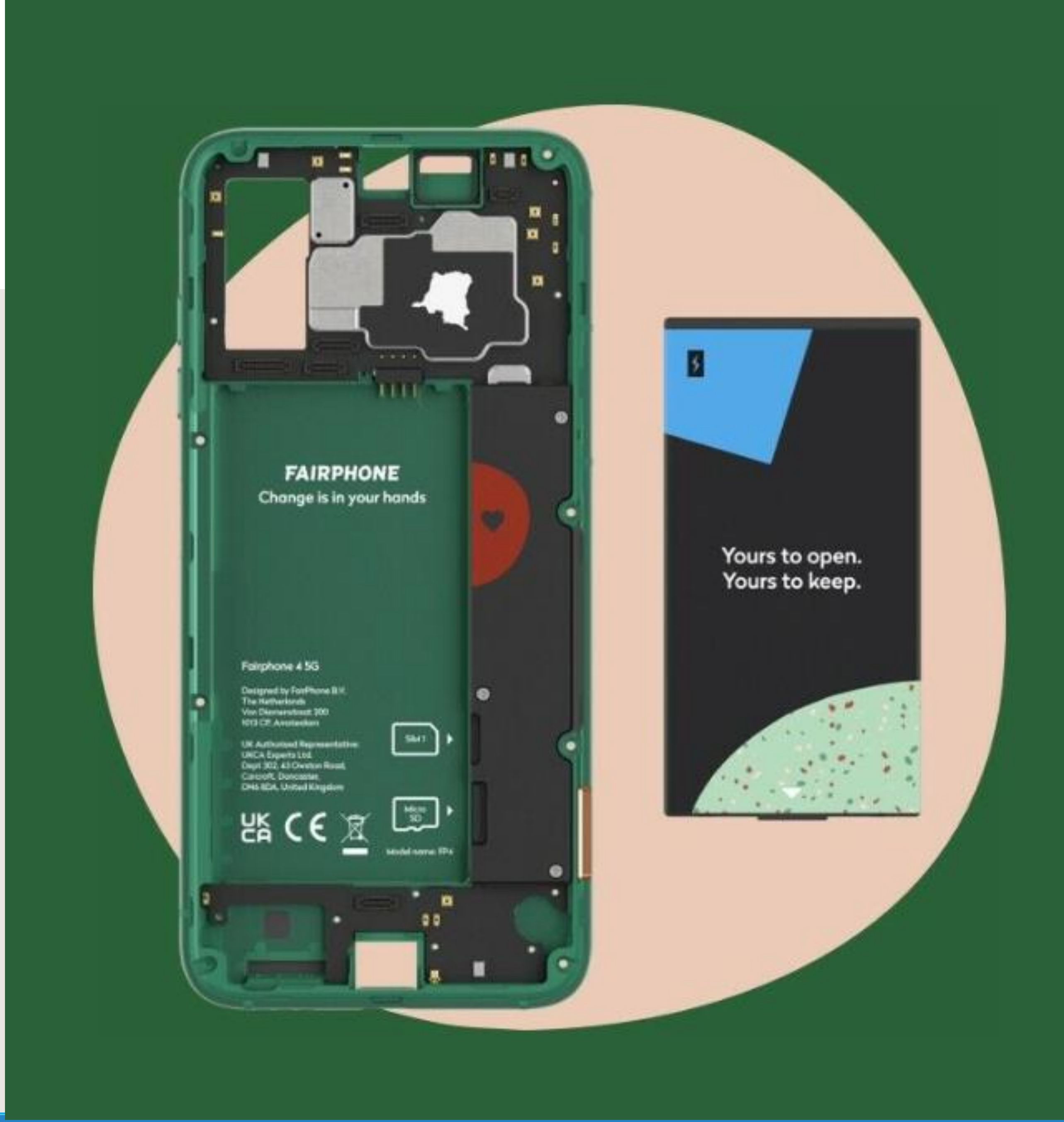
Minerals are the backbone of AI, its “lifeblood” is electricity.

Speaking of AI in immaterial metaphors – e.g. “the cloud”, hides and abstracts its very real social, material and environmental costs

Crawford (2021)

Exercise

Fairphone (2013, NL)



https://www.fairphone.com/en/impact/

https://www.youtube.com/watch?v=zEvMaTlo7pU&ab_channel=Fairphone

Fairphone

1. What kind of problems does Fairphone try to address?
2. How effective are the strategies they propose?
3. Do you know anything about where your phone comes from? – I challenge you to find out.

Break

Data centers.
A sociotechnical perspective

A Facebook data centre in Luleå, Sweden (C) Nicola Jones.

Data centers

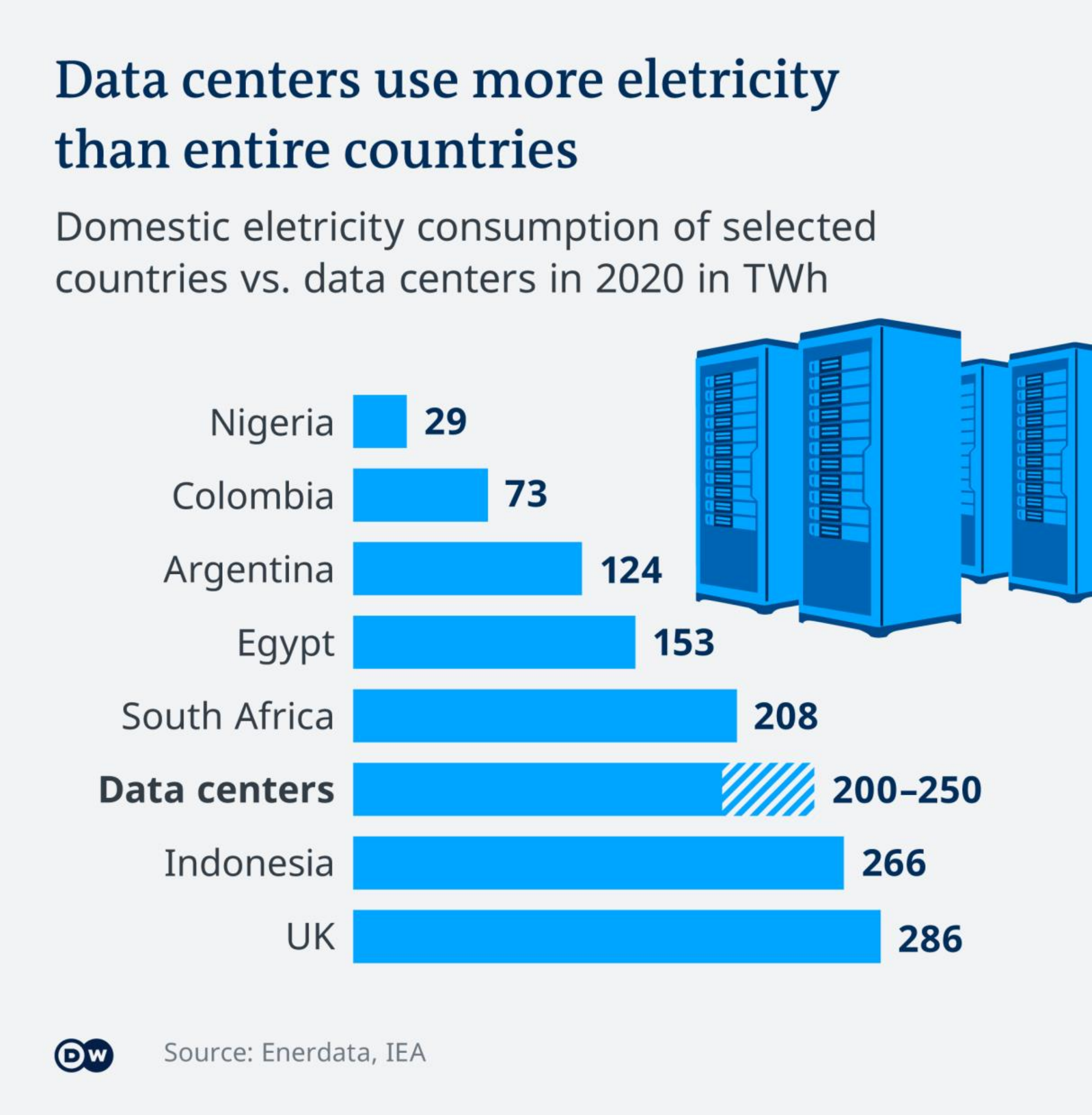
Globally – circa 1% of energy consumption, in the EU – 2.7% (excl. cryptocurrency)

E.g. Facebook: in 2016 - 1.83 mln MWh of electricity, in 2020 - 7.17 mln (6.9 mln for datacenters) – 390% increase

Electricity for running machines cool (40%)

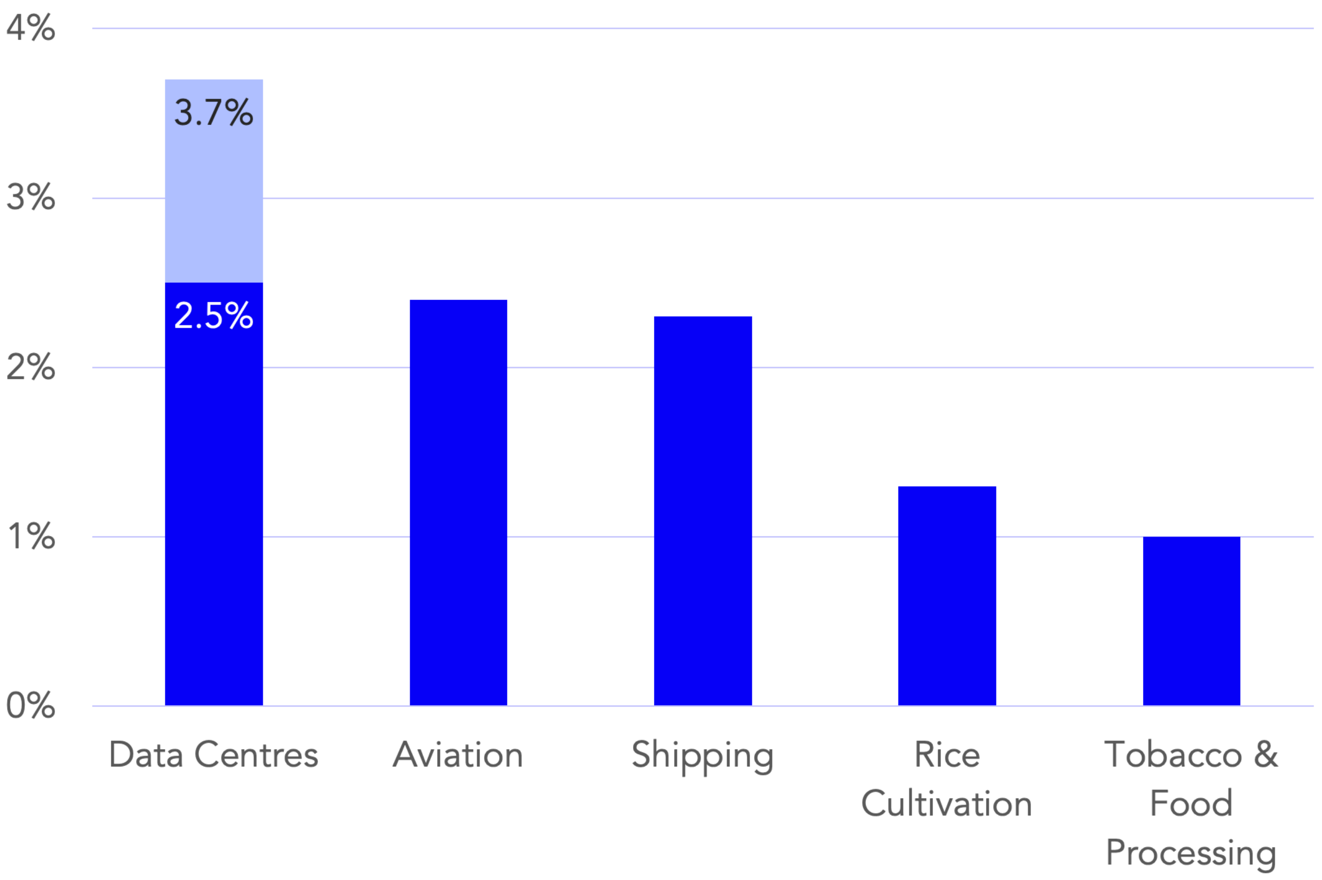
EU - the goal of climate-neutral data centers by 2030.

Rooks, T. (2022). Data centers keep energy use steady despite big growth. *DW*



Global cloud computing emissions exceed those from commercial aviation

Share of global CO_2 emission generated by sector/category



Source: Climatiq Analysis, The Shift Project, OurWorldinData

climatiq

Data centers as a sociotechnical issue

Not every company can afford moving to more sustainable hyperscale data centers: frequent use of outdated equipment, old cooling systems, etc.

Running not only what’s need – but also “just in case” scenarios, a system of hyperredundancies: constant use of aircos even when not needed, diesel generators, redundant servers ready to take over, always on. Result – much energy waste.

Koomey at al. did a study of US corporate data centers in the US 2017 – 25% were “zombies,” using energy without any work, forgotten to be unplugged or as redundancies.

Data center operators risk their livelihoods if the operation is interrupted, e.g. due to overheating. Huge financial costs of every second of downtime -> constant fear and mental pressure of operators. Use of wasteful flood cooling strategies. Emotion, human judgement, instinct also in the data centers, somebody needs to care for the centers.

Gonzalez Monserrate, S. (2022). The cloud is material. *MIT*.

CLIMATE & ENVIRONMENT

In a small Dutch town, a fight with Meta over a massive data center

By Tracy Brown Hamilton

© Britta Pedersen/dpa/picture-alliance

The politics of data centers: Dutch case

Dec 2021 – Zeewolde plans to host a Meta data center, appr. 1.3 sq.km (130 hectars), powered by clean energy

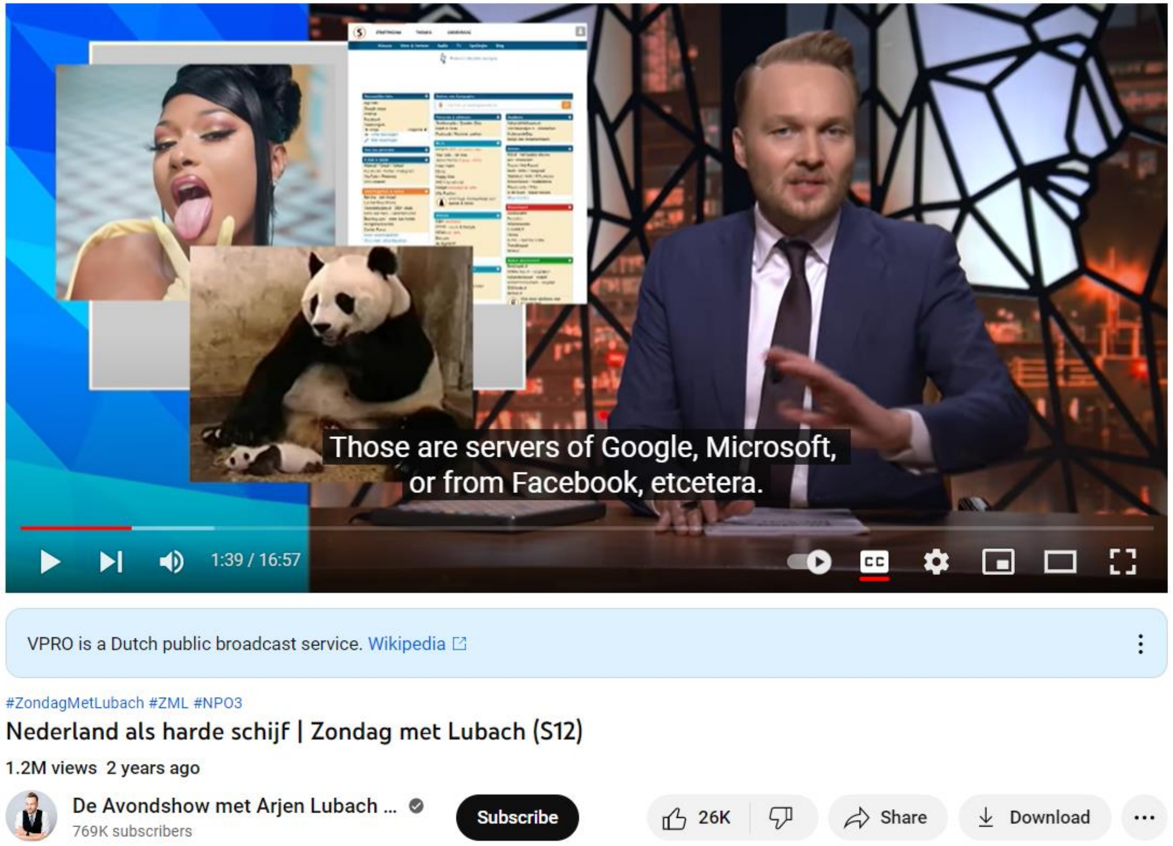
NL hosts over 200 data centers, incentivizes by low taxes, low energy prices – 2% NL energy use, 10% wind power

Huge local protests, demise of the local government, national “pause” on data centers for 9 months

Meta’s data center at the core of Dutch dilemma:

competitive global edge by hosting big data centers vs. country’s sustainability goals (+ local community values)

WHY?



The politics of data centers: Dutch case

The sociomaterial lens:

The earth as the context - Zeewolde – a land recovered from the sea, rich mineralized soil; the biggest clean energy producer in NL, more than it consumes through wind farms

Technology - expected energy use of Zeewolde data center 1.3 TWh yearly = the city of Amsterdam. All the other earlier points

Sociocultural environment - The law passed in Dec 2021 could reassign rich agricultural land for industrial use -> to Meta. Local protests from the activists (e.g. Susan Schaap) and farmers: 1. richest agricultural land in the country; 2. canals to cool the data centers and pollution concerns; 3. huge noise pollution

Local elections in March 2022 removed the supporting party and put the opposition in charge. Meta immediately issued a statement saying they want friendly relations with their neighbors and will “pause” the project in Zeewolde.

Government lacks clear policy. For now, a limit on size, 10 ha vs. 130 ha planned in Zeewolde. Google plans to build such a data center in Groningen region, Winschoten

Responsible engineering (and use) perspectives

AI for environment?

Used for tracking deforestation, evaluating susceptibility to coastal inundation, aid in the design of next-generation batteries, to forecast renewable power production, crop yields and transportation demands. By controlling and improving the operational efficiency of complex systems, such as industrial heating and cooling systems, robots can be used to save resources and energy (Quoted from Kaack et al., 2021, p. 521)

Careful not to give an idea of technological fix – "attempts to use engineering or technology to solve a [complex sociocultural] problem (often created by earlier technological interventions)" (Cook, 2009)

Engineering possibilities

We should always start with the question “Do we need to build this system?”

For whom is it needed? What are its hidden costs?

Are marginal increases in error rates worth huge parallel energy and emissions costs? E.g. ResNet and image classification in Kaack et al., 2022. Keep energy consumption as a criteria in mind when choosing between different options.

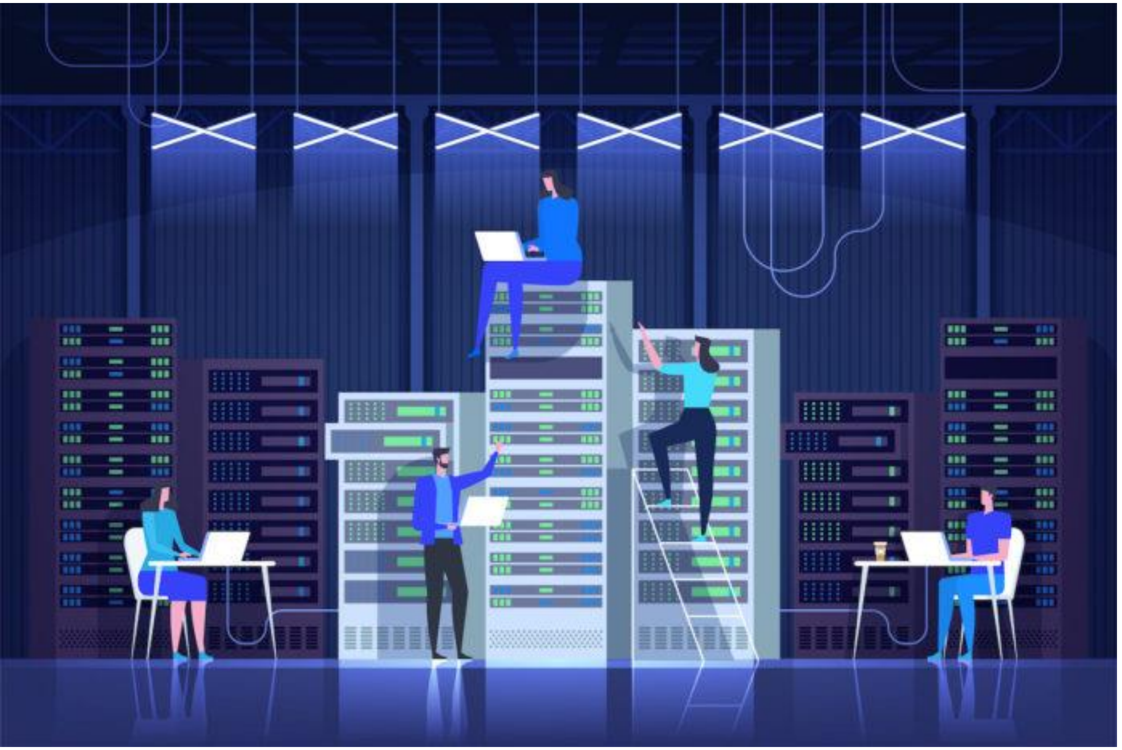
Keeping oneself proactively in check:

Keeping oneself proactively in check

- ✓ measuring and reporting on energy use and emissions, e.g., ML Emissions Calculator (Anthony et al., 2020)
- ✓ usage patterns disclosure
- ✓ model metadata disclosure
- ✓ shifting your work hours to times of day with higher shares of renewable energy.

Which strategies can you think of? Pros and cons

Policy possibilities



- Data centers rules and enforcement

- Supply chain transparency and enforcement

- Noise pollution regulation update – “a right to quiet”, a new “right to be forgotten”? Value change in the making

- Concrete measures, e.g. “banning high-definition color cameras on phones could reduce data traffic in Europe by 40%” (Betterlin, I. in Jones (2018)); “materials passports” to promote reuse of devices and detail their histories

- Concrete measures about data centers: forbidding using drinking water, mandating solar panels, increased noise reduction measures by e.g. thicker walls, aesthetically pleasing design, making waste heat usable (e.g. into electricity)

Rebound effects

Whenever we attempt to decrease something, it can without intention increase in other means due to behavioural and systemic effects – e.g. efficient electricity bulbs, low-calory food.

Also sustainability and robotics: robots can help to achieve climate goals and at the same time change these goals, making them hard to reach.

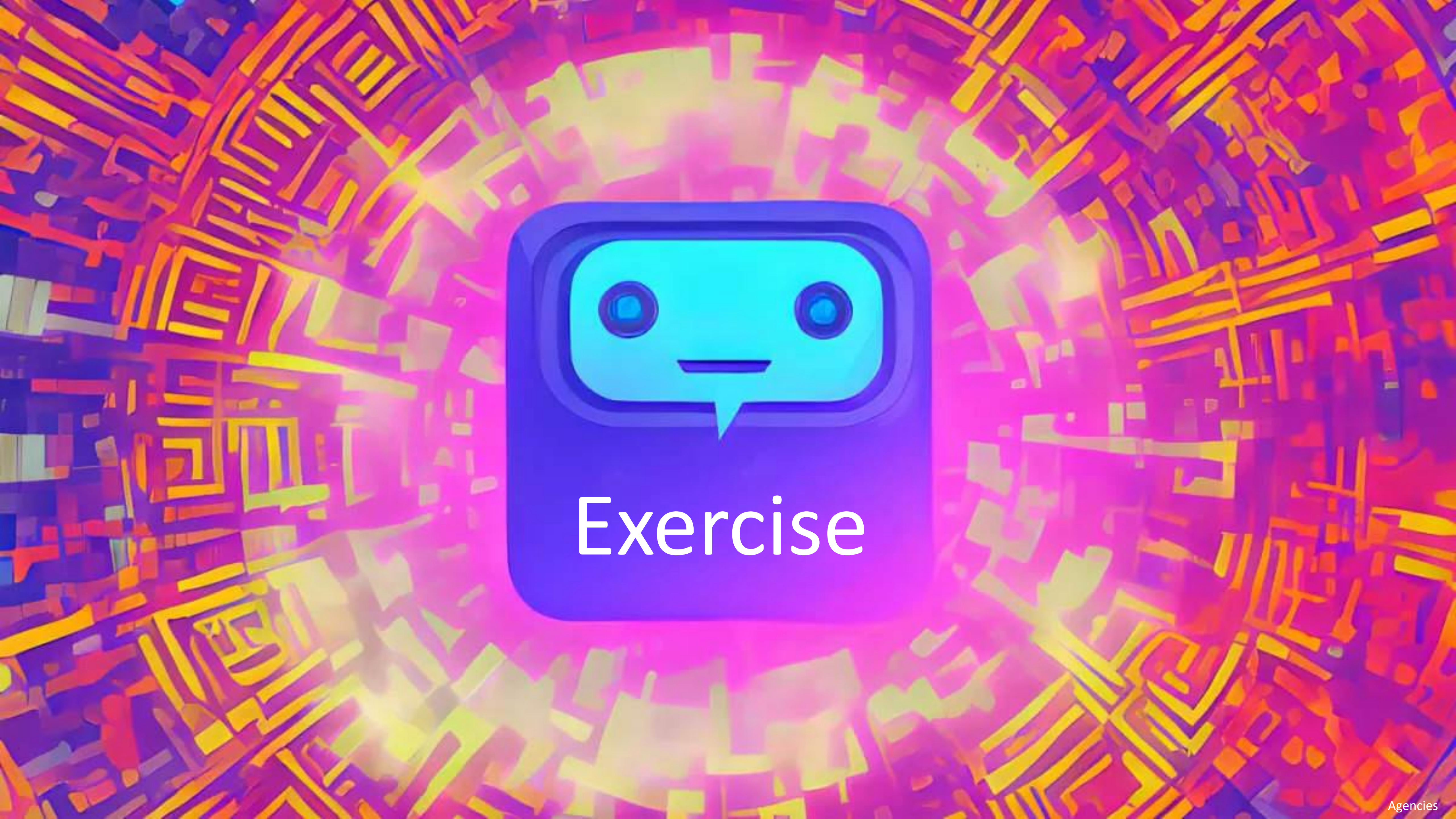
“ […] Autonomous vehicles can improve fuel efficiency, but they may also lead to higher rates of individualized vehicle travel, potentially increasing overall energy use and emissions” (Kaack et al., 2022, p. 522)

The virtue of moderation

The average web page tested produces approximately **0.5 grams** CO2 per page view. For a website with 10,000 monthly page views, that's **60 kg** CO2 per year.

Keeping yourself in check – e.g. Website Carbon Calculator

In parallel, we can all attempt to change our individual consumption behaviours. Like the small costs of running individual inference models add up, so do our individual small actions in decreasing our thirst for data (e.g. 30% of US traffic is for Netflix)



Exercise

AI and the environment

1. What can be improved in terms of the environmental impact of AI, e.g. ChatGPT or another concrete application? Think of several potential strategies.

2. Discuss the feasibility of these strategies.

3. Provide your group's opinion on how to realistically deal with the environmental problems of AI.

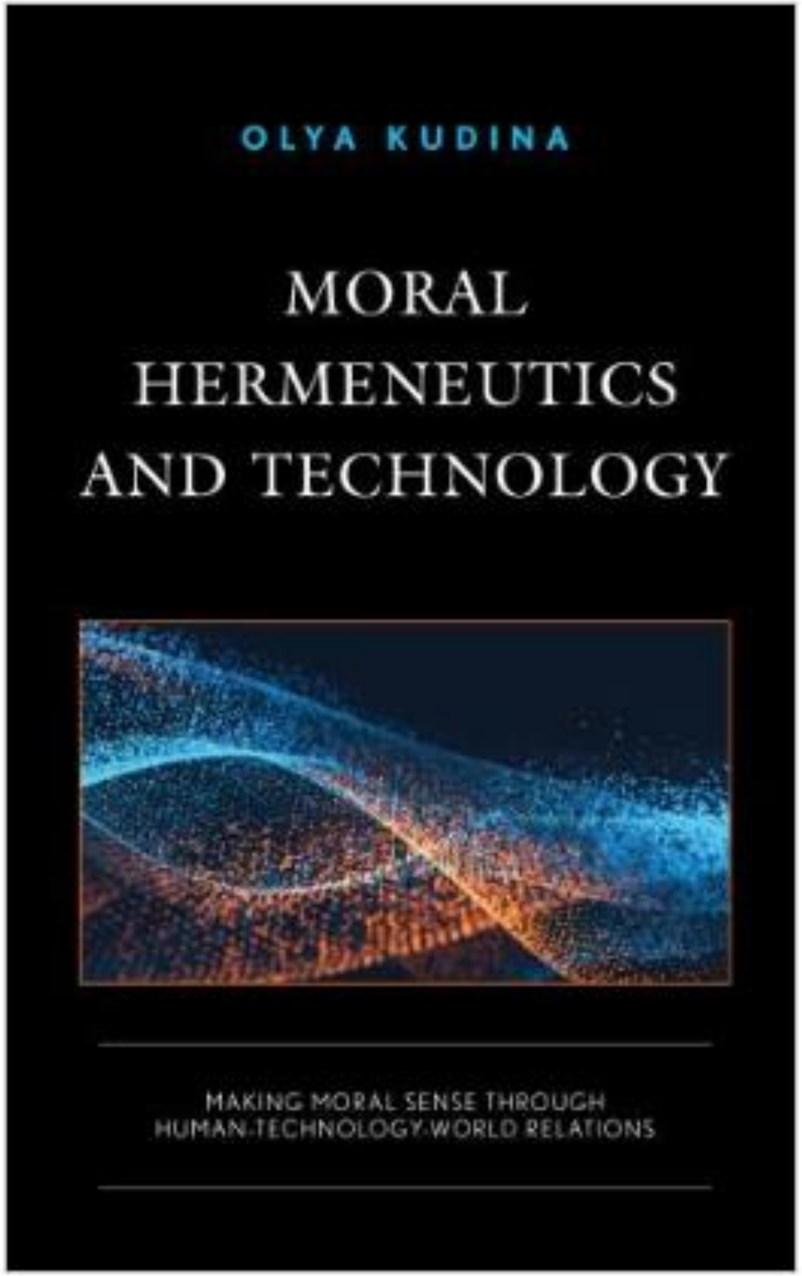
The ethics in Sustainable AI rhetoric

Moral hermeneutics and technology:

Making moral sense through human-technology-world relations.

December 2023. Lexington Books Rowman & Littlefield.

Open access to an e-book!



Summary

AI systems consist of earth-bound elements and have very real and visible social and environmental costs.

The sociotechnical systems perspective allows to make the invisible or less spoken of components of robotics visible and available for reflection.

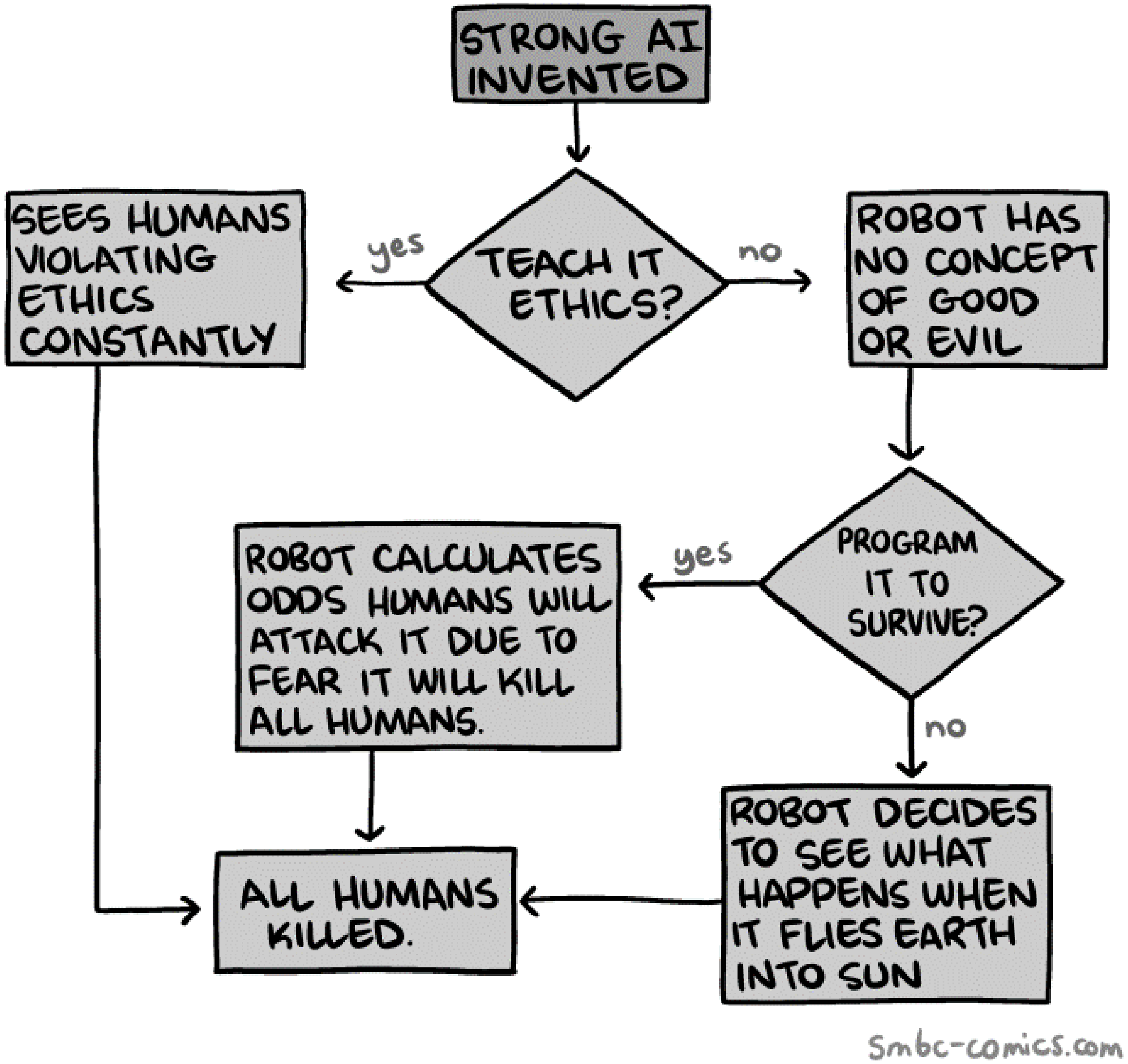
AI systems can help towards the climate goals – but they can never pretend to fix the environmental problems because a. they were partly caused by these tech and b. these are complex social and cultural problems that require complex policy and use solutions, not only tech.

We can no longer afford to ignore the environmental damage that AI systems perpetuate – need to think of making them more sustainable, from any perspective available to us.

CHATGPT IN HIGHER EDUCATION

Please, fill in this questionnaire and share with your students to understand how ChatGPT changes the value of education and what universities should do about it (or not) – 10-15 min max.

https://forms.gle/msWy5jnBWSLN1LWA7



Thank you!

o.kudina@tudelft.nl

olyakudina.com

@OlyaKudina