



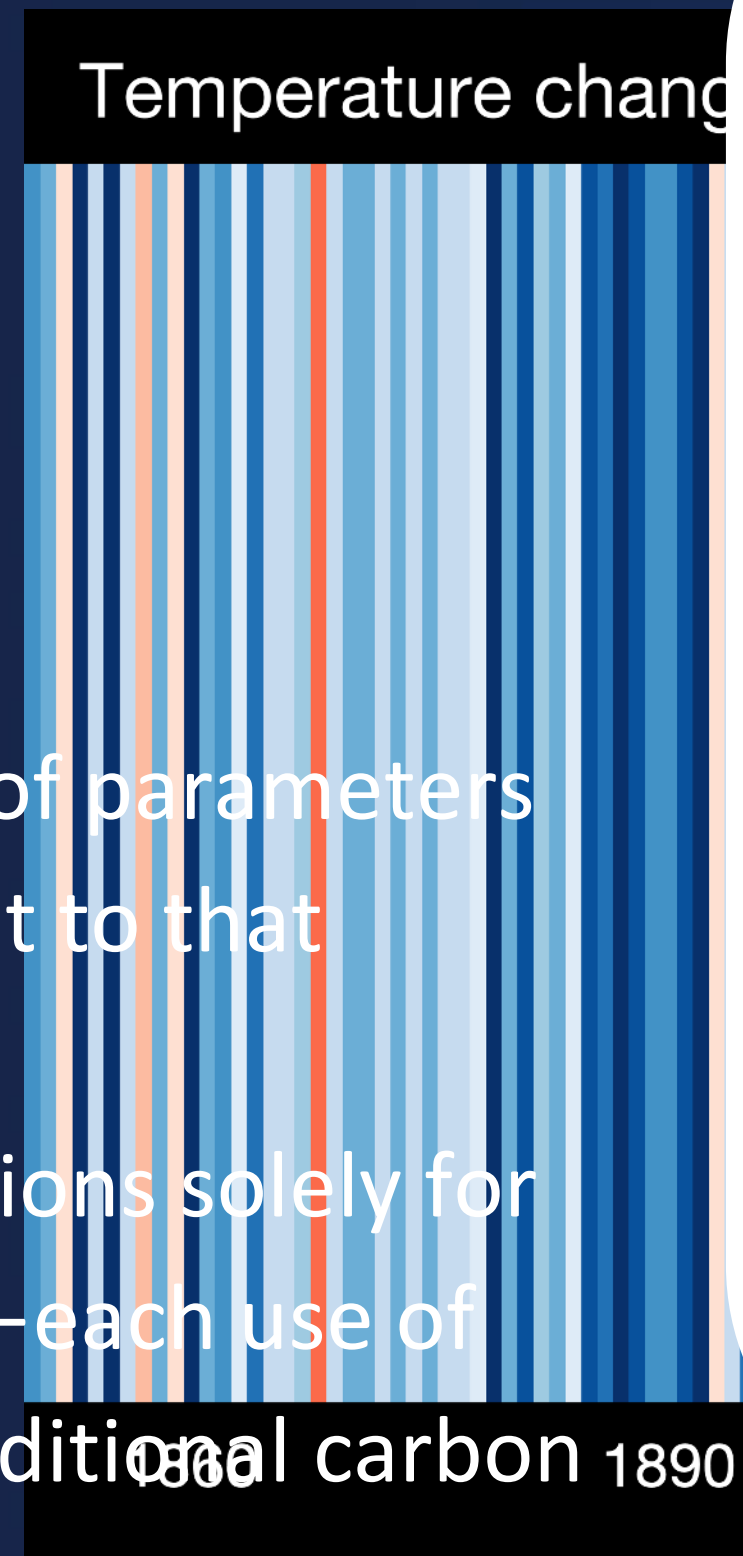
# The ethics of AI and ML

Michelle Fabienne Bieger

7<sup>th</sup> December 2023

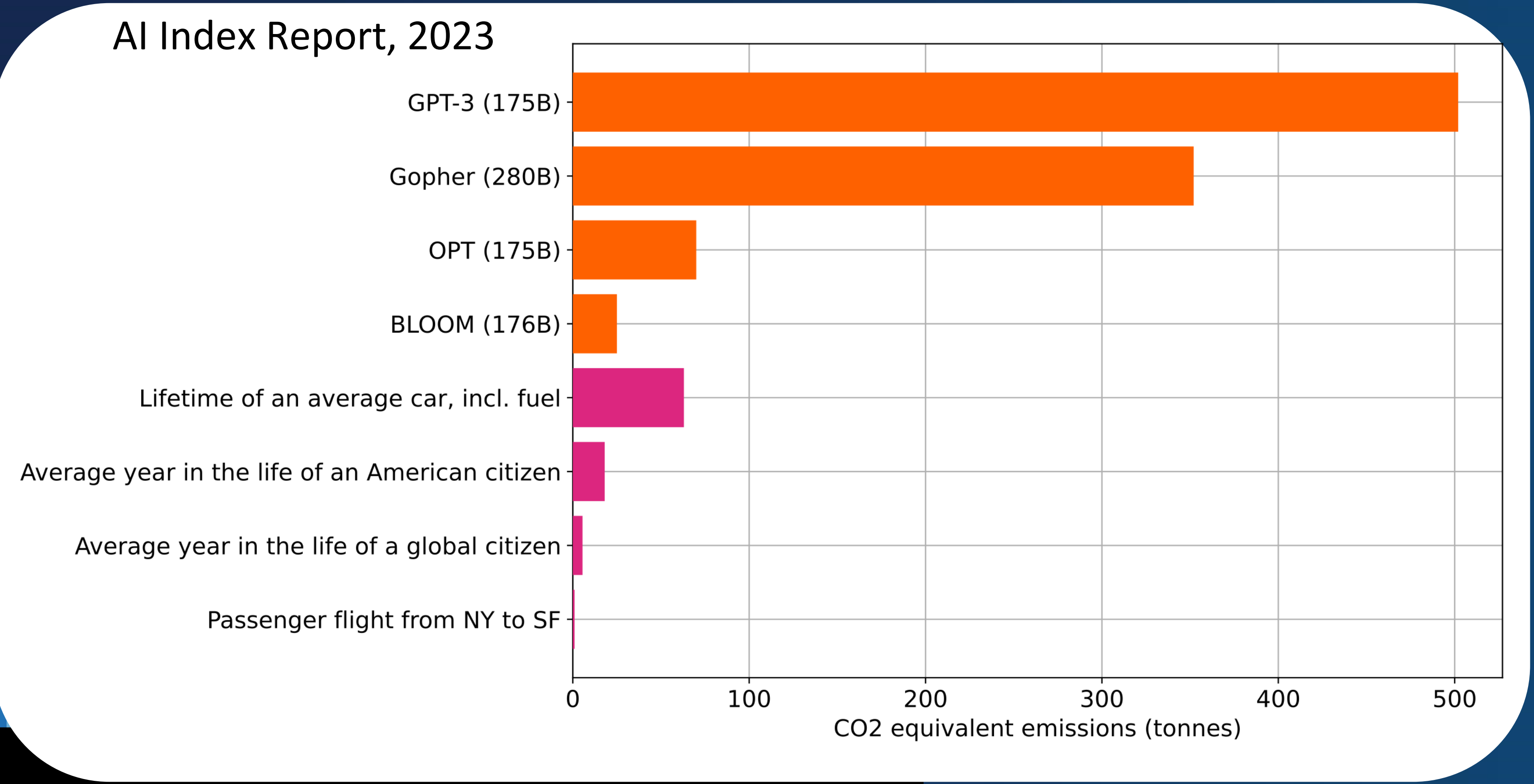
digiLab

# AI and the environment



- B refers to number of parameters that have been input to that training dataset
- These are the emissions solely for training the model—each use of the model incurs additional carbon costs
- Not just limited to carbon costs!

AI Index Report, 2023



# Data provenance

Shadow libraries have formed the training datasets of the biggest names in LLMs—e.g. ChatGPT, Meta’s Llama competitor, Bloomberg’s BloombergGPT (reporting from Wired)

Many of the ML/LLM training datasets have been trained on “the Internet”

What does the Internet look like?



Gender bias:  
Less than 15% of Wikipedia contributors are women  
34% of Twitter users are women  
Only 33% of Reddit users are not men

Language bias:  
Predominant language on the Internet is English



Consent

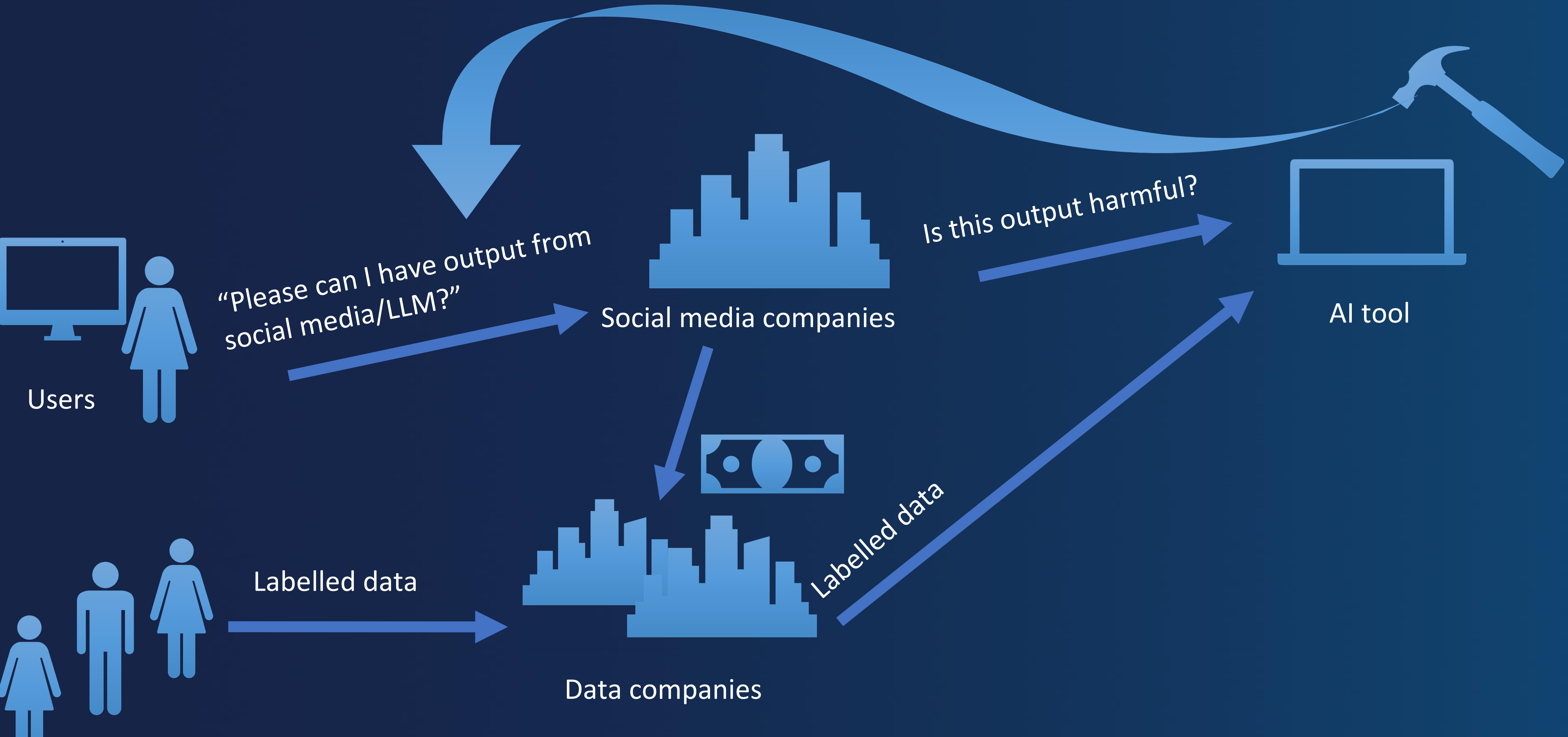


Notice



Regulatory obligation

# Getting data training right



# Case study: Karya—getting training right



Karya: founded by  
Vivek Seshadri and  
Manu Chopra

1

## Living wages

Most data workers are not paid sufficiently for their labour, with many making significantly below the minimum wage. This leaves workers unable to cover their basic living expenses.

Minimum wages sustain a low-income lifestyle, data employment should give a **living wage** to encourage personal and societal growth for workers and their communities.

2

## No data outsourcing!

Many companies win data generation contracts through personal connections without having direct connections to data workers themselves. These companies become middle-men often taking a handsome finders fee. This creates two issues: low data quality and the potential exploitation of data workers. We must work together to shorten data value chains and bring the profits of data to its makers!

3

## Ensuring & investing in data worker welfare

Technology has shown the potential to be transformative and dangerous. Companies must prioritize the safety of data workers over profits, addressing both physical and psychological well-being. To bring the value of data full circle, companies should ensure that the benefits of their technologies extend to all socio-economic communities.

The Ethical Data Pledge

# Case study: Karya—getting training right



# Bias within the model itself



In the US health care system

Black patients were assigned the same level of risk by the algorithm despite being more ill than white patients, **reducing the number of Black patients who identified as needing more care by more than half**

Bias introduced through use of using health care costs as a proxy for health needs—**but this reflects a bias already in-built**, as the system already spends less money on Black patients with the same healthcare needs as their white counterparts

# Over-reliance on AI models

Explainability needs to come with **actionable recourse**

- Allow space in the system for human input to correct or alter AI-generated output
- Automated systems should come with means to identify and address potential errors
- Principle 3 of the Santa Clara moderation principles: companies need to provide *“meaningful opportunity for timely appeal”*



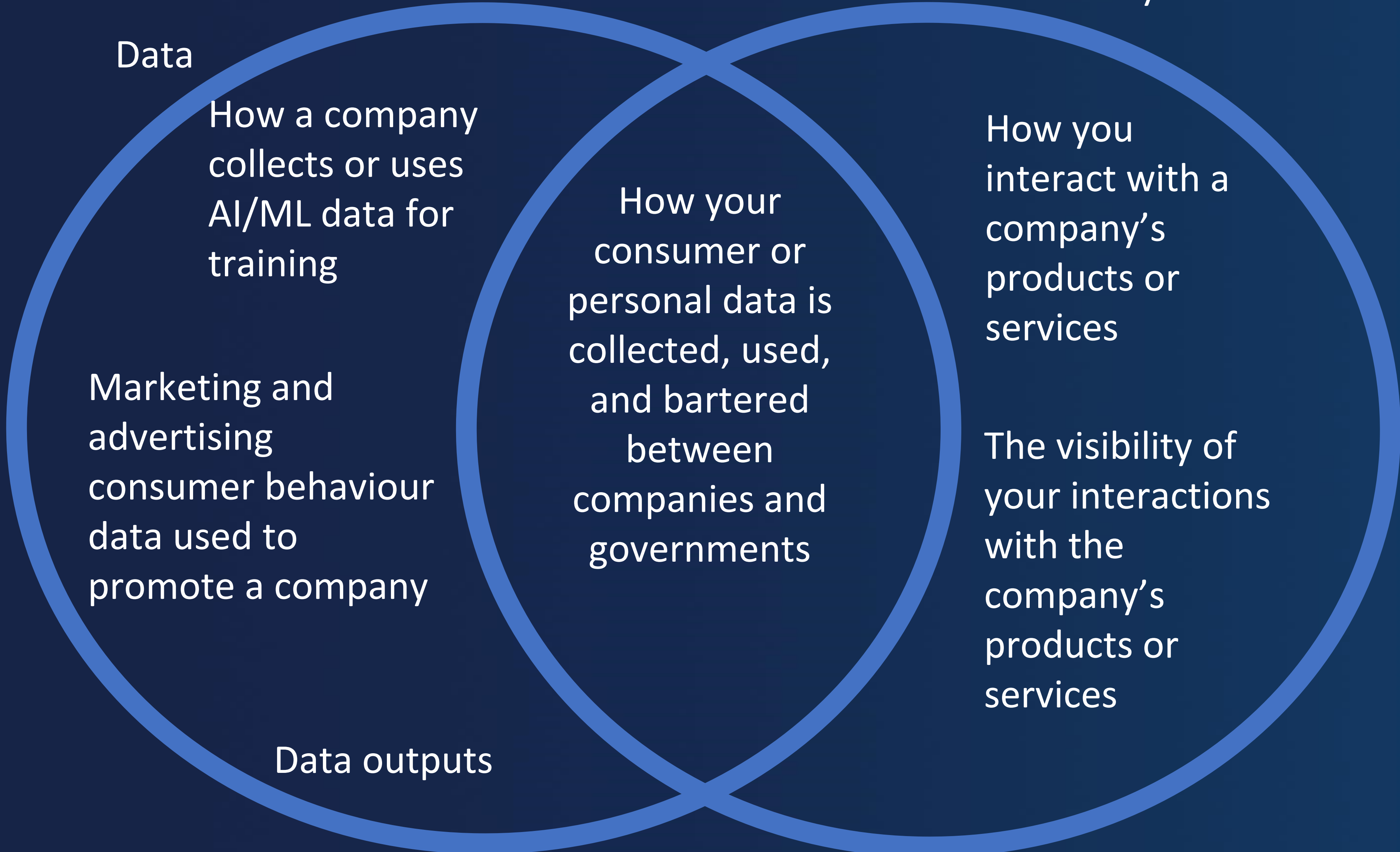
Patient Tammy Dobbs, diagnosed with cerebral palsy

Required 56 hours of care a week, but once automation came in, this was changed to 32 despite no change in her circumstance

It required legal procedures to change this despite attending nurses agreeing she required more hours



# Data and privacy concerns



Three tenants to properly handling data privacy

~~Consent~~

~~Notice~~



Regulatory obligation

# Case study: Roomba

Guo, December 2022, MIT Technology Review

Check your privacy policies—you give permission to Roomba to create a virtual layout of your house + take photos. And that's just your Roomba...

Pictures are sent to outsourced workers who can and do post them online in unsecured company and social media servers



Regulatory obligation

# So what?

Establish business values and proactively interrogate what use cases challenge those values

Like in medicine, consider the use of an ethics board or committee

Lead with solutions that centre humans and humanity and leave room for accountability

Seek out inclusivity and centre diversity in ML teams. More voices at the table mean more POVs

## Stay informed!

Regulation is often behind technology so journalists tend to be the first to interrogate bias and abuses

### Recommendations:

- Billy Perrigo, Karen Hao
- Rumman Chowdhury, Timnit Gebru, Safiya Noble, Joy Buolamwini, Meredith Whittaker, Emily Bender, Margaret Mitchell
- Wired, Time, 404
- The EU AI Act Newsletter, Mystery AI Hype Theatre, All Tech is Human