

Data, Politics and Society

W5 – Data and the Environment



Where we at?

W1

W2

W3

Data: The Good, The Bad, The Ugly

W4

W5

Societal and environmental impacts of data and technology

Hello World

Hello World - Hannah Fry

hannahfry.co.uk/book/hello-world/

HANNAH FRY

Mathematician, science presenter and all-round badass



The book cover for "Hello World" by Hannah Fry is displayed. It features a teal background with yellow text and diagrams. The title "Hello world." is written in large, bold, yellow letters. Above it, there are two yellow boxes connected by arrows: "How to Be Human" and "in the Age of the Machine". Below the title, there is a quote from Alex Bellos: "A gem of accessible science writing. A wise guide to the increasingly data-driven world." The author's name, HANNAH FRY, is at the bottom.

Hello World

How to be Human in the Age of the Machine

Black Swan, 2019

It's a book about how we've slowly handed over control to computers – how there are algorithms and artificial intelligence hiding behind almost every aspect of our modern lives – and what that means for our society.

Cambridge Analytica might have made the most headlines, but these algorithms are everywhere. In our hospitals, our courtrooms, our police stations and our supermarkets. This is a book that takes stock of where we are now, and where we are headed in the not-too-distant future.

Today

- Climate crisis
- Role of data and technology
- Responsible data science

Climate crisis

Climate crisis made summer drought 20 times more likely, scientists find

Record northern hemisphere drought in 2022 hit crops and power stations, worsening food and energy crises

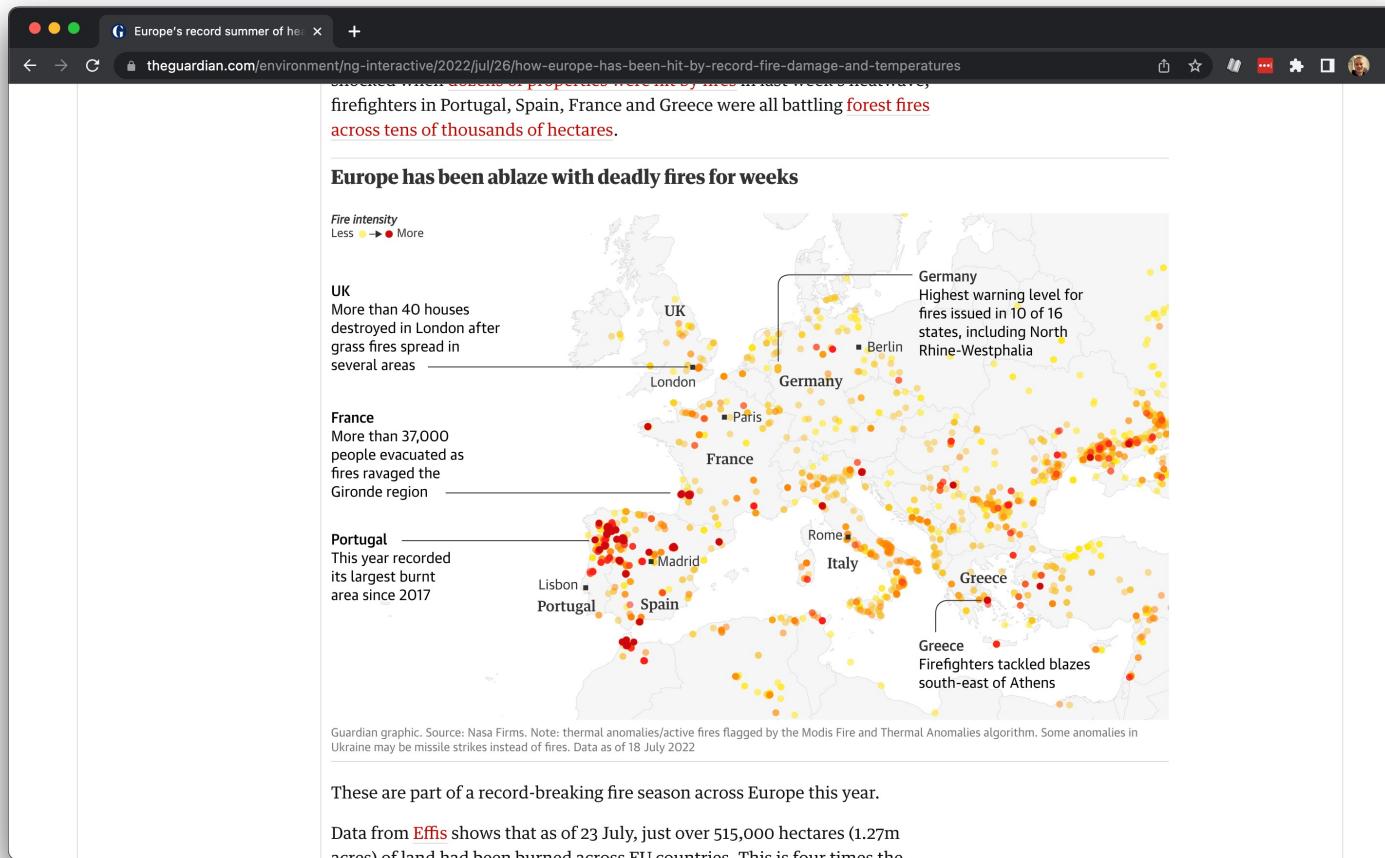


Cracked soil at Alto Rabagao dam in northern Portugal. The dry conditions caused water shortages and wildfires across North America, Europe and Asia. Photograph: Jose Coelho/EPA

The climate crisis made the record drought across the northern hemisphere this summer at least 20 times more likely, scientists have calculated. Without human-caused global heating, the event would have been expected only once every four centuries.

The drought hit crop production and power supplies, exacerbating the food and energy crises already sparked by Russia's war in Ukraine. Droughts will become even more severe and more frequent unless the burning of fossil fuels is phased out, the research found.

Climate crisis



Climate crisis



Climate crisis

Climate crisis makes extreme Indian heatwaves 100 times more likely - study

Latest analysis adds to evidence that the impacts of human-caused global heating are already damaging many lives around the world



New Delhi, India, experiences record-high temperatures. Photograph: Anushree Fadnavis/Reuters

Record-breaking heatwaves in north-west India and [Pakistan](#) have been made 100 times more likely by the climate crisis, according to scientists. The analysis means scorching weather once expected every three centuries is now likely to happen every three years.

The region is currently suffering intense heat, with the Indian capital New [record on Sunday above 49C](#) and the peak temperature in

Waiting for securepubads.g.doubleclick.net...

Climate crisis

G South Africa's April floods made twice as likely by climate crisis, scientists say

the guardian.com/environment/2022/may/13/south-africa-floods-climate-crisis-global-heating

Climate crisis

This article is more than 4 months old

South Africa's April floods made twice as likely by climate crisis, scientists say

Brutal heatwave in India and Pakistan also certain to have been exacerbated by global heating, scientists say

Damian Carrington
Environment editor
@dparrington
Fri 13 May 2022 12.01 BST

f t e



A temple in Chatsworth, outside Durban, was severely damaged by the flooding in April.
Photograph: AP

The massive and deadly floods that struck **South Africa** in April were made twice as likely and more intense by global heating, scientists have calculated. The research demonstrates that the climate emergency is resulting in devastation.

Catastrophic **floods and landslides** hit the South African provinces of KwaZulu-Natal and Eastern Cape on 11 April following exceptionally heavy

Climate crisis



Climate crisis

Storm Babet: second severe flood alert in Scotland after two people die

Red 'danger to life' warning covers Angus and southern Aberdeenshire, as person confirmed killed by falling tree

Storm Babet: latest news updates



00:01:07

Storm Babet: second person dies as substantial flooding hits Scotland – video

The Met Office has issued a second “danger to life” red warning for rain covering the region in eastern [Scotland](#), an area already suffering unprecedented flooding.

The weather agency said the very rare red warning of severe flooding and disruption covered Angus and southern Aberdeenshire and was in place for the whole of Saturday. It came as a second person was confirmed to have been killed after a falling tree hit a van near Forfar on Thursday evening.

<https://www.theguardian.com/uk-news/2023/oct/31/eight-shocking-revelations-from-cummings-and-cain-at-the-covid-inquiry>

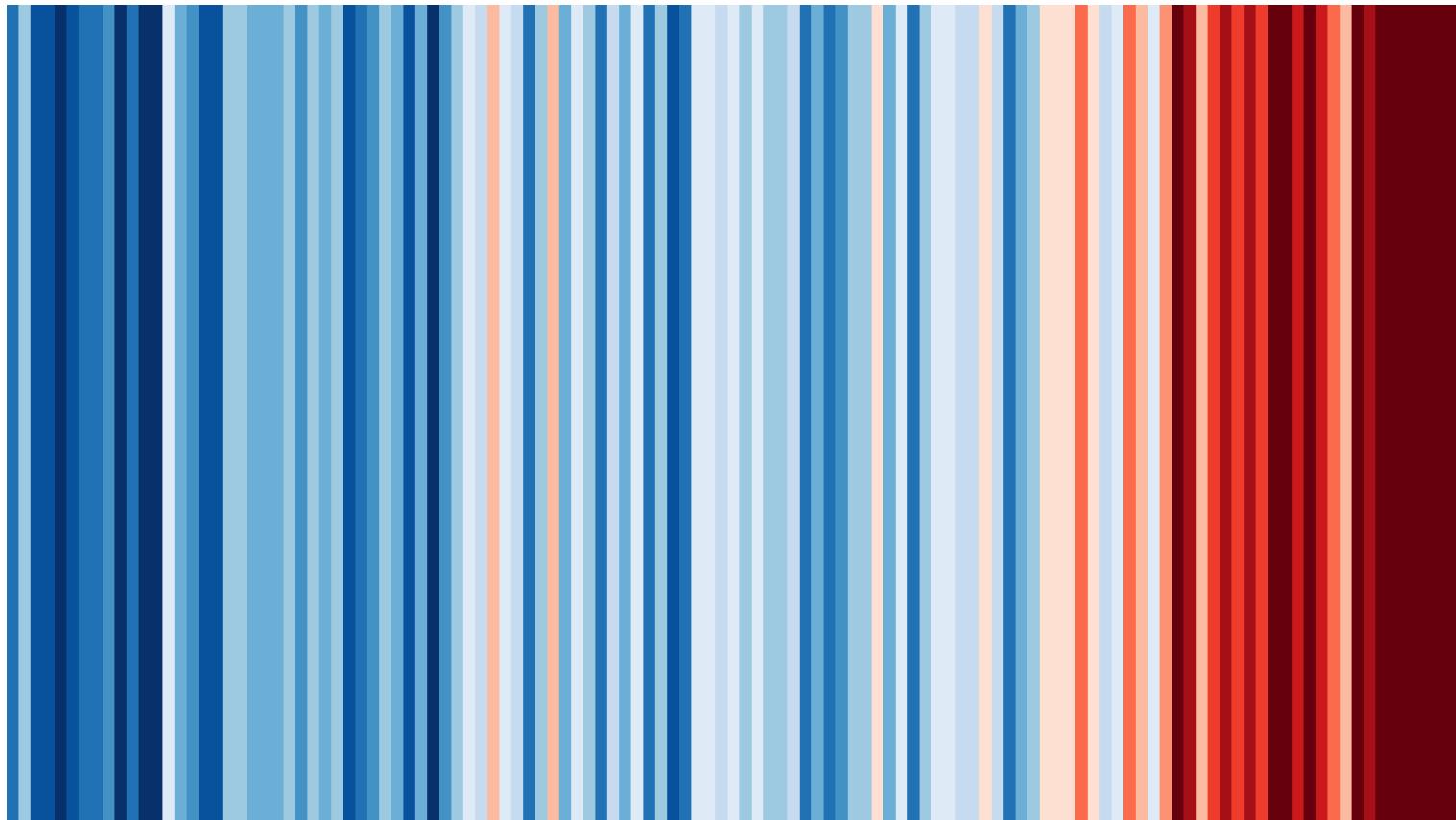
Most viewed

- Eight shocking revelations from Cummings and Cain at the Covid inquiry
- Early retirement in England increasingly preserve of wealthy, report shows
- 'Shame on you!' Robert De Niro shouts at former assistant in court over workplace abuse claims
- My Airbnb host threw me out on the street at midnight
- Genius among morons Dominic Cummings gives Halloween display of his ego *John Crace*

Climate crisis

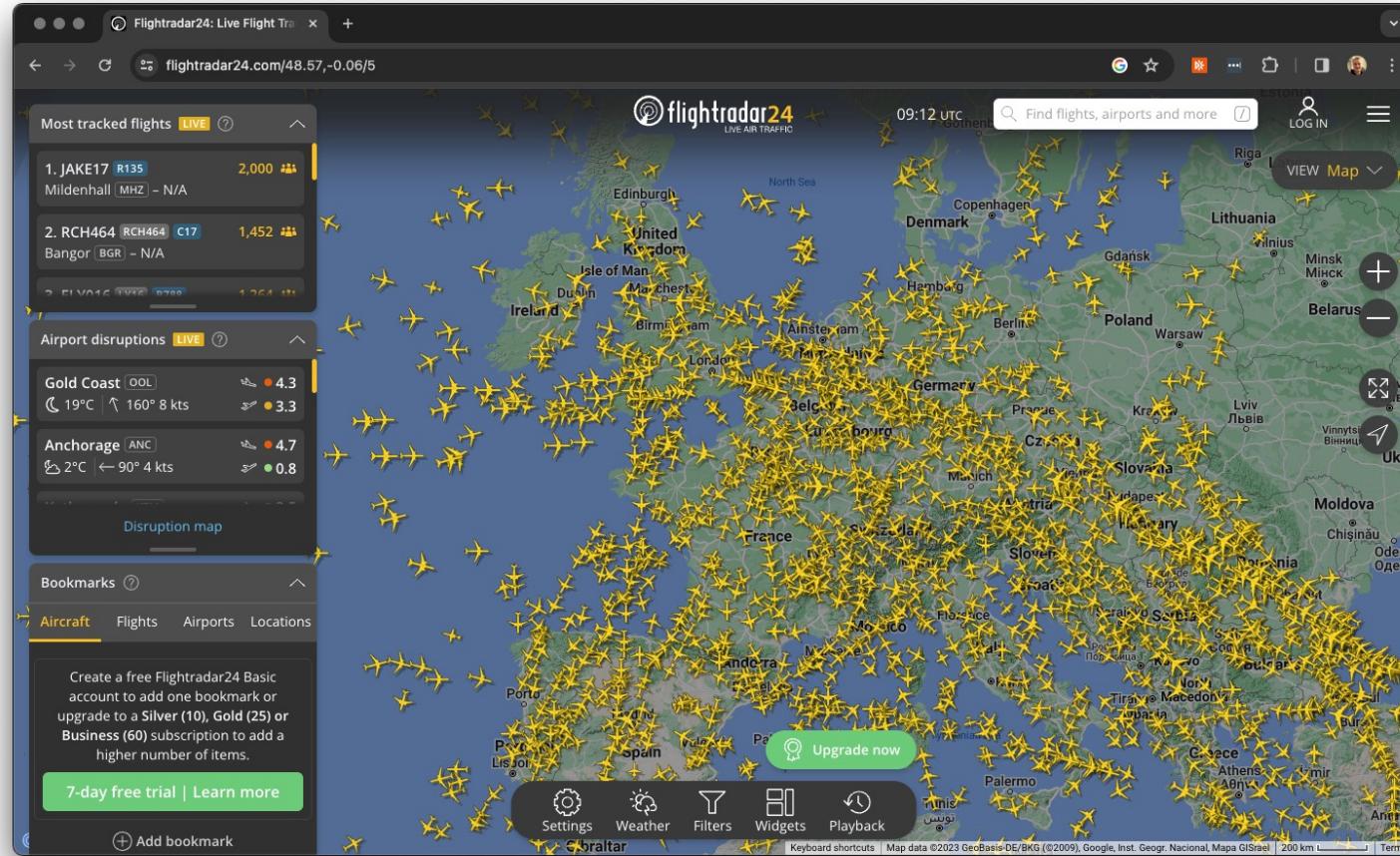


Climate crisis

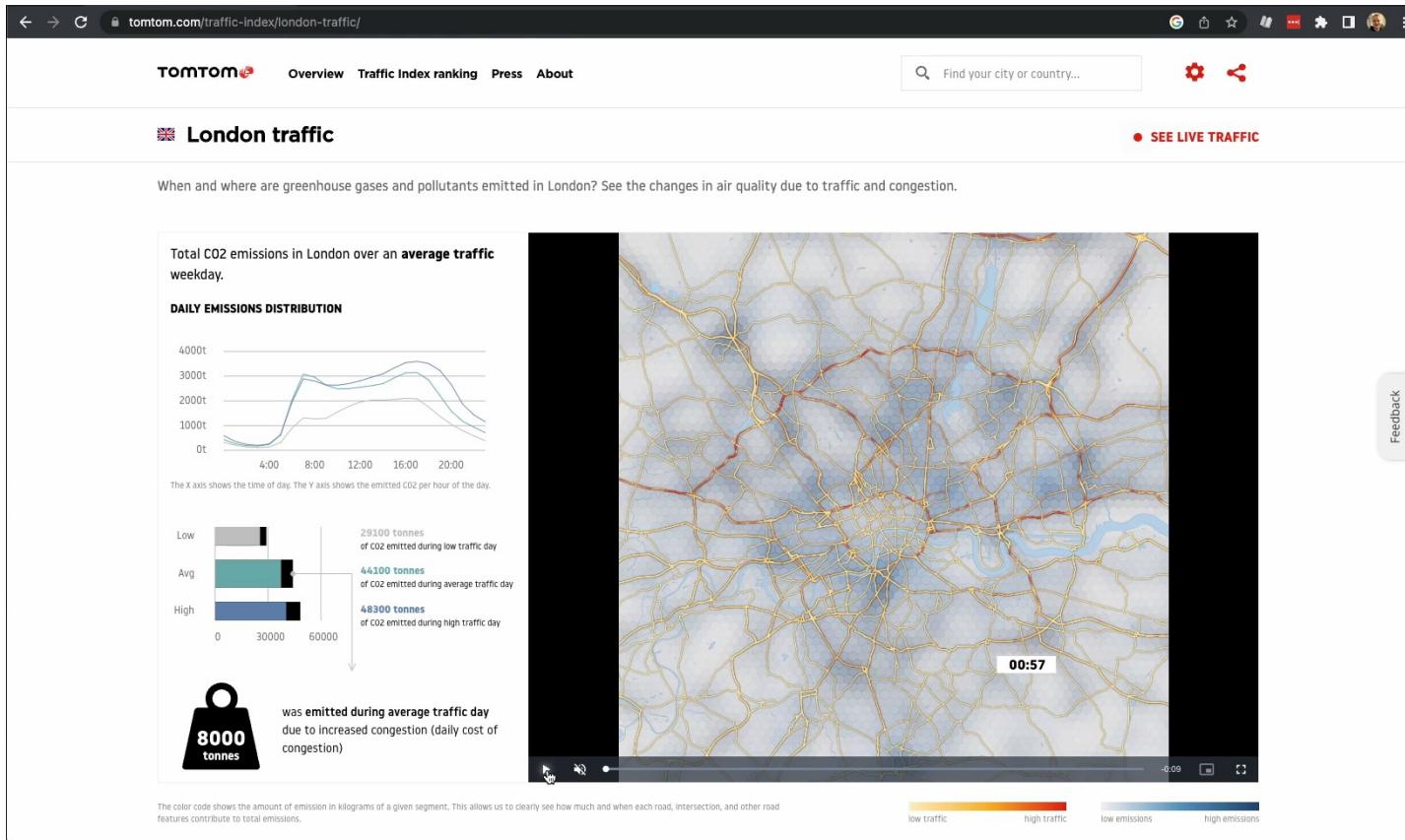


showyourstripes.info, created by Ed Hawkins

Climate crisis?



Climate crisis?



Climate crisis?

The war on motorists: the secret history of a myth as old as cars themselves



Charles Stewart Rolls (back left) in a Rolls Royce in 1905. He founded the car maker the previous year with Henry Royce. Photograph: Hulton Archive/Getty

Local resistance

Ulez expansion: 45% fewer 'dirty' vehicles now on London's roads, says TfL

Sadiq Khan hails 'huge progress' as progress report finds more than 95% of vehicles are now compliant

Ulez: what is it, how much does it cost and why is it so controversial?



The Ulez zone was expanded on 29 August to all of Greater London. Photograph: Leon Neal/Getty Images

The number of the most polluting vehicles driven in London has fallen by almost half since the capital's ultra-low emission zone (Ulez) was expanded, taking almost 80,000 older cars off the roads.

About 77,000, or 45%, fewer non-compliant cars and vans were detected on

Most viewed

- Eight shocking revelations from Cummings and Cain at the Covid inquiry
- Early retirement in England increasingly preserve of wealthy, report shows
- 'Shame on you!' Robert De Niro shouts at former assistant in court over workplace abuse claims
- My Airbnb host threw me out on the street at midnight
- Genius among morons Dominic Cummings gives Halloween display of his ego *John Crace*

Local resistance

Low-traffic schemes are driving congestion and pollution

Hilary Walker, Simon Jones and Lois Keith on how low-traffic neighbourhoods are blighting the lives of many residents. Plus a letter from **Alun Gordon** on how LTNs have helped his east London community



A low-traffic zone in Hackney, east London. Photograph: Graeme Robertson/The Guardian

George Monbiot overplays the delights of the community effect of low-traffic neighbourhoods in Oxford and underplays their disadvantages ([Ignore the culture warriors - low traffic neighbourhoods don't close streets, they liberate them, 3 August](#)). Those living in the arterial roads, which now contain higher volumes of traffic often at a standstill, are experiencing high levels of pollution. The air quality in Hackney is among the worst in the country.

Local resistance

G Spat at, abused and run off the road...
theguardian.com/lifeandstyle/2022/aug/30/why-do-some-people-hate-cyclists-so-much

Cycling



Helen Pidd

Tue 30 Aug 2022 05.57 BST

f t e

**Spat at, abused and run off the road:
why do some people hate cyclists so
much?**

Why do drivers not thank us? ... Helen Pidd cycling in Manchester city centre. Photograph: Christopher Thomond/The Guardian

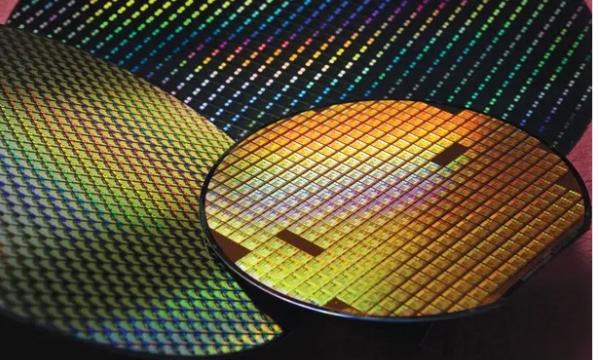
Data and technology

Although the transport sector is still one of the larger contributors to greenhouse gas emissions, the climate impact of information and communication technologies has been rapidly increasing.

Data and technology

The computer chip industry has a dirty climate secret

As demand for chips surges, the semiconductor industry is trying to grapple with its huge carbon foot print



A TSMC chip wafer. The semiconductor industry is starting to reckon with its big climate footprint. Photograph: Taiwan Semiconductor Manufacturing Company

The semiconductor industry has a problem. Demand is booming for silicon chips, which are embedded in everything from smartphones and televisions to wind turbines, but it comes at a big cost: a huge carbon footprint.

The industry presents a paradox. Meeting global climate goals will, in part, rely on semiconductors. They're integral to electric vehicles, solar arrays and wind turbines. But chip manufacturing also contributes to the climate crisis.

Data and technology

The screenshot shows a web browser displaying an article from the journal *Scientific Reports*. The article title is "Economic estimation of Bitcoin mining's climate damages demonstrates closer resemblance to digital crude than digital gold". It was published on 29 September 2022, is open access, and is available in [PDF](#). The abstract discusses the economic estimates of climate damages from Bitcoin mining, comparing it to digital crude oil.

Abstract

This paper provides economic estimates of the energy-related climate damages of mining Bitcoin (BTC), the dominant proof-of-work cryptocurrency. We provide three sustainability criteria for signaling when the climate damages may be unsustainable. BTC mining fails all three. We find that for 2016–2021: (i) per coin climate damages from BTC were increasing, rather than decreasing with industry maturation; (ii) during certain time periods, BTC climate damages exceed the price of each coin created; (iii) on average, each \$1 in BTC market value created was responsible for \$0.35 in global climate damages, which as a share of market value is in the range between beef production and crude oil burned as gasoline, and an order-

View all journals | Search | Login | Sign up for alerts | RSS feed

Download PDF

Sections | Figures | References

Abstract | Introduction | Results | Discussion | Methods | Data availability | References | Acknowledgements | Author information | Ethics declarations | Additional information | Supplementary Information

Data and technology

- Natural resources
- Data centres responsible for data storage and processing

Natural resources

Revealed: The shocking amount x +
independent.co.uk/climate-change/news/revealed-the-shocking-amount-of-resources-needed-to-make-products-such-as-mobile-phones-coffee-and-ts...

**Revealed: The shocking amount of
resources needed to make products
such as mobile phones, coffee and T-
shirts**

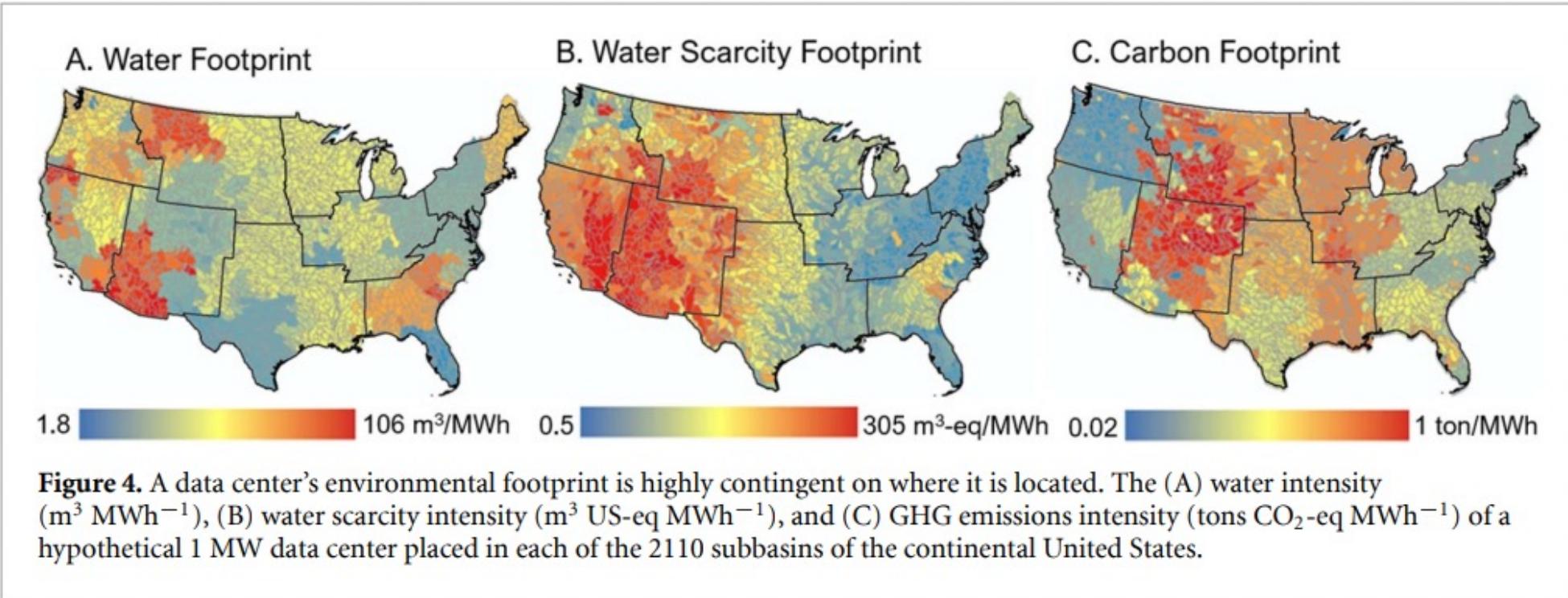
Smartphones, the report claimed, use nearly 13 tons of water when being manufactured

Tom Bawden • Thursday 07 May 2015 20:21 • [Comments](#) 


Data centres

- Siddik *et al.* 2021
- Data are stored in large data centers.
- Large amounts of energy required: 1.8% of electricity use in the US.
- Large amounts of greenhouse gas emissions: 0.5% of total emissions in US.
- Energy demand is between 15-100 times as large as those of typical commercial buildings.

Data centres



Data centres

Siddik *et al.* 2021:

"Though the amount of data center computing workloads has increased nearly 550% between 2010 and 2018, data center electricity consumption has only risen by 6% due to dramatic improvements in energy efficiency and storage-drive density across the industry. However, it is unclear whether energy efficiency improvements can continue to offset the energy demand of data centers as the industry is expected to continue its rapid expansion over the next decade."

Data centres

- Current estimate that the tech sector contribute around 3% of global greenhouse, of which around 45% can be attributed to data centers.
- At the same time: lack of transparency on the share of green and dirty energy being used.
- Some estimates suggest that the wider tech industry's carbon footprint could increase to 14% of global emissions by 2040; think crypto currency, energy intensive 5G networks, autonomous vehicles, Internet of Things.

Data centres

But it is not only the storage and access of data; data processing with **data hungry** and computationally intensive models is having a growing influence.

Data centres

- Strübel *et al.* 2019
- Training a state-of-the-art model now requires substantial computational resources which demand considerable energy.
- Research and development of new models multiplies these costs by thousands of times by requiring **retraining to experiment** with model architectures and hyperparameters.
- Estimated carbon footprint from training a large Natural Language Processing model: 300,000 kilograms.

Data centres

Consumption	CO₂e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
<hr/> Training one model (GPU) <hr/>	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

Table 1: Estimated CO₂ emissions from training common NLP models, compared to familiar consumption.¹

Data centres

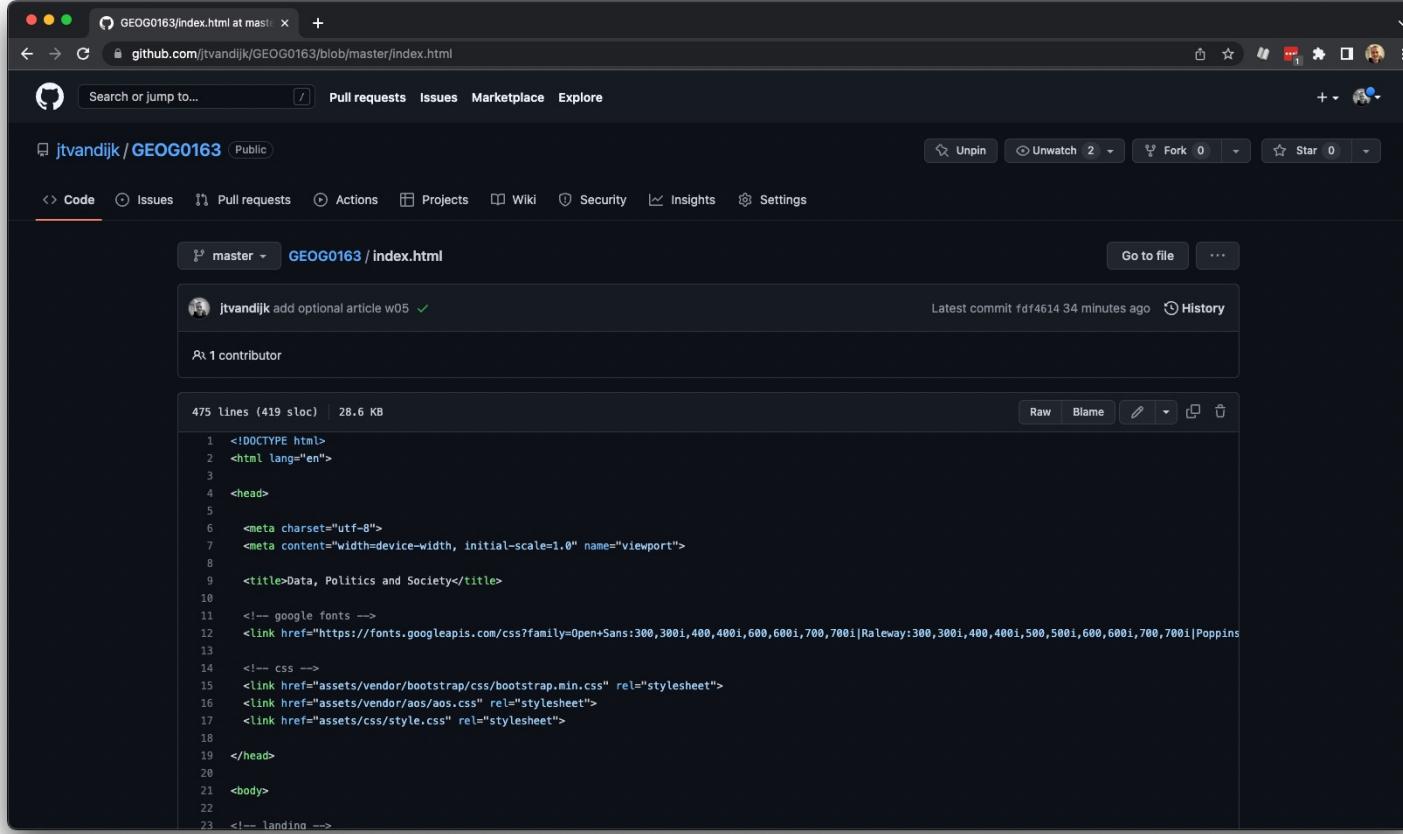
Consumer	Renew.	Gas	Coal	Nuc.
China	22%	3%	65%	4%
Germany	40%	7%	38%	13%
United States	17%	35%	27%	19%
Amazon-AWS	17%	24%	30%	26%
Google	56%	14%	15%	10%
Microsoft	32%	23%	31%	10%

Table 2: Percent energy sourced from: Renewable (e.g. hydro, solar, wind), natural gas, coal and nuclear for the top 3 cloud compute providers (Cook et al., 2017), compared to the United States,⁴ China⁵ and Germany (Burger, 2019).

Data processing tools

- Increased accessibility in terms of using cloud-distributed computing and move away from laptop/desktop-based calculations.
- Lots of tools are part of a typical data science workflow.

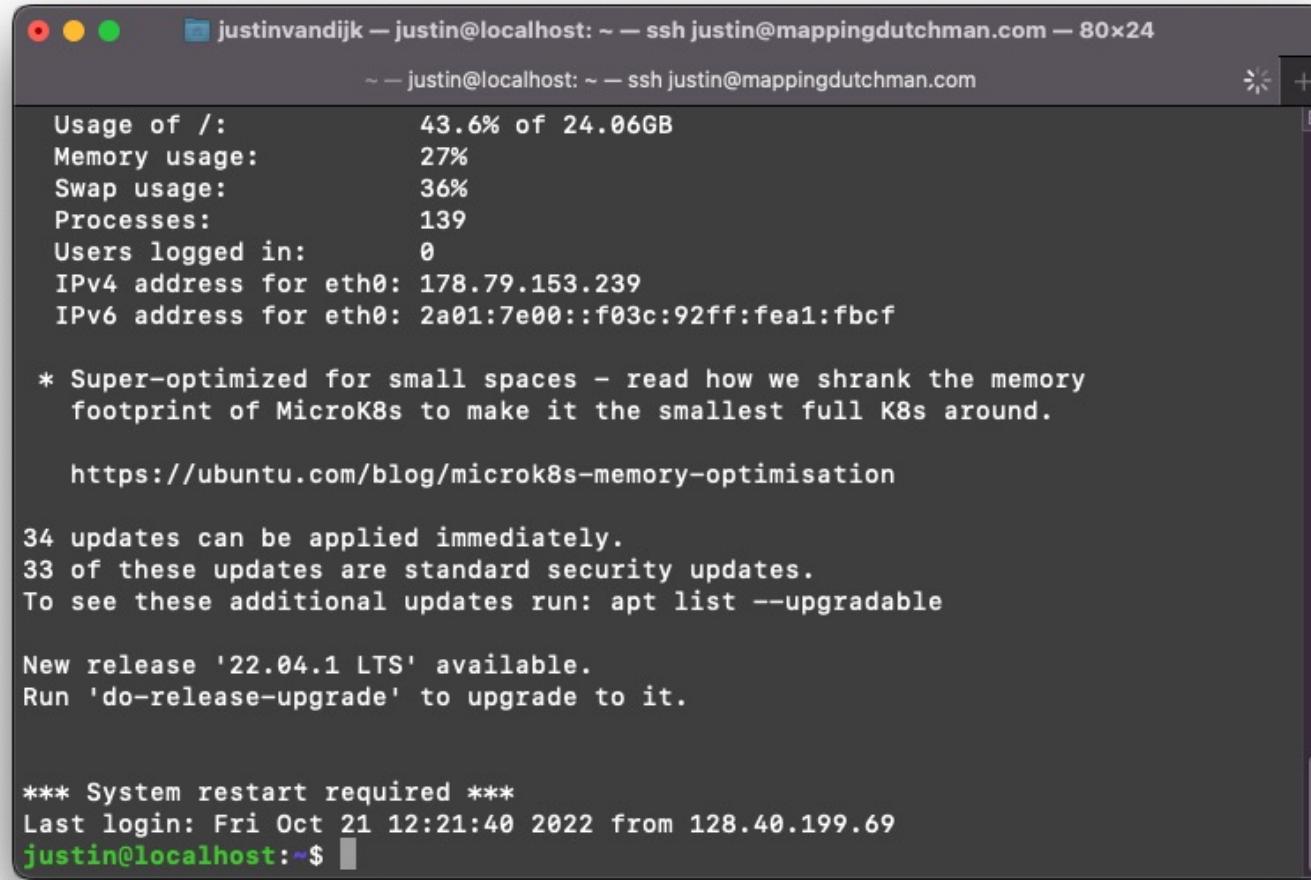
Data processing tools



A screenshot of a GitHub repository page for 'jtvdnijk / GEOG0163'. The repository is public and contains one file, 'index.html'. The file has 475 lines (419 sloc) and is 28.6 KB in size. The code is displayed in a monospaced font, showing HTML and CSS. The commit history shows a single commit from 'jtvdnijk' adding an optional article w05, with a timestamp of 34 minutes ago.

```
1 <!DOCTYPE html>
2 <html lang="en">
3
4 <head>
5
6 <meta charset="utf-8">
7 <meta content="width=device-width, initial-scale=1.0" name="viewport">
8
9 <title>Data, Politics and Society</title>
10
11 <!-- google fonts -->
12 <link href="https://fonts.googleapis.com/css?family=Open+Sans:300,300i,400,400i,600,600i,700,700i|Raleway:300,300i,400,400i,500,500i,600,600i,700,700i|Poppins" rel="stylesheet">
13
14 <!-- css -->
15 <link href="assets/vendor/bootstrap/css/bootstrap.min.css" rel="stylesheet">
16 <link href="assets/vendor/aos/aos.css" rel="stylesheet">
17 <link href="assets/css/style.css" rel="stylesheet">
18
19 </head>
20
21 <body>
22
23 <!-- landing -->
```

Data processing tools



A screenshot of a macOS terminal window titled "justinvandijk — justin@localhost: ~ — ssh justin@mappingdutchman.com — 80x24". The window shows system status and upgrade information.

```
Usage of /:           43.6% of 24.06GB
Memory usage:        27%
Swap usage:          36%
Processes:           139
Users logged in:    0
IPv4 address for eth0: 178.79.153.239
IPv6 address for eth0: 2a01:7e00::f03c:92ff:fea1:fbcf

* Super-optimized for small spaces - read how we shrank the memory
  footprint of MicroK8s to make it the smallest full K8s around.

https://ubuntu.com/blog/microk8s-memory-optimisation

34 updates can be applied immediately.
33 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

New release '22.04.1 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

*** System restart required ***
Last login: Fri Oct 21 12:21:40 2022 from 128.40.199.69
justin@localhost:~$
```

Data processing tools

The screenshot shows a Jupyter Notebook interface titled "arch.ipynb" running on Google Colab. The notebook content is as follows:

- GPU Architecture**: A section containing text about GPU performance, architecture comparison, and a note about the Tesla T4 GPU.
- Terminal Output**: A command-line interface showing the output of the command `!nvidia-smi -q -i 0 | grep "Product Name"`, which returns "Product Name : Tesla T4".
- Streaming Multiprocessor**: A diagram illustrating the internal structure of an SM. It shows four identical blocks, each containing:
 - A "Fetch / Decoder" unit.
 - Four "FP32" units labeled "x16".
 - Two "Tensor Core" units labeled "x2".
 - "Registers" at the bottom.Below these blocks is a horizontal bar labeled "(GKGD)".

What can we do?

What else can we do in a research context to reduce our environmental impact?

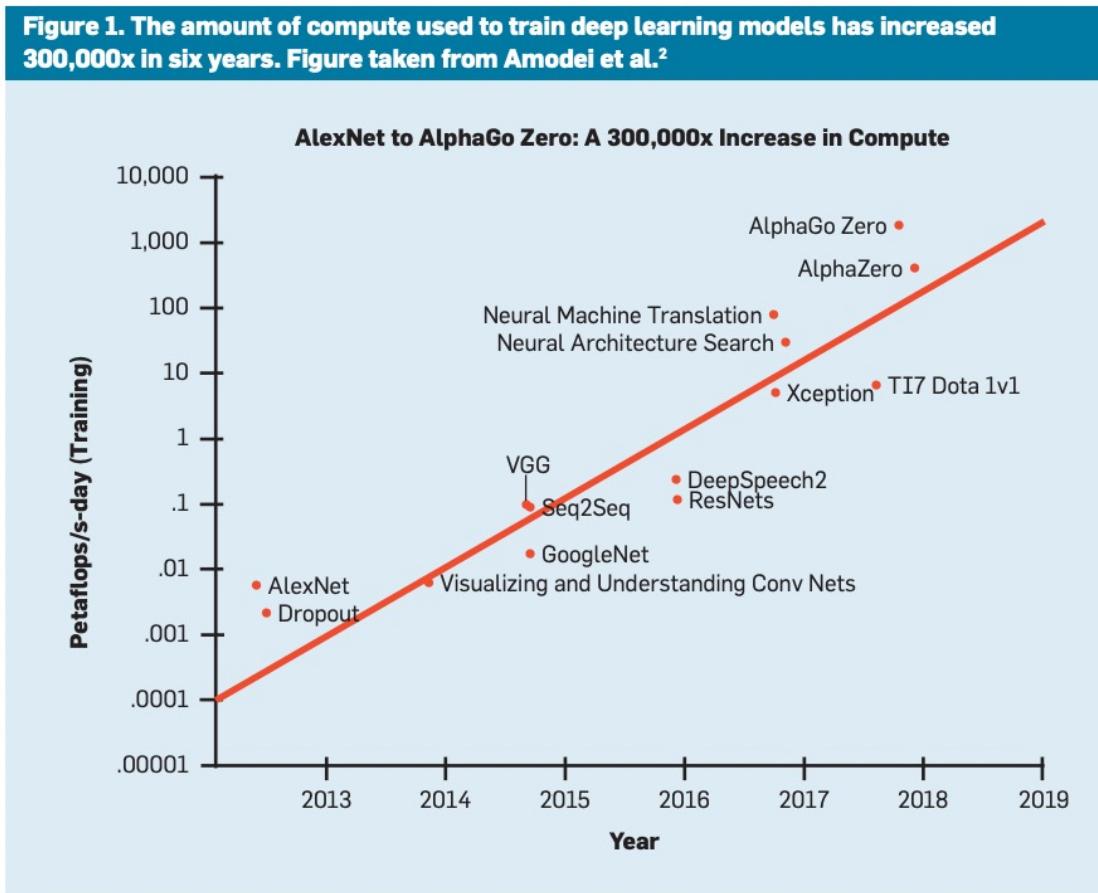
What can we do?

- Strübel *et al.* 2019
- Authors should report training time and sensitivity to hyperparameters.
- Academic researchers need equitable access to computation resources.
- Researchers should prioritize computationally efficient hardware and algorithms.

What can we do?

- Improvements in the field of AI: object recognition, game playing, speech recognition, and machine translation.
- The computational costs of state-of-the art AI research has increased 300,000x recent years: prioritising accuracy over efficiency. Red AI.

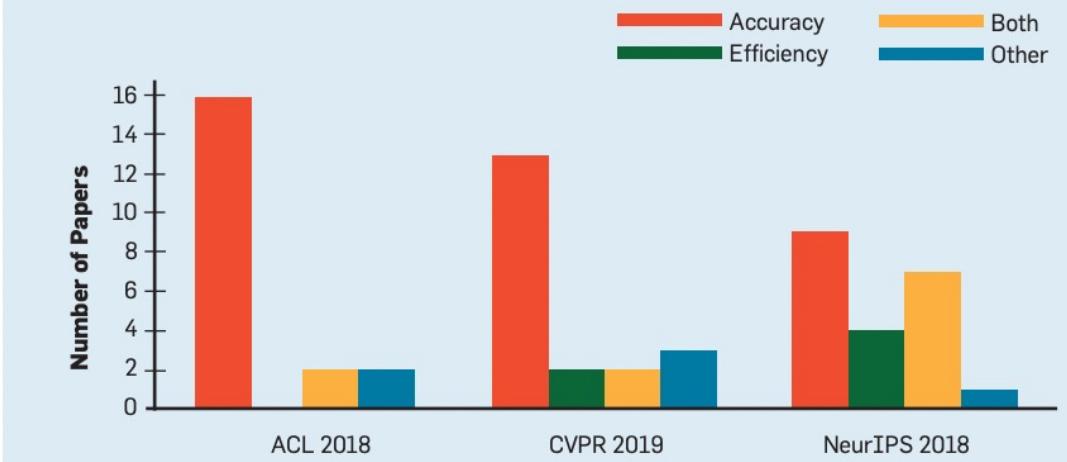
What can we do?



Schwartz *et al.* 2020

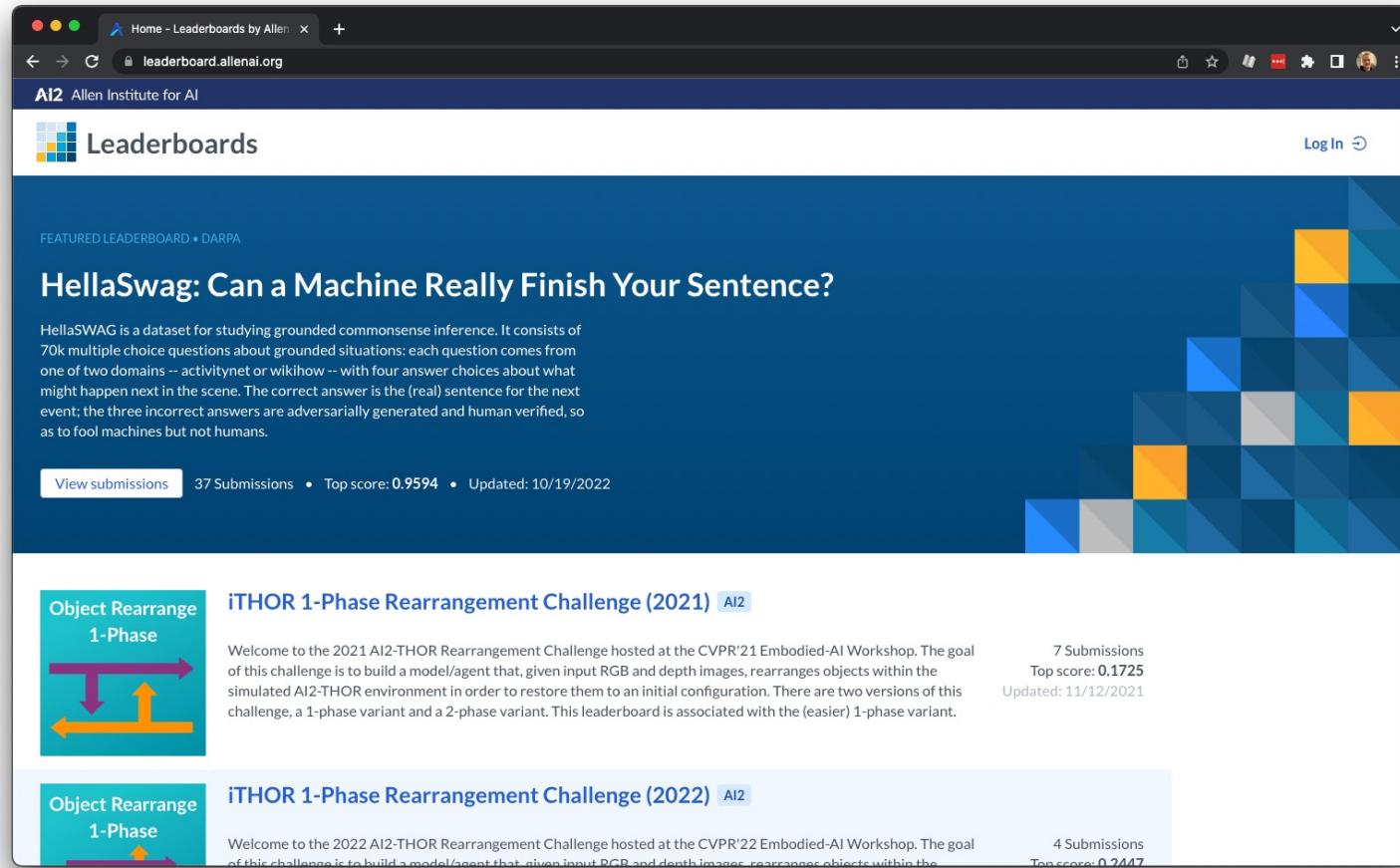
What can we do?

Figure 2. AI papers tend to target accuracy rather than efficiency. The figure shows the proportion of papers that target accuracy, efficiency, both or other from a random sample of 60 papers from top AI conferences.



Schwartz *et al.* 2020

What can we do?



The screenshot shows the AI2 Leaderboards by Allen website. The top section features the "HellaSwag: Can a Machine Really Finish Your Sentence?" challenge, which is a featured DARPA leaderboard. It includes a brief description of the dataset, a "View submissions" button, and statistics: 37 Submissions, Top score: 0.9594, Updated: 10/19/2022. To the right is a decorative graphic of colored squares (blue, yellow, grey) arranged in a grid pattern. Below this are two more challenges: "iTHOR 1-Phase Rearrangement Challenge (2021)" and "iTHOR 1-Phase Rearrangement Challenge (2022)". Each challenge has a small icon illustrating the task (two objects being rearranged), a title, a brief description, and submission statistics.

HellaSwag: Can a Machine Really Finish Your Sentence?

HellaSWAG is a dataset for studying grounded commonsense inference. It consists of 70k multiple choice questions about grounded situations: each question comes from one of two domains -- activitynet or wikihow -- with four answer choices about what might happen next in the scene. The correct answer is the (real) sentence for the next event; the three incorrect answers are adversarially generated and human verified, so as to fool machines but not humans.

[View submissions](#) • 37 Submissions • Top score: 0.9594 • Updated: 10/19/2022

iTHOR 1-Phase Rearrangement Challenge (2021)

Welcome to the 2021 AI2-THOR Rearrangement Challenge hosted at the CVPR'21 Embodied-AI Workshop. The goal of this challenge is to build a model/agent that, given input RGB and depth images, rearranges objects within the simulated AI2-THOR environment in order to restore them to an initial configuration. There are two versions of this challenge, a 1-phase variant and a 2-phase variant. This leaderboard is associated with the (easier) 1-phase variant.

7 Submissions
Top score: 0.1725
Updated: 11/12/2021

iTHOR 1-Phase Rearrangement Challenge (2022)

Welcome to the 2022 AI2-THOR Rearrangement Challenge hosted at the CVPR'22 Embodied-AI Workshop. The goal of this challenge is to build a model/agent that, given input RGB and depth images, rearranges objects within the

4 Submissions
Top score: 0.2417

What can we do?

- Transfer learning
- Efficiency of coding language
- Be aware of diminishing returns
- Together: Green AI?

How to measure efficiency?

- Carbon emissions
- Electricity usage
- Elapsed real time
- Number of parameters
- Reporting on number of floating-point operations
- Workflow?

What can we do?

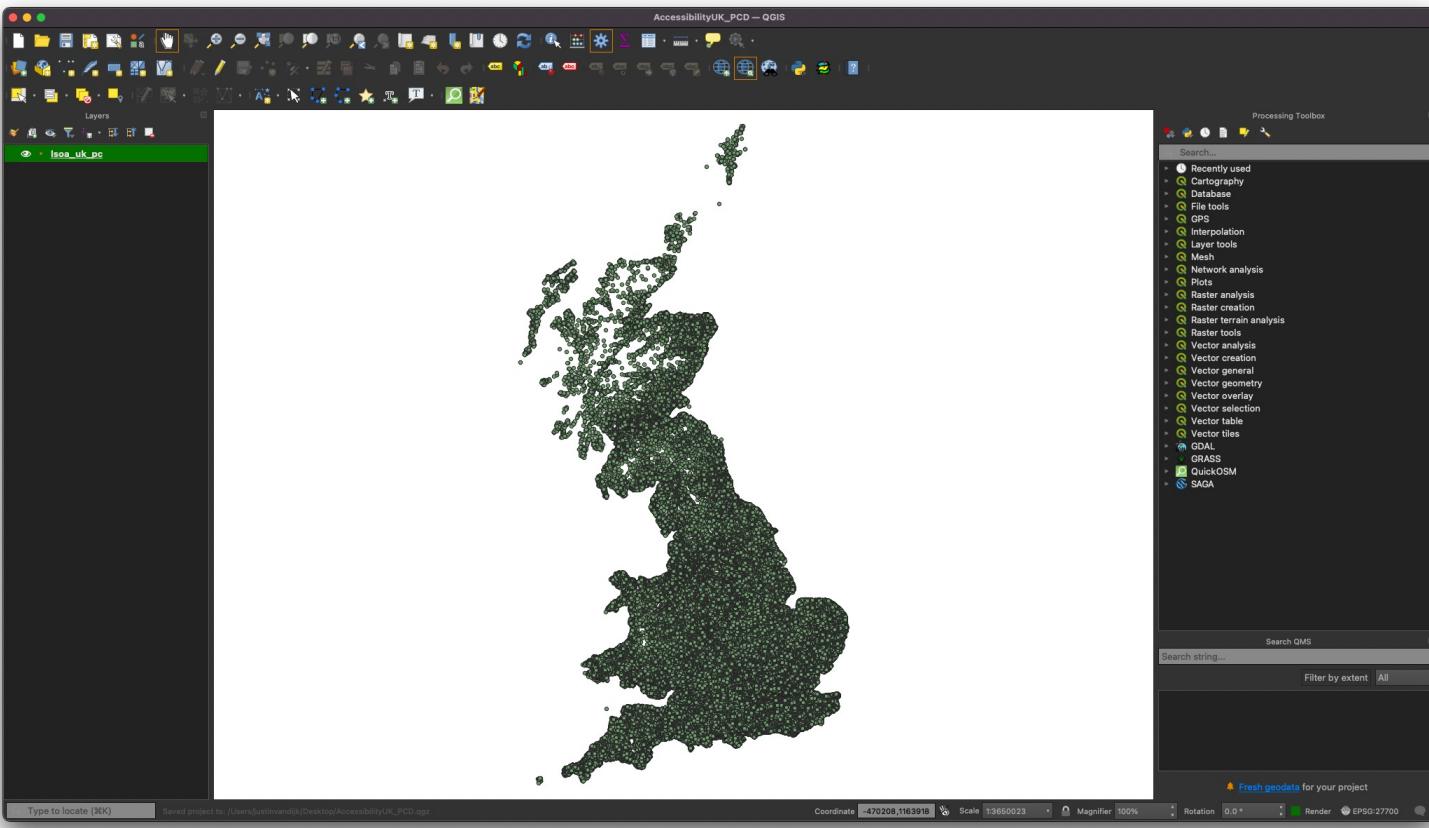
Task

- For each Output Area in Great Britain, calculate the average accessibility by unit postcode to the nearest supermarkets or vegetable shops with fresh foods.

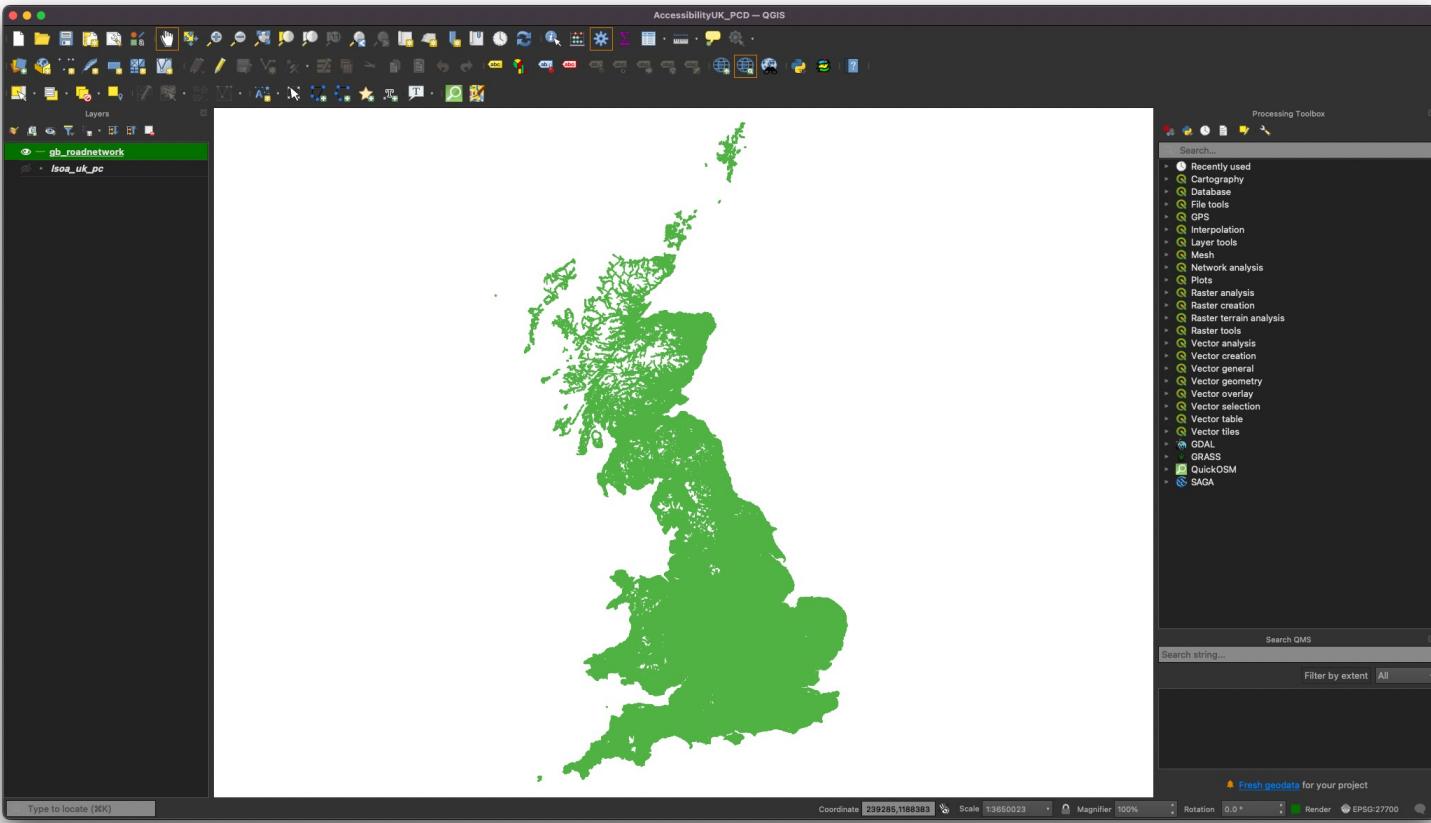
Data

- 2.5 million postcodes in Great Britain
- 3.7 million individual road segments in Great Britain

What can we do?



What can we do?



Conclusion

- The climate crisis is real and it is here.
- Massive impact of the "information society" not only in terms of natural resources needed for chips and infrastructure, but also in terms of electricity and water required for data storage and processing – heavy carbon footprint.
- Responsibility as a data scientist / researcher to maximise efficiency of analytical work; even at small scale simple strategies can be employed.

Recap

- Part I: Data and its role in Society.
 - Mainly: raised issues and concerns, “taking stock”.
 - Fundamental questions: what are data and how can they be used?
-
- Part II: Mitigating the risks of working with large-scale datasets.
 - Mainly: what are “we” currently doing to address some of these issues and concerns.
 - Fundamental question: is what “we are doing” enough?

Seminar preparation

This week we will be paying attention to the *essay outline* part of your first coursework assignment through a peer-feedback assignment. For this Friday this means that everyone should bring their draft essay topic outline (either printed or on a digital device). This should include the main research question or aim and some bullet points on how you intend to address the question or aim.

This is an excellent opportunity to get some feedback – the better your draft outline, the more useful your feedback will be.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk



Enjoy reading week

