University of BRISTOL

UNDERSTANDING THE DYNAMICS OF THE ENVIRONMENTAL IMPACTS OF AI

Chris Preist

Professor of Sustainability and Computer Systems,

University of Bristol

bristol.ac.uk

“Digital technologies impact positively and negatively on GHG emissions through:

their own carbon footprint; technology application for mitigation; and induced larger social change.”

(IPCC 6AR WGIII Box 11, 2022)

Systemic

‘Induced Social Change’

Enabling

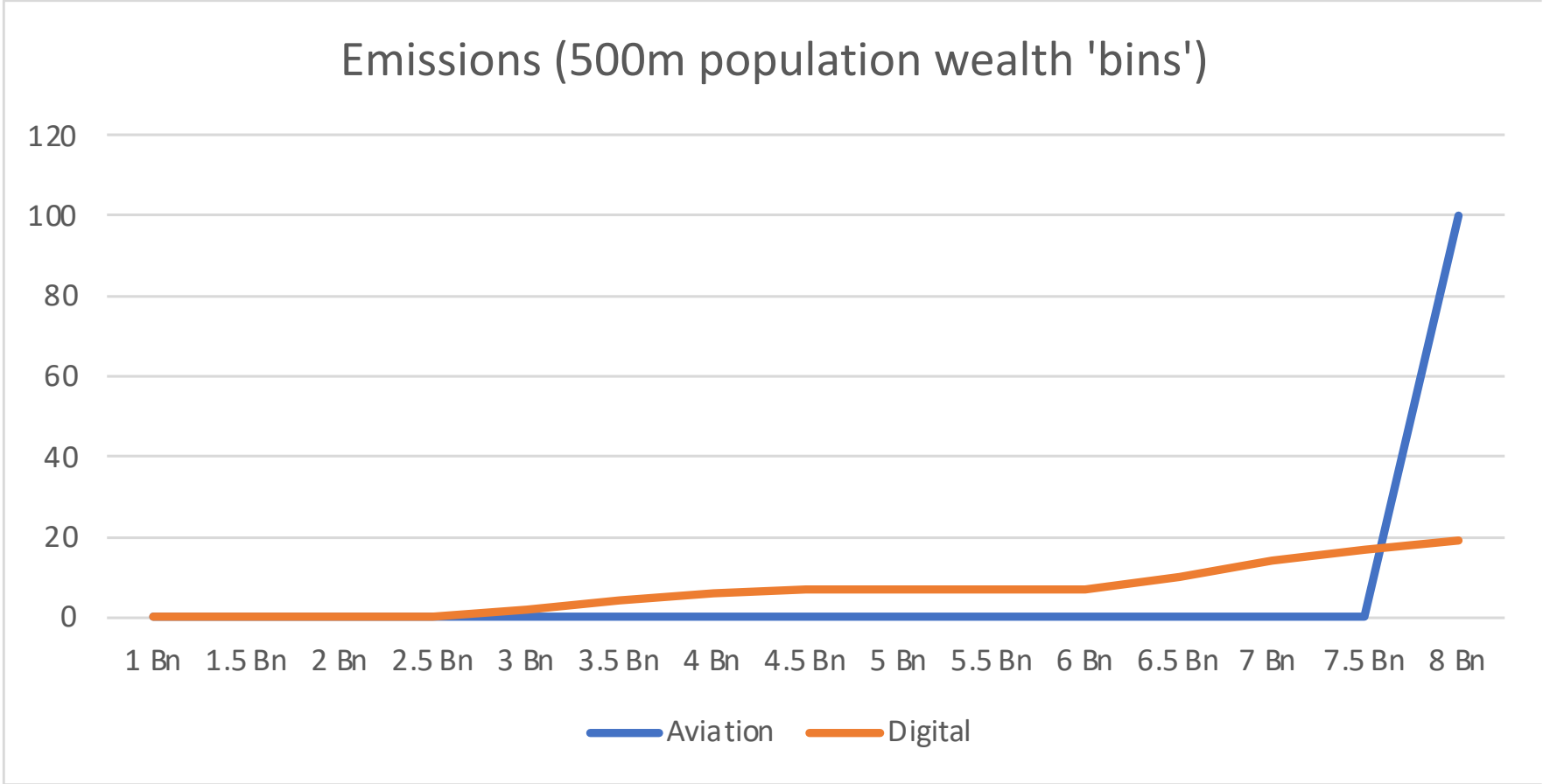
‘Technology application’

Direct

‘Their own footprint.’

The Big Picture – Global Direct Emissions

Each is roughly 2.5% of Global CO2 emissions

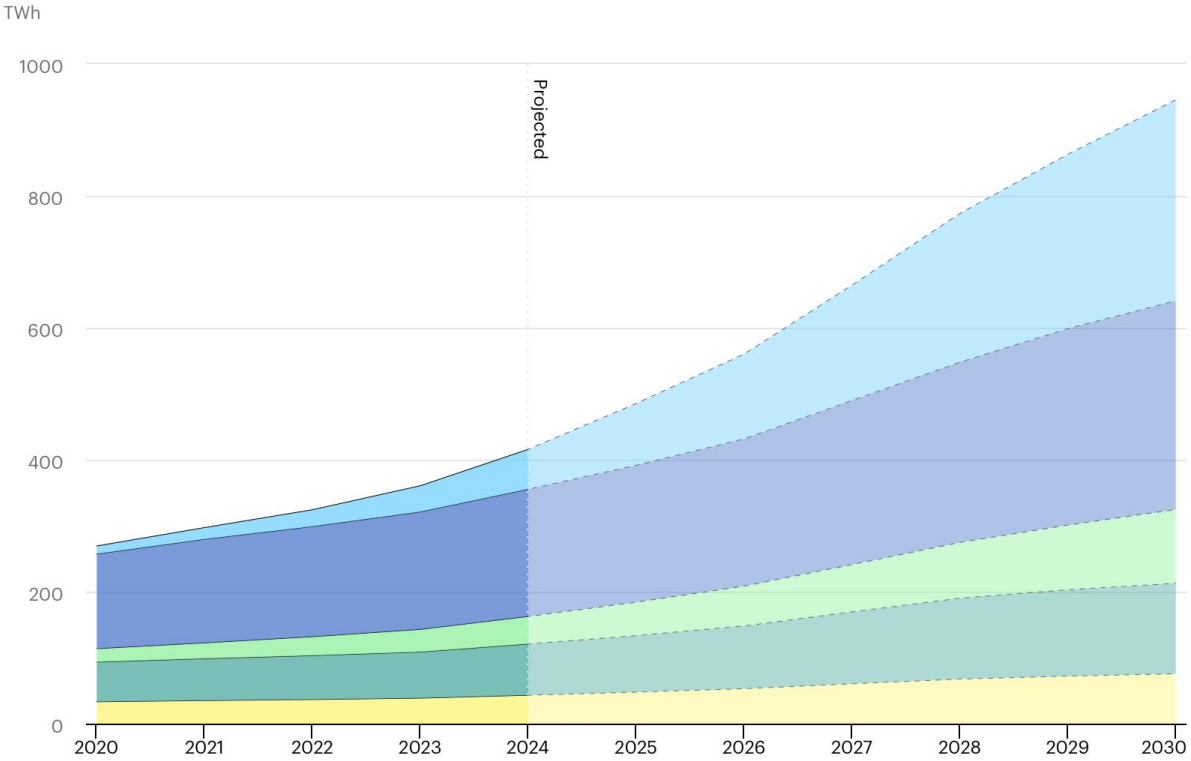


Digital consumption loosely based on: https://datareportal.com/global-digital-overview

CNE

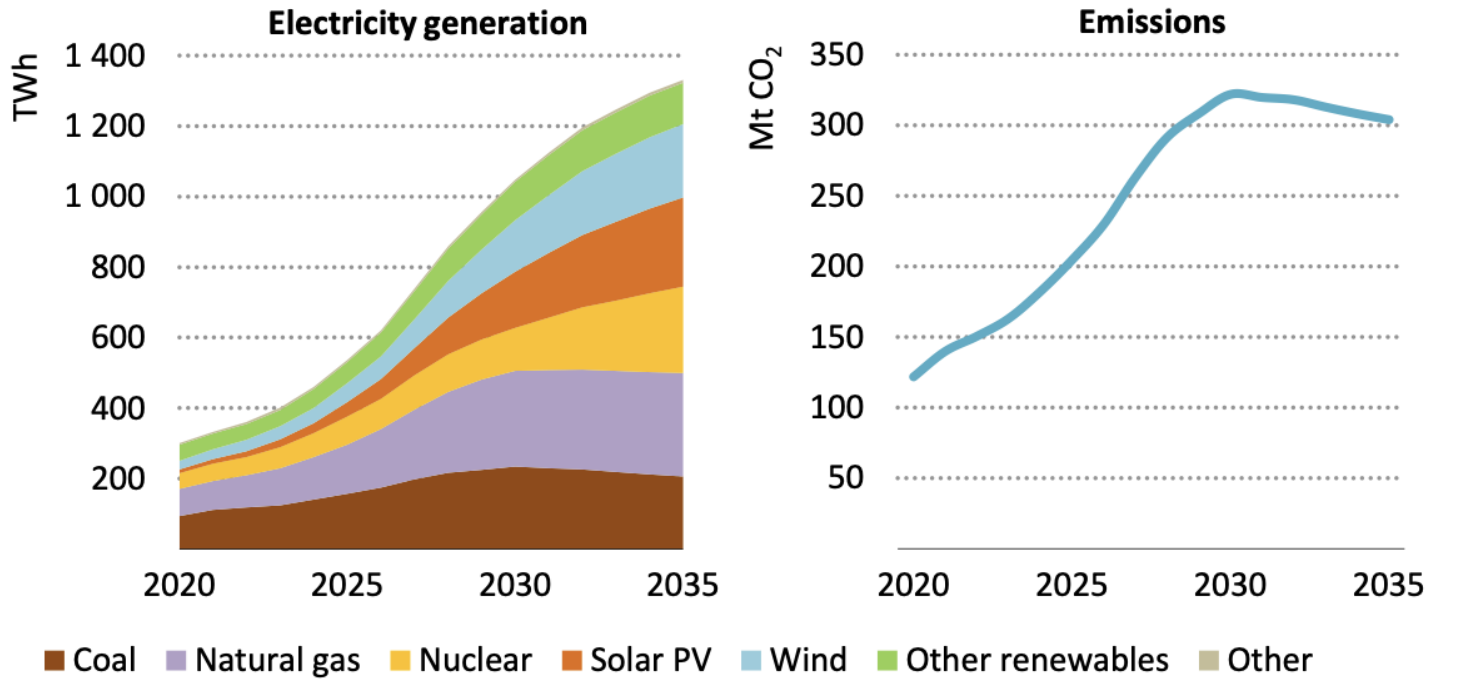
AI Data Centers Are Coming for Your Land, Water and Power

Think of them as AI factories, churning out your responses from ChatGPT, Gemini, Claude and all the other generative AI tools. The costs are staggering.


TWh
1000
800
600
400
200
0
Projected
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
IEA. Licence: CC BY 4.0
Accelerated servers
Conventional servers
Other IT equipment
Cooling
Other infrastructure

IEA (2025)

Figure 2.20 ▻ Global electricity generation for data centres and the associated CO_2 emissions in the Base Case, 2020-2035



Between now and 2030, renewables meet nearly half of the increase in global data centre electricity demand, followed closely by natural gas and coal-fired electricity generation

Plans submitted for 1GW data center in North Lincolnshire, UK

Will span 1.5 million sqm if built

May 30, 2025 By: Matthew Gooding Have your say

Detailed plans for a 1GW data center campus in North Lincolnshire, UK, have been revealed.

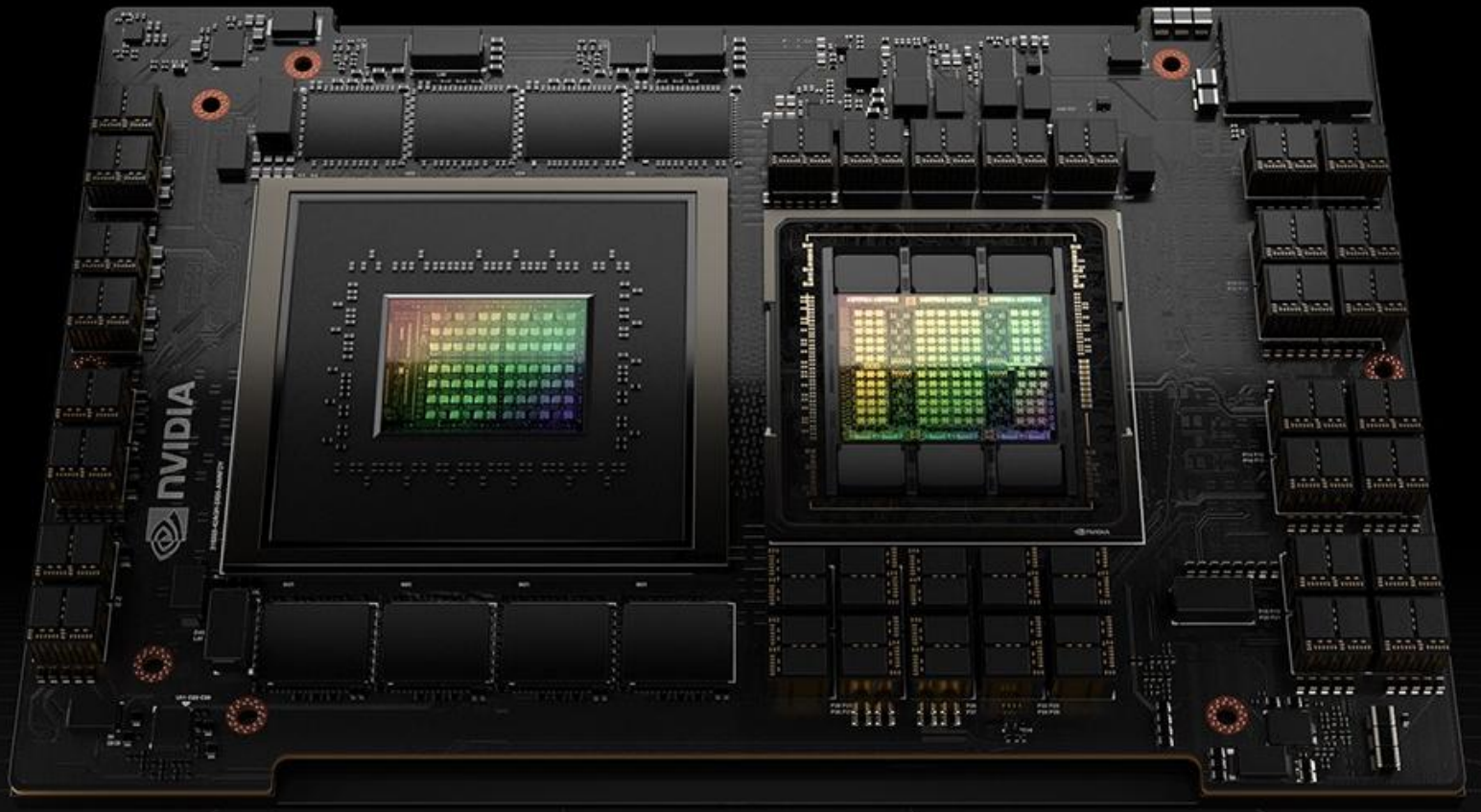
The Elsham Tech Park would be built on land south and east of Elsham Wolds Industrial Estate, adjacent to the A15 road outside Scunthorpe. Elsham Wolds is a former RAF base that has been converted into an industrial zone.

DCD reported in April that a pre-planning application for the site had been submitted, and a full application has now been lodged with North Lincolnshire Council by a newly formed company, Elsham Tech Park Ltd.

The site plan for Elsham Tech Park
– North Lincolnshire District Council planning portal

If approved, the campus would span 1.5 million sqm (16.1 million sq ft), making it one of the largest in the UK.

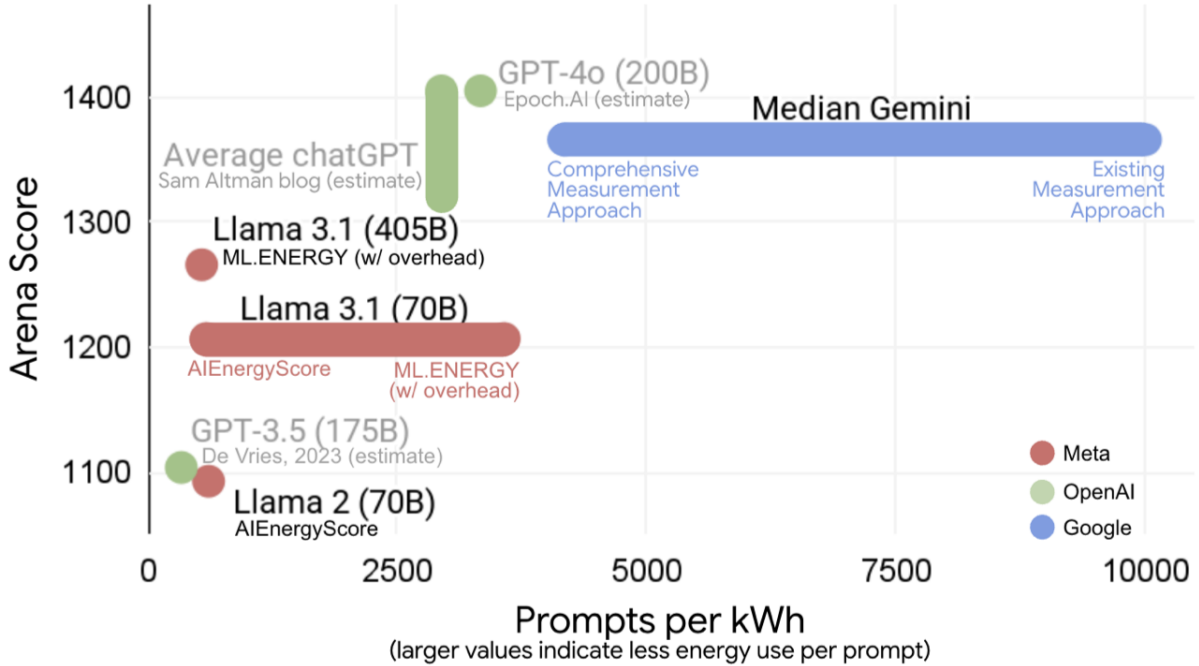
It will feature 15 data halls with a combined power draw of 1GW, the application said. An on-site energy center will generate 49.9MW to fulfill part of the development's power requirements, though the application does not specify what


NVIDIA

How many cups of water could you boil with a ChatGPT query?

- 0.01
- 0.1
- 1
- 10
- 100

Ballpark inference figure: 0.5Wh



bristol.a

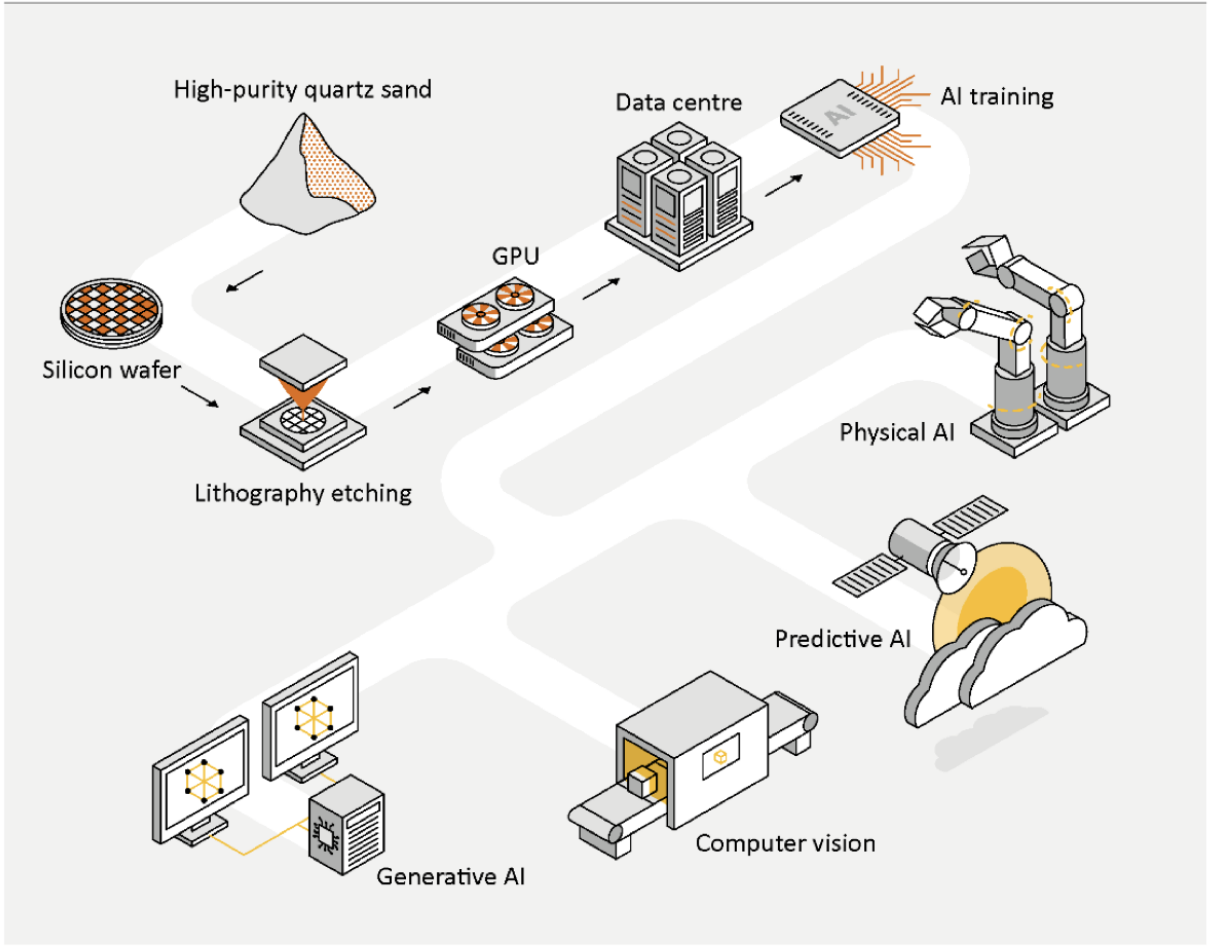
Scoping Life Cycle Assessment of AI

- What are the environmental impacts of using a chatbot?

Functional Unit: One ‘typical’ chatbot interaction.

System Boundary: We only consider the impact of the Data Centre and its activity

’Scoping’ assessment: ballpark figures to identify ‘hotspots’



IEA Energy and AI

AI is supported by a highly complex global supply chain

Environmental impacts of Data Centre activity

- Embodied impacts – Supply chain to construct the data centre.
- Operational impacts – (Impacts of Electricity and Water Usage)
  - Creation of the digital service: Training.
  - Use of the digital service: Inference

Impacts of one ‘typical’ chatbot interaction =

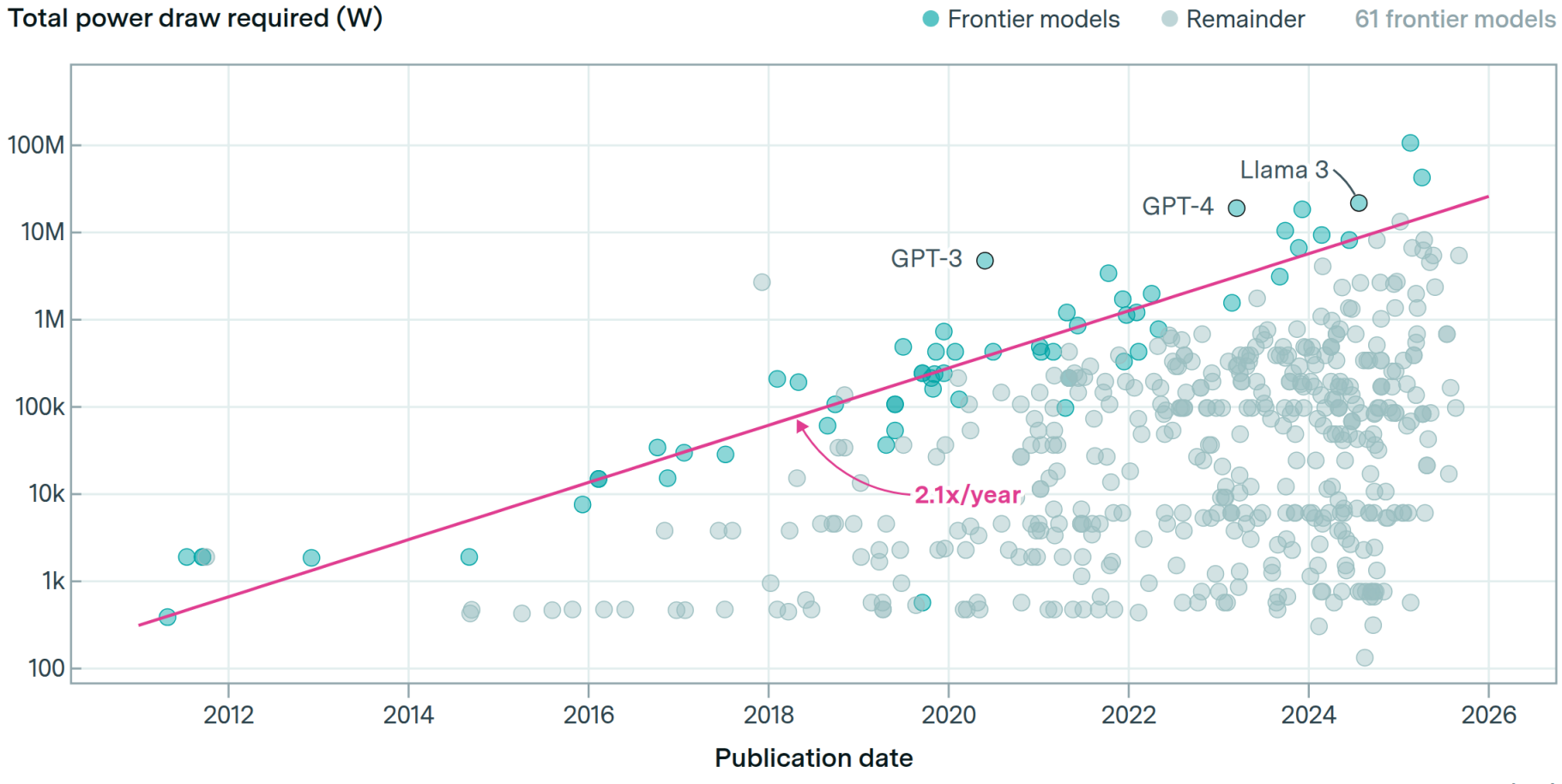
Impact of inference used + impact of proportional share of training + impact of proportional share of data centre embodied

bristol.ac.uk

TRAINING

Power required for training frontier models

EPOCH AI



EPOCH AI

- Estimate 246 Million NVIDIA H100-hours to train Grok 4
- Estimate of 310GWh electricity, including interconnect and cooling
- 154,000t CO_2e from its use of gas turbines
- This is about 3.5 years of emissions from a passenger jet
- About 2 days of LHR flights

Allocation of Training Impacts

A ‘Share’ of training impacts should be allocated to each use of the model.

e.g. Estimate total number of queries to the model, and allocate proportionally to each query

Find rough model ‘lifetime’, and queries per day to estimate this.

It will be approximate, ballpark.

Allocation of Training Impacts

- Amortize training energy costs over all queries.
- Assume a model has a 1 year lifetime
- ChatGPT had 18Bn queries per week in July 2025
- Hence training cost per query roughly 0.3Wh for ChatGPT
- Grok estimated to have market share 0.1 that of ChatGPT (Meeker p262)
- Hence training cost per query around 3Wh

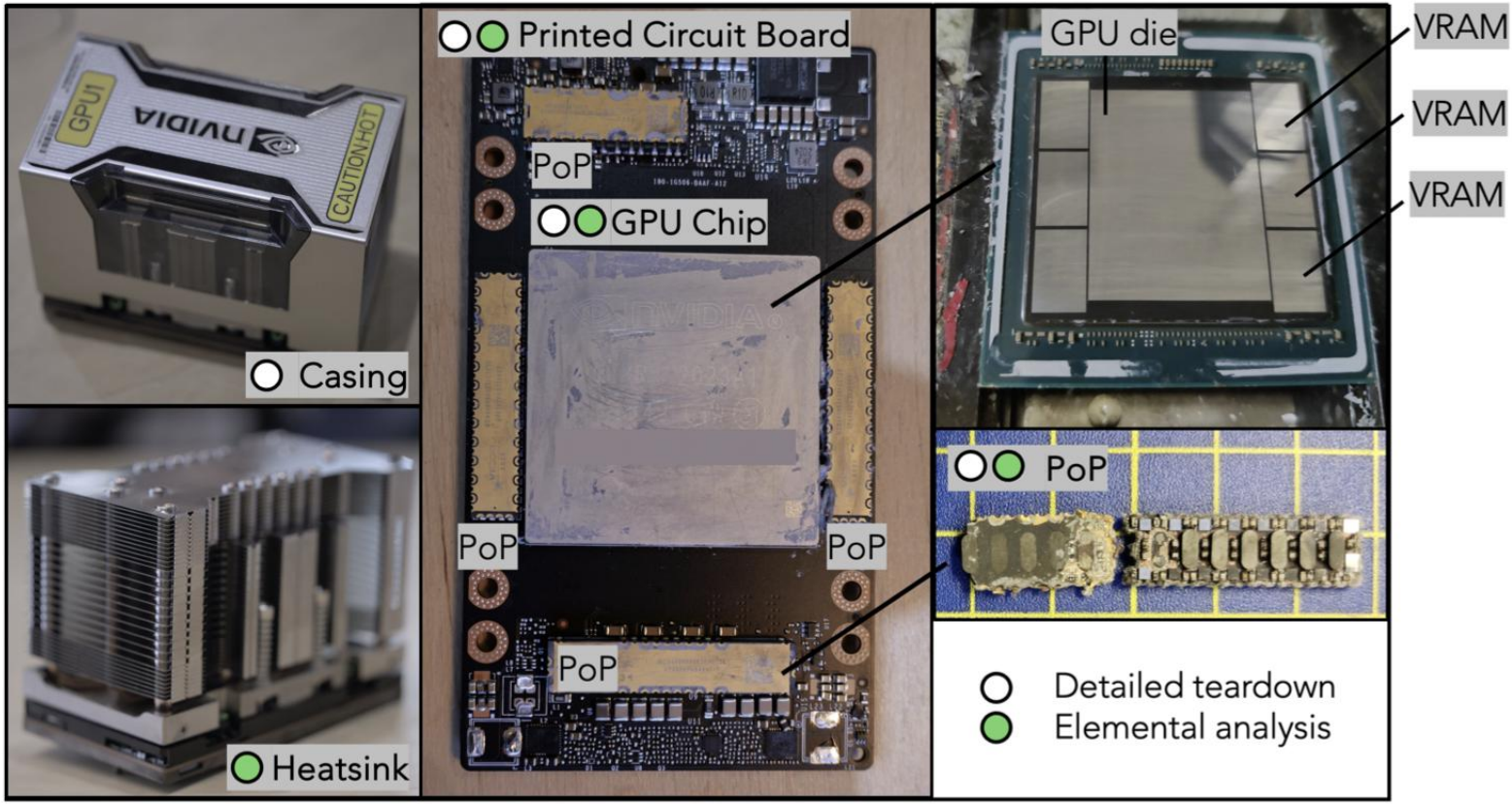
Embodied Impacts

- Dominated by chip manufacturing.

- NVIDIA in house assessment HGX H100 (8) – 1312 Kg CO_2e (= 164 Kg CO_2e per H100 plus chassis share)
  - https://images.nvidia.com/aem-dam/Solutions/documents/HGX-H100-PCF-Summary.pdf

- Academic Tear-Down assessment of A100 GPU – 127.6 Kg CO_2e

- Falk S, Ekchajzer D, Pirson T, Lees-Perasso E, Wattiez A, Biber-Freudenberger L, et al. More than Carbon: Cradle-to-Grave environmental impacts of GenAI training on the Nvidia A100 GPU [Internet]. arXiv; 2025 [cited 2025 Sept 17]. Available from: http://arxiv.org/abs/2509.00093



bristol.ac.uk

Allocation of Embodied Impacts

- A ‘share’ of manufacturing impacts should be allocated to each query to the model.
- Different Options
  - Allocate according to ‘Total number of queries’
  - Allocate according to ‘inference time taken’
  - Allocate according to ‘share of total energy used’

Allocation of Embodied Impacts

Embodied impacts of a single H100 within an HGX: 164 Kg CO_2e

Assume lifetime of 3 years, = 26280 hours

Power consumption of an H100 SXM, assume (conservative) 50% cpu load: 350W

Hence embodied impacts can be allocated:

164000 / (26280 x 350) = 0.018 g CO_2e per Wh.

Additional data centre equipment would roughly double this.

Say 0.04 g CO_2e per Wh

Scoping estimate per ChatGPT query

- 0.5Wh (to 5Wh if complex reasoning) inference
- 0.3Wh (to 3Wh or more for less used models) training
- 0.04 g CO_2e per Wh for embodied impacts

UK mix: 0.14 – 1.0g CO_2e

US mix: 0.35 - 3.0g CO_2e

Widening the scope

- We have not considered user device and network usage
- We have not considered impact of research and software development, beyond the actual training.
- We have not considered the impact of web crawler bots used to harvest the data needed

- We have not considered the (indirect) impacts of the uses of AI
- We have not considered how it changes society and the global economy

This Data Centre alone could offer *every* person in the world a prompt every 4 hours *using current technology*

bristol.ac.uk

Plans submitted for 1GW data center in North Lincolnshire, UK

Will span 1.5 million sqm if built

May 30, 2025 By: Matthew Gooding Have your say

Detailed plans for a 1GW data center campus in North Lincolnshire, UK, have been revealed.

The Elsham Tech Park would be built on land south and east of Elsham Wolds Industrial Estate, adjacent to the A15 road outside Scunthorpe. Elsham Wolds is a former RAF base that has been converted into an industrial zone.

DCD reported in April that a pre-planning application for the site had been submitted, and a full application has now been lodged with North Lincolnshire Council by a newly formed company, Elsham Tech Park Ltd.

The site plan for Elsham Tech Park
– North Lincolnshire District Council planning portal

If approved, the campus would span 1.5 million sqm (16.1 million sq ft), making it one of the largest in the UK.

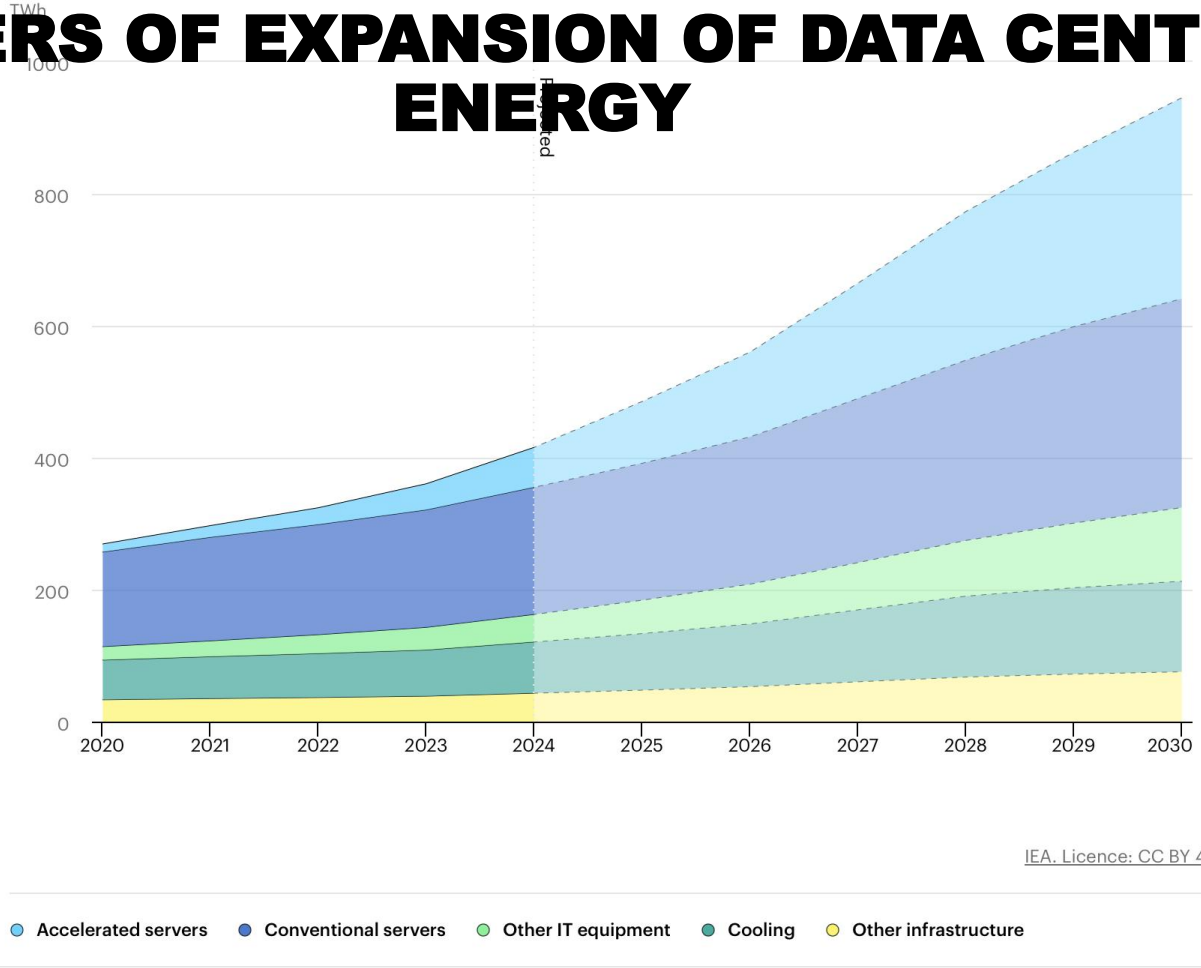
It will feature 15 data halls with a combined power draw of 1GW, the application said. An on-site energy center will generate 49.9MW to fulfill part of the development's power requirements, though the application does not specify what

Sources of efficiency improvements

Hardware improvement (Moore's law) and ASICs
>> NVIDIA H100 2-6x improvement on A100
>> Blackwell (optimistically) claiming 20-30x on H100

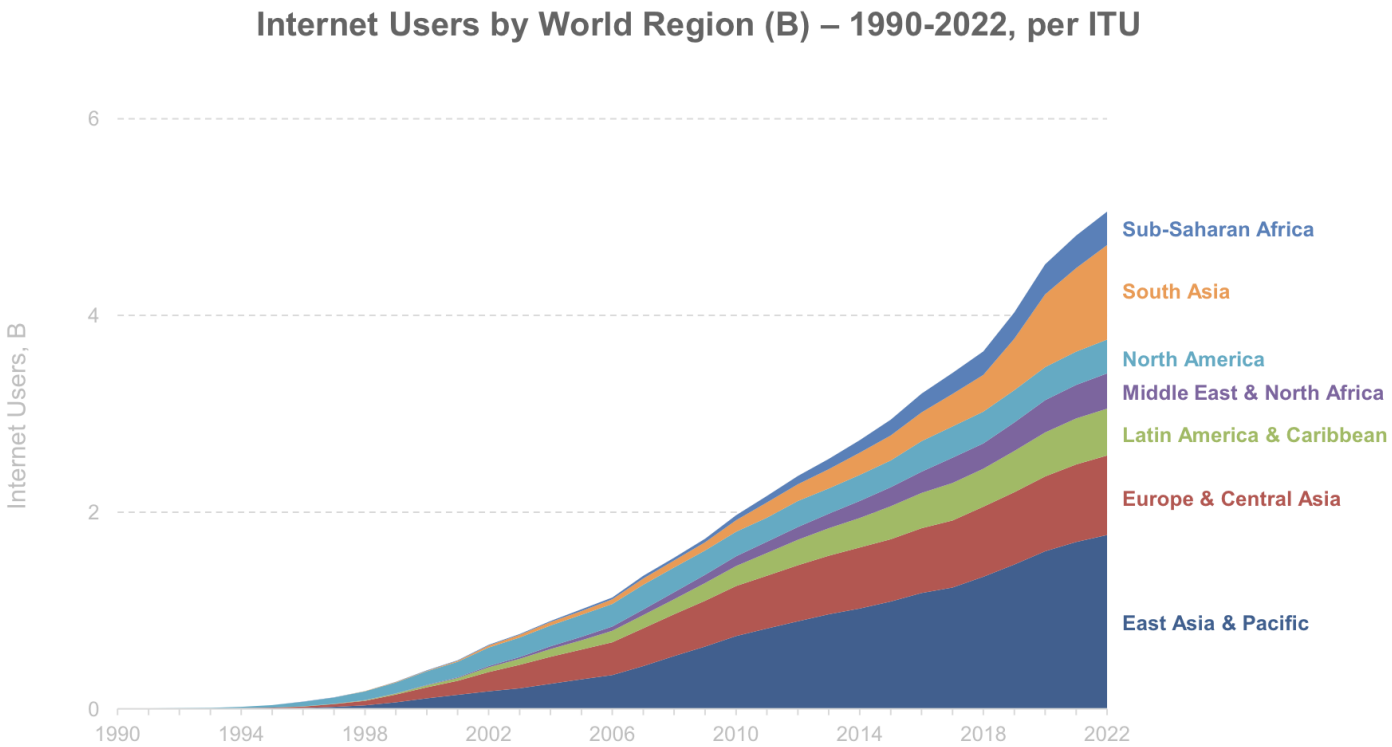
- Model Quantisation (Smaller numbers)
- Knowledge Distillation (Smaller Models)
- High level architectures (eg 'Mixture of Experts' architecture)
  >> Google report a 33x improvement in a year

DRIVERS OF EXPANSION OF DATA CENTRE ENERGY



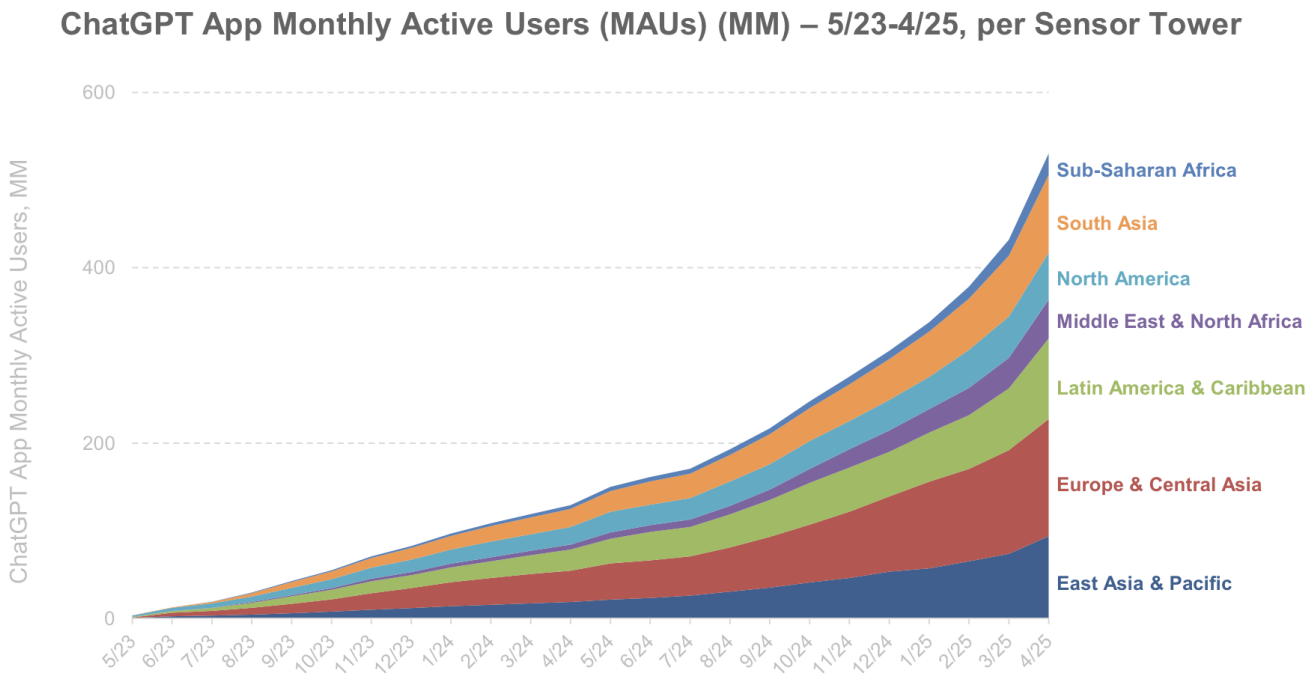
IEA (2025)

Global Internet Users =
Epic Growth Over Past Thirty-Three Years, per ITU



Note: 2021 data interpolated due to data gaps for select nations. Regions are per United Nations definitions. Data is occasionally unavailable for select nations in select years, which may lead to trendline choppiness or minor discrepancies vs. global user figures. Source: United Nations / International Telecommunications Union (3/25)

ChatGPT Mobile App @ 530MM MAUs in Twenty-Three Months = Global Growth We Have Not Seen Likes Of Before



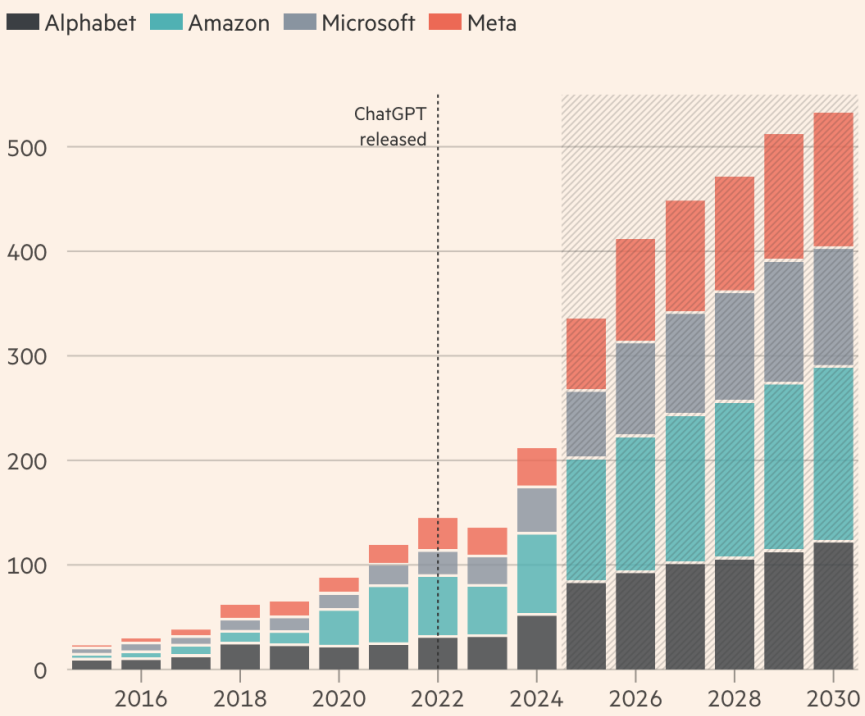
Note: Regions are per United Nations definitions. ChatGPT app not available in China, Russia and select other countries as of 5/25. Includes only Android, iPhone & iPad users. Figures may understate true ChatGPT user base (e.g., desktop or mobile webpage users). Data for standalone app only. Source: Sensor Tower (5/25)

Companies Are Pouring Billions Into A.I. It Has Yet to Pay Off.

Corporate spending on artificial intelligence is surging as executives bank on major efficiency gains. So far, they report little effect to the bottom line.

Big Tech's spending boom

Capital expenditure, $bn



FINANCIAL TIMES

Source: 10-K filings, S&P Global Market Intelligence • Forecasts = shaded area

Financial Times (Aug 21, 2025)

Considerations

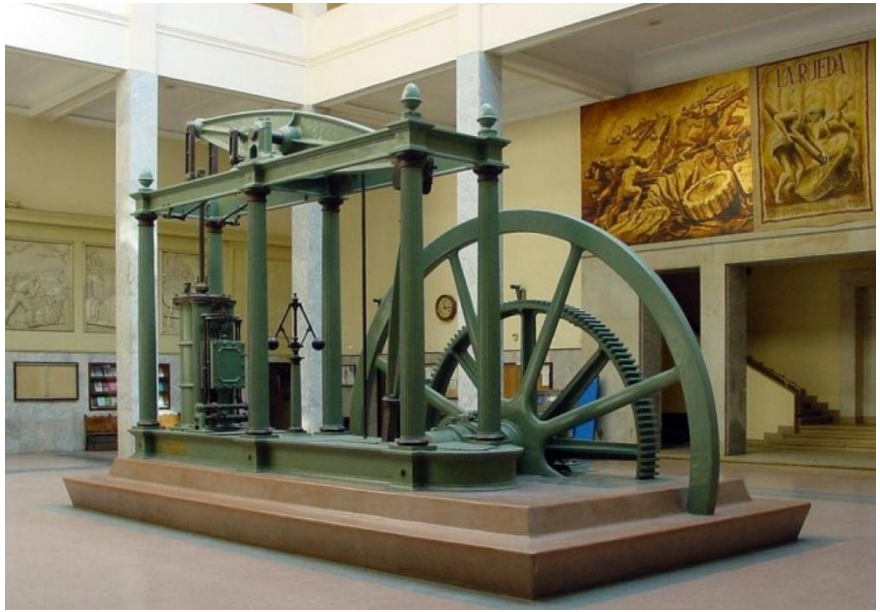
- The impact of rapid increased electrical power demand, locally and nationally. (And possibly water – though may be less an issue)
  - Data Centres as ‘Critical National Infrastructure’
- Rapid obsolescence? Circularity and repurposing?
- Digital Demand: What advice should be given to users? (Businesses, consumers)
  - Current well-intentioned misinformation
  - How to ensure it is evidence based and proportional

Wider Questions

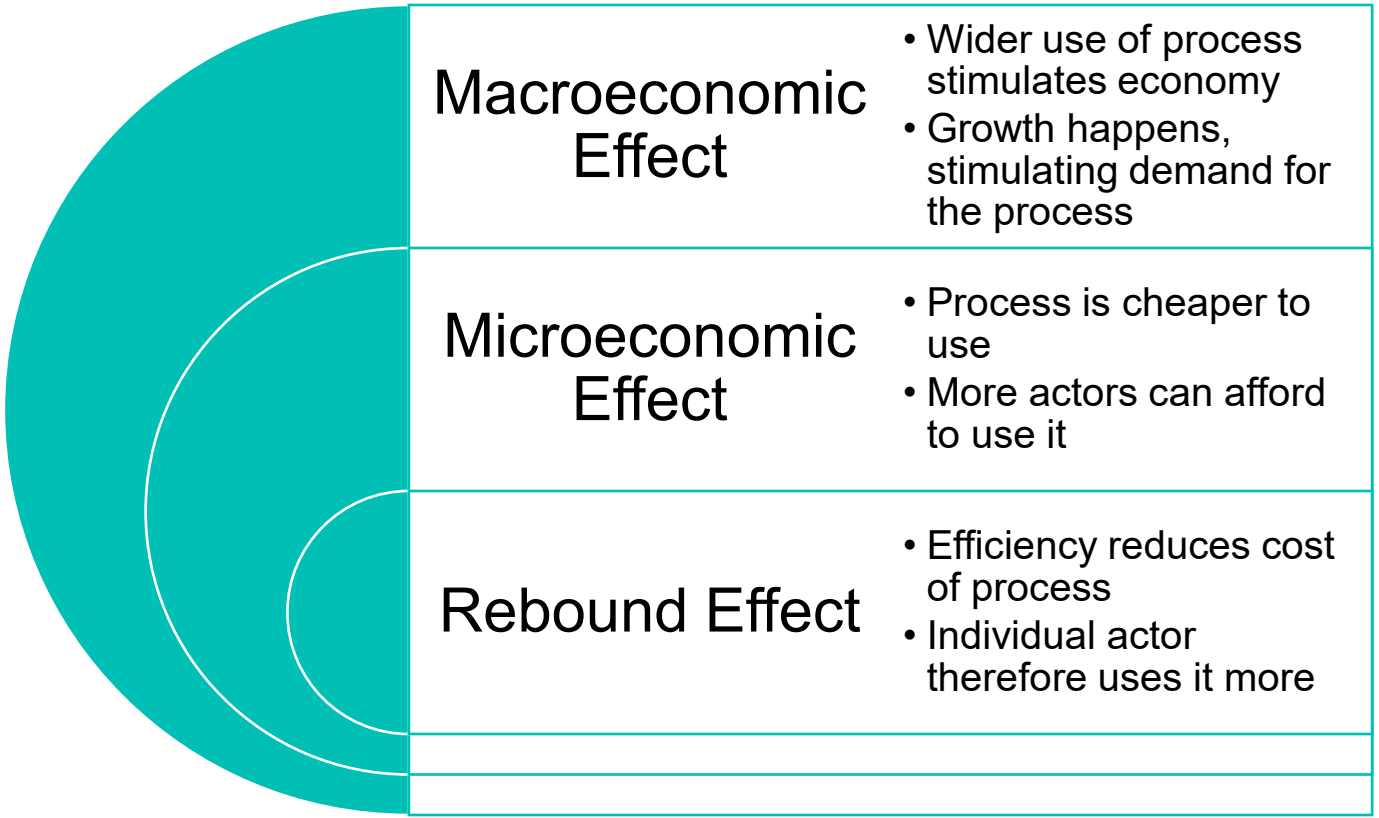
- What (social and political) factors are driving and shaping investment in the UK? And beyond? How might they change?
- What if it is a bubble? What if AI is merely ‘very useful’?
  - Model size stabilizes. Less training, more inference.
  - A phone is enough? Redundant infrastructure?
- What if it is far more?
  - AI in ‘Everything, Everywhere, All at Once’
  - Strategic Infrastructure? Rapid upgrade?

Supplementary slides if needed

Jevon's Paradox and the Rebound Effect



An increase in *efficiency* of a process using some input (fuel/energy) can paradoxically lead to an increase, rather than a reduction, in consumption of that input.


Macroeconomic Effect
• Wider use of process stimulates economy
• Growth happens, stimulating demand for the process
Microeconomic Effect
• Process is cheaper to use
• More actors can afford to use it
Rebound Effect
• Efficiency reduces cost of process
• Individual actor therefore uses it more

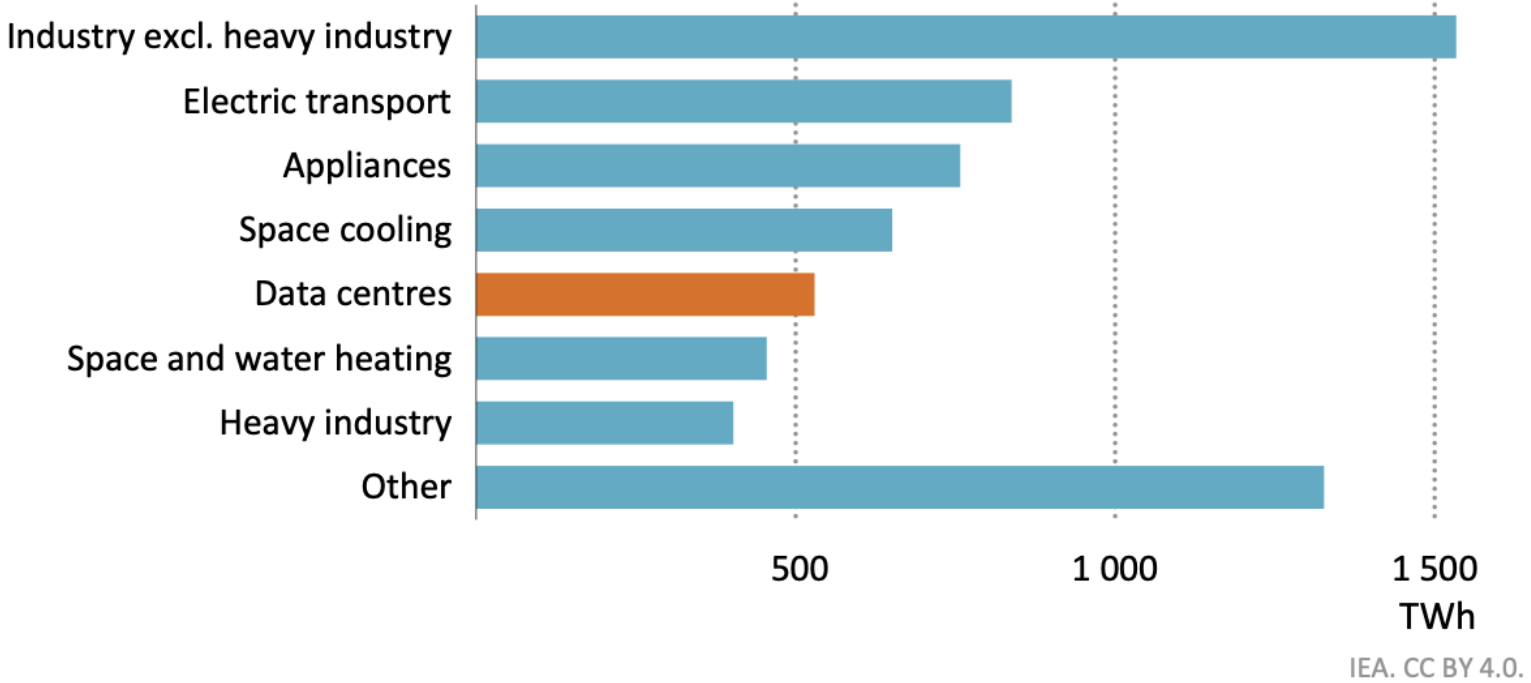
Water Usage: At the DC

- Direct, to cool the data centre: reduces energy use
- ‘Water Usage Effectiveness’ of a DC, measured in litres/KWh
- Old systems use about 3 l/KWh
- Historical USA average estimate: 1.8 l/KWh
- Recent USA average estimate: 0.4 l/KWh
- Microsoft reported average: 0.3 l/KWh
- DCs with Liquid Cooled AI chips - such as Isambard AI - often use an air cooler with adiabatic assist:
  - Varies with temperature, averages at 0.1-0.3 litres/KWh

Water Usage (l/KWh): At the power stations

- Indirect, in the process of generating the electricity used.
- Varies with the mix of generation technologies
- Coal (1.5), Nuclear (1.9) high.
- Hydroelectric dams highest and very variable – evaporation from lakes
- Gas lower (0.6)
- Renewables lowest (Near 0)

Figure 2.13 ▷ **Increase in electricity demand by sector in the Base Case, 2024-2030**



Data centres contribute more to global electricity demand growth than heavy industry or space and water heating

br