The Cloud and the Climate

Navigating AI-Powered Futures

Digital Humanities Climate Coalition

September 2024

Authors: Jo Lindsay Walton, Nathalie Huegler, Faosiyat Tiamiyu-Tijani, Kinda Al Sayed, Olivia Byrne, Josephine Lethbridge, Polina Levontin, Kerry O'Donnell, Ben Price, Nic Seymour-Smith, Benjamin Sovacool

Additional review and research: Rory Brown, Ishmael Burdeau, Can Hankendi, Alex Cline, Ayse Coskun, Hamish Hutchings, Janne Kalliola, Michelle Lin, Rudolf van der Berg, Mike Wooldridge

Digital Humanities Climate Coalition | Sussex Digital Humanities Lab | Climate Acuity

Suggested citation: Digital Humanities Climate Coalition (2024), *The Cloud and the Climate: Navigating AI-Powered Futures*. DOI: 10.5281/zenodo.13850067

Acknowledgements

The Cloud and the Climate is a publication of the Digital Humanities Climate Coalition, the Sussex Digital Humanities Lab, and Climate Acuity. We gratefully acknowledge the support of Designing the Future of Cloud Data Emissions (Innovate UK), Automating Climate Mitigation Advisory Services (AHRC IAA), and Designing Sustainable Digital Futures (Design Museum Future Observatory). All views are the authors' own. Thanks to: Anne Alexander, James Baker, Caroline Bassett, Alan Blackwell, Mark Butcher, Jennifer Cheung, Alice Eldridge, Benja Faecks, Didi Hoffmann, David Kohnstamm, Dan McQuillan, Adam Newman, Christopher Ohge, Florence Okoye, Lisa Otty, Charlotte Rae, Ben Roberts, Charlotte Pascoe, Hannah Smith, Jaskaran Singh, Arne Tarara, Sharon Webb, Oliver Winks, Sussex Energy Group, Better Images of AI, Climate Action Unit, Climate Visuals, Green Software Foundation. Special thanks to Greenpixie.

Cover image: 'Tree' by David Man and Tristan Ferne. For full image credits, see end of report.

Feedback, queries, and comments: j.c.walton@sussex.ac.uk.

CONTENTS

Data labelling

Executive Summary

This report gives a snapshot of current critical issues around AI, the cloud, and climate change. Awareness of the environmental impacts of AI and the cloud has been growing rapidly. While there are potential environmental benefits to unlock if we get our policy and practice right, indications from the latest AI revolution point toward a backsliding in climate and environmental commitments.

Topics like AI and the climate are incredibly exciting and important for many of us. Yet the underlying evidence can be very complex, messy, and — unless you happen to be professionally or personally very into it — boring. We have done our best to make this report accessible to both technical and non-technical audiences, while preserving the nuance.

There is some focus on the UK context, but this report should be of interest all around the world, especially since the ‘cloud giants’ Amazon, Google, and Microsoft have such a global reach.

Somehow this report ended up getting quite long. If you don’t quite make it to the end, make sure not to miss the ‘Actions and Resources’ section. The Digital Humanities Climate Coalition welcomes queries, comments, and opportunities for dissemination and collaboration.

Let's begin with some key messages. These comprise the report in a nutshell. If they feel dense and jargon heavy, don't worry: many terms will be explained later on. We also have two glossaries here to help: one focusing on the technical side of cloud computing, the other on the sustainability governance side. So what are our key messages?

1. **We are in a climate crisis.** Progress has been made, but the world as a whole is not yet on track to meet its decarbonisation targets. The negative impacts are already happening, including loss of human lives, economic loss that hits human freedom and flourishing, and irreversible changes to nature. Once net zero is achieved and maintained, global warming is likely to stop. See Net Zero and Overshoot.

2. **The digital has a physical basis.** When we think of AI and 'the cloud' (where most AI is run), it’s easy to imagine something intangible. But in reality, it relies on tangible resources—data centres built from concrete, metals, and other materials, powered by electricity that is produced by burning fossil fuels or by renewable sources. See Artificial Intelligence and the Cloud.

3. **The climate footprint of the cloud is growing when it should be shrinking.** Innovation and the spread of best practices are improving efficiency. Plentiful options exist to make the cloud, and the data centres behind it, more sustainable and efficient. However, the overall growth of data centres and other network infrastructure is overwhelming the efficiency gains. Drivers of growth include new technology such as AI, as well as efforts to close the digital divide and a growing world population. See Tech’s pathway to net zero; Green Data Centres; The Current Carbon Footprint of AI: A Microsoft Case Study; Modelling the Future of AI and the Climate?

4. For some companies, cloud use can be both a significant cost and a significant environmental impact. **GreenOps** is emerging as a paradigm for joining up cloud teams with sustainability teams. To aid with this, there are **open source tools and third-party sustainability data providers** that can sit between cloud service providers and their clients. See GreenOps.

5. The cloud giants, Amazon, Google, and Microsoft, all have track records of celebrating efficiency improvements that are outweighed by growth. There is an urgent need for **policy to limit and to more fairly allocate the use of IT resources**, instead of relying on efficiency improvements and on voluntary sustainability initiatives. See Modelling the Future of AI and the Climate?

6. Google and Microsoft in particular have aimed to show leadership on sustainability, and have been **influential in shaping the overall paradigm and pace of energy transition**. Understood more holistically, however, their activities have not been consistent with a rapid, orderly, and just climate transition. Communications around sustainability progress has misled and in some cases constituted greenwashing. They have not managed to stay within their pledged carbon budgets. Regrettably some elements within each company now appear to be preparing the ground for further future failures. A course correction is possible and necessary. See Google; Microsoft.

7. **Policy interventions are needed, but may not happen.** The cloud giants are all in the top ten biggest companies in the world by market capitalisation. These powerful global entities shape policymakers' beliefs, attitudes, and regulatory approaches. While we advocate for bold, responsible policy to bring the tech sector into line with climate goals, we acknowledge the possibility that governments may lack the power to achieve this. See The Cloud Giants and the Greenhouse Gas Protocol.

8. **None of the cloud giants is a monolith**, and there is great potential for the more progressive elements within Amazon, Microsoft, and Google, along with industry, policy, and other societal stakeholders, to work together to find the approaches, tools, evidence, and resources to drive change. See Actions and Resources.

9. Although the **carbon-aware computing** approach has helped to raise awareness, and generated efficiency gains, so far it has not really considered the systemic picture of achieving global net zero. Useful dialogue is emerging under the term **"grid-aware computing,"** but there still remains a gap when it comes to joining up individual organisational strategies with the bigger picture. See Carbon-Aware Computing and Grid-Aware Computing.

Scientist studying coral reefs in Virgin Islands National Park

10. The cloud giants are working to shape a revision of the **Greenhouse Gas Protocol,** which determines best practice in carbon emissions accounting. Amazon's proposed shift to more **“project-based” / “consequential” accounting** may pose significant greenwashing risks. **24/7 hourly matching** has been championed by Google and Microsoft to address greenwashing issues in energy procurement. This approach has some promise, but there has been as yet not enough critical scrutiny and debate. This approach does not address the risk of well-resourced purchasers hoarding clean energy resources that might otherwise be allocated to more socially beneficial uses. Voluntary initiatives can also undermine and/or delay much-needed decisive regulation. See 24/7 Hourly Matching; The Cloud Giants and the Greenhouse Gas Protocol; Carbon Offsets; Relevant UK policy on sustainable AI.

11. **Greenwashing and lack of transparency remain serious issues.** All cloud giants have shortcomings as regards their own net zero / net negative transparency. Claims to be carbon neutral and 100% renewables-powered may be technically defensible, but are likely to be misunderstood by stakeholders who expect good faith and clear communication. Transparency might be improved by mandating VCMI and Core Carbon Principles, within-region hourly matching of renewable energy purchased and used, and Beyond Value Chain Mitigation, by prohibiting claims which obfuscate the use of carbon credits at the level both of organisation and product / service, and by creating more channels to report and correct greenwashing short of legal action. However, by itself the improved availability, accuracy and transparency of information will not lead to better outcomes. Stronger policy remedies are also required. See Sustainability Transparency of the Cloud Giants; The Cloud Giants and the Greenhouse Gas Protocol; Carbon Offsets; Amazon; Google; Microsoft.

12. **The ICT sector is not a leading contributor to global warming, but it still must decarbonise rapidly.** Much larger contributions are made by activities like agriculture, constructing buildings and heating and lighting them. The scale of ICT is still significant enough that rapid decarbonisation is urgently needed. Some sectoral emissions are more expensive to abate than others. ICT is acting as an enabler both of decarbonisation and of continued reliance on fossil fuels. See Tech’s pathway to net zero; Modelling the Future of AI and the Climate?;

13. **More AI means more cloud.** Very recently, AI has grown very prominent in thought leadership, and had a considerable influence on investment and corporate strategy, across a range of sectors. An AI intensive future appears to be assumed by many business leaders and policymakers. As AI is typically trained and deployed using the cloud, an AI intensive future implies the future expansion of data centres and networks, and their associated climatic and environmental impacts and opportunity costs. Microsoft's research and thought leadership on AI's carbon impact requires urgent methodological revision, and currently appears more oriented to controlling narratives than open collaboration to solve real challenges. See The Current Carbon Footprint of AI: A Microsoft Case Study.

14. The interdisciplinary field of **Critical Data Center Studies** has sprung up in recent years, drawing on the environmental humanities, and the anthropology and sociology of data centres, although there are as yet few signs of this important research impacting policy and practice. Not all scholars in this field appear to have appetite for attempting to impact policy and practice. 24/7 hourly matching deserves more attention from Critical Data Centre Studies. See Green Energy Procurement; The Cloud in Context.

15. **AI's sustainability benefits have been misleadingly presented.** It is important to distinguish digital technology *for* sustainability from the sustainability *of* digital technology. Digital technologies including AI plausibly have a part to play in a rapid and just transition to net zero, and in adapting to a warmer world. To clarify their role, some broad distinctions can make a big difference (see next point), even in the absence of detailed frameworks for weighing up the environmental costs and benefits. However, some AI is also devoted to purposes that are inconsistent with a rapid and just transition to net zero. See AI for Sustainability?; A Literature Review Reviewed; Modelling the Future of AI and the Climate?

16. When discussing the use of AI for sustainability, **instead of talking about AI in general, we are better off clearly distinguishing different types of AI**. Many use cases of AI for sustainability depend on relatively lightweight discriminative models, not power-hungry GenAI models. **Likewise we should clearly distinguish between climate adaptation and climate mitigation**. Many AI use cases will help to cope with a warmer world, but won't accelerate decarbonisation. The messaging that it is 'too early to tell' can obscure these distinctions and shield AI from responsible scrutiny. See The Current Carbon Footprint of AI: A Microsoft Case Study; Modelling the Future of AI and the Climate?

17. In the literature on using AI for sustainability, there is some evidence of academic research that is **well below publishable standard,** representing **likely academic misconduct,** going largely unnoticed and unremedied. The literature on AI for sustainability needs to be thoroughly reviewed to determine the nature and scale of this problem. Peer review processes need to be examined, and aggregator sites (such as Google Scholar) need better mechanisms for verifying journal quality and reporting suspicious content. See AI for Sustainability?, A Literature Review Reviewed, and More AI for Sustainability Spam.

18. **We can develop better tools and frameworks for weighing together the environmental costs and benefits of AI systems.** Claims for the sustainability benefits of AI often go untested, and seldom are weighed against sustainability costs. Sustainability benefits and costs, and associated uncertainties of each, ought to be the default way of thinking about AI for sustainability. There may be beneficial 'spillover' effects, where AI research that is not immediately justifiable by its environmental cost/benefits helps drive sustainability in a different domain—but there can be harmful spillover effects too. See AI for Sustainability?, A Literature Review Reviewed, and More AI for Sustainability Spam.

19. The cloud giants can and should improve the **transparency of their own AI foundation models** (and/or of companies such as OpenAI, as major investors). It is not clear that they will do so without strong policy incentives. See Amazon; Google; Microsoft.

20. Greater emphasis on the distribution of IT resources can help to address the **digital divide,** with due consideration for the risks of **adverse digital inclusion.** Currently about two-thirds of the world's population online, and one-third who do not use the internet. **Internet use is growing.** Some estimates suggest that there will be around a billion more users added in the next five years. See The Cloud vs. On-Prem vs. Hybrid; The Cloud in Context.

21. Critiques of data centre environmental sustainability often fail to contextualise data centres within the wider environmental footprint of ICT. The embodied carbon and energy use of **networks** and **user devices** comprise a large part of this footprint. See Networks and users devices.

22. **Responsible AI tools and frameworks are maturing, but there is work to be done.** AI Ethics, Responsible AI, and AI Safety are currently the dominant paradigms for regulation and other applied critical thinking around AI. As yet the critical perspectives of the arts and humanities and social sciences (e.g. critical data center studies, critical data studies, critical internet studies, critical AI studies, Science and Technology Studies, the Digital Humanities) are at best marginal to these paradigms. Responsible AI tools have yet to reflect in any actionable, granular way the environmental impacts of AI. On the positive side, such tools appear to be rapidly evolving; however, there is no guarantee of uptake. See Relevant UK policy on sustainable AI; The Cloud in Context.

23. **The public profile** of the cloud's environmental impacts has grown rapidly in recent years, as demonstrated for example by a 2023 BBC Panorama episode on the topic, and extensive press coverage in 2024 of the environmental impacts of AI. However, the cloud is not yet a significant part of the public climate change imaginary, in the same way as for example oil rigs, gas pipelines, SUVs, short haul flights, recycling, etc.

24. This report has a **UK focus**, although it is relevant globally. We believe that the UK is well-placed to **show leadership in cloud sustainability and AI sustainability going forward**, based on factors including net zero ambitions, existing tech and sustainability capacity and expertise, and strategic R&D and infrastructural priorities of both the previous and new governments. See Relevant UK policy on sustainable AI.

25. Everywhere we look we find passionate individuals working for change in their local context, and actively seeking collaborators and fellow travellers. The communities working on digital sustainability are exceptional in their willingness to share and collaborate, often in a spirit of great generosity and open-mindedness. Despite the challenges, **we are optimistic.**

Information Technology

“The digital is material.”[1]

Information Technology (IT) is a broad term. It encompasses things like computers, phones, and other devices, software such as enterprise application software and middleware, networks made up of servers, cabling, datacentres, switches, routers, and hubs, and all the other equipment that stores, processes, and transmits data.

IT is ubiquitous for pretty much everybody in the developed world, and the backbone of many industries. In healthcare, for instance, IT is essential for storing and accessing patient records, running diagnostics, and facilitating remote care. In retail, IT includes point-of-sales systems, online shopping platforms (everything from multinational corporations to local independent shops now has an online retail presence), your inventory management, and so on.

We begin in this way to emphasise that IT is *physical*. It has some very direct impacts on climate and environment. When I access a website for the first time, it takes energy to transmit that data from a server somewhere onto my device. My device is also using energy, and so is the server as it waits patiently for my request. Where does this energy come from? Maybe it comes from burning coal or gas, releasing carbon into the atmosphere and heating the climate. Maybe it comes from biomass or nuclear. Hopefully it comes from clean renewable energy, such as solar, wind, or hydropower — even then, there will be some emissions associated with the construction of the solar panels, the wind farms, the turbine lagoons. Building that device, from extracting and processing raw materials, to assembling parts and packaging and shipping, also uses energy and resources. Then there's how you dispose of it at the end of its working life.

On the other hand, IT is also more than physical. It is more than just the *stuff* itself — batteries and wires and circuits and satellites and undersea cables. It is also people and the things that we do with these technologies. IT is also a set of practices, institutions, laws, regulations, norms, cultures, and so on. We might say that Information Technology also consists of a **sociotechnical imaginary** (to use Sheila Jasanoff’s term).

How can we get a flavour of this sociotechnical imaginary? Just for starters, we can pay attention to the aesthetic and rhetorical features of the way tech workers do what they do. These features may seem incidental, but they often teach us things about what the tech industry tends to pay attention to or to value.

We can also think about who works in IT. This doesn’t mean subscribing to stereotypes, of course, but simply acknowledging that the population of IT workers at any given moment represents some particular distribution of social and cultural characteristics. **IT is people:** SysAdmins manage and maintain an organisation's IT infrastructure; Developers create software applications; Database Admins manage databases; Help Desk Technicians support end-users and troubleshoot issues; Data Scientists are applied statisticians who work with big data to try to extract insights; Security Analysts protect data from cyber threats;

---

[1] *Digital Humanities and the Climate Crisis: A Manifesto*. <https://dhc-barnard.github.io/envdh/>

Chief Information Officers align IT systems with business goals and manage overall technology infrastructure; roles like Chief Marketing Officer, Social Media Manager in tech companies, etc. work to manage image and relationships with the wider world; and much more.[2]

What are the social and educational backgrounds and connections of the humans who fill these roles? What are their daily lives like? How do they perceive the relationship between their professional and personal lives? Do their experiences outside of work actually sometimes show up within their working lives (and vice-versa)? Do the kinds of experiences they have encourage them to see the world in a particular way? What kinds of cultures are common in the places they work? What are their identity characteristics, such as gender, ethnicity, nationality, class? (For example, "data centers still employ shockingly few women"[3]). How might identity and lived experience shape the way they perceive climate and the environment? Some research suggests that within wealthy countries, "women are more likely than men are to express concern about our changing climate" (Bush and Clayton 2022). How are identity characteristics distributed across different roles *within* IT? How much variation is there in the answers to these questions, and in what ways? How might the answers vary in different places, or have changed in recent years?

Exploring questions like these can help us get used to the idea that digital technologies are not only material, but also social and cultural. We can also add that IT tends to be a space filled with tantalising puzzles, elegant solutions, and an endless effervescence of innovations. There are many ways to improve the efficiency of IT, and this report touches on some of them. But it's worth stating up front, nice and clear, that all this promise may *itself* be perilous. **If we focus too much on using resources efficiently, we may lose sight of how resources are distributed. We may also lose sight of the potential to use fewer resources, or to reduce the rate at which resource use grows.** As Freitag et al. (2021) write:

> There is a pressing need to devise a strategy for constraining consumption of ICT so that efficiency improvements lead to actual emissions reductions and enable productivity to be maintained in a carbon-constrained world. It is likely that unabated growth in demand for ICT will more than offset the emissions saved through improved efficiency of these technologies. The only condition under which these rebound effects would not apply is if a constraint were applied, such as a constraint on consumption or an economic constraint through rising carbon costs (e.g., a carbon tax or a cap on emissions).

Finally, there is one important category of people we haven't yet mentioned: the users of IT. The expectations, habits, capabilities and needs of users are also part of the global ICT system, how it functions and how it relates to climate and environment.

---

[2] Those are just a few examples, of course. There are many more, and the people who fill these roles actually do far more than is suggested by these terse summaries!

[3] Jacqueline Davis, "Data Centers are Short-Staffed Boys' Clubs," *Uptime Institute*, November 1, 2023, <https://journal.uptimeinstitute.com/data-centers-are-short-staffed-boys-clubs/>

What is the internet made of?

A non-comprehensive list includes:

1. Routers: Devices connecting end users to their Internet Service Providers and managing traffic between the user and the broader internet.
2. Fibre-Optic Network Equipment:
   - Fibre-Optic Cables: Backbone of high-speed internet, transmitting data as light pulses. These include undersea cables.
   - Network Routers: Devices routing data packets between different networks.
   - Undersea Cables
     - Shelters or Housings: Physical structures housing and protecting networking equipment.
3. Servers Hosting Digital Services: Powerful computers storing, processing, and serving up data and applications.
4. Switches: Used within networks to connect devices like computers and servers within a LAN.
5. Data Centres: Facilities housing servers and related equipment, processing and storing internet data. See 'Types of Data Centres' below.
6. Content Delivery Networks (CDNs): Distributed networks of servers providing fast delivery of internet content.
7. Internet Exchange Points: Locations where different internet networks connect and exchange traffic.
8. Modems: Devices for data transmission over cable or phone lines, often used with routers in home networks.
9. Cellular Network infrastructure: Especially in areas where wired connections are less developed or feasible, cellular towers and related infrastructure are significant for internet connectivity.
10. Power Plants and Power Infrastructure: Without electricity, there would be no internet.
11. Our own bodies — our eyes, hands, fingers, fingerprints, faces, our brains and hearts and everything they're connected to — might also be seen as fundamental parts of the internet.
12. Ideas, concepts, practices, norms, discourses, habits, imaginaries, ideologies. If these were very different, the internet would be very different too. The social is material, but its materiality includes things you can't straightforwardly touch.

Net Zero and Overshoot

We are in a climate crisis. According to the latest IPCC reports, to have a good chance of achieving the 2015 Paris Agreement commitment of holding average global temperatures to well below 2.0 degrees (and preferably 1.5 degrees) above the baseline period, global net emissions would need to peak now, drop steeply to about half their current levels by 2030, and hit net zero by 2050 (IPCC 2022).

It now appears likely that we are in an 'overshoot' scenario where emissions will continue to rise, or not drop quickly enough, and average global temperatures will regularly exceed 1.5 degrees (WMO 2024; Betts et al. 2023; Lamboll et al. 2023; Cuff 2024). Such warming is not irreversible, although it is difficult to reverse, and some of its effects may be irreversible.

In theory, some of this warming could be reversible if we do better than net zero, and get to net negative (Dunne 2024). In a net negative situation, more carbon is being removed than emitted, and we could start to cool average temperatures back down. We are very far off from this scenario. There is a risk that we put too much store in speculative future developments, believing we can rapidly get to carbon net negative later in the century, and so fail to take urgent action now.

Achieving and maintaining net zero (or better) will likely be enough to stop global warming (Matthews and Caldeira 2008; Hausfather 2021 IPCC 2022). There is some uncertainty around this point, however (Chaisson 2011; Corner et al. 2023).

Terms like net zero and carbon neutral are also applied at smaller scales—a country, a city, a company, a product, a process. These usually need to be investigated carefully, as there are lots of definitional subtleties, and plenty of opportunities for misunderstanding or for greenwashing.

Climate change impacts are not only in the future, but also the past and present. For example, the UN estimates more than 20m people have been forcibly displaced by climate events each year since 2008, and this number is expected to increase. In this report, we assume readers have a foundational knowledge of climate change and the need for urgent decarbonisation and climate adaptation.

Planetary Boundaries

This report focuses mostly on climate change. However there is also more to the environmental crisis than climate change alone. For example, the influential 'planetary boundaries' model identifies nine interconnected areas in which human activities may be impacting the planet's capacity to support life, including climate change, freshwater change, biosphere integrity (including biodiversity), release of novel entities into the environment, and others (Stockholm Resilience Centre 2023). IT has impacts across all these dimensions.

Artificial Intelligence

Artificial Intelligence has been around for decades (or depending on how you define it, for centuries). However, in recent years there has been something of an AI revolution, driven by the availability of large amounts of data and processing power. What is popularly described as 'AI' is a particular type of AI called Machine Learning (and usually specifically Deep Learning). We look at AI in more detail toward the end of this report.

Tech commentator Paris Marx writes, "generative AI is an environmental disaster that's accelerating natural destruction and the climate crisis at the very moment alarms are sounding about the precious little time that remains to turn things around" (Marx 2024).[4]

AI has now hit mainstream awareness. A significant milestone was ChatGPT, launched by OpenAI with support from Microsoft in 2022. ChatGPT is based on a Large Language Model (LLM); a key turning point in the underpinning research was the 2017 paper "Attention Is All You Need" (Vaswani et al. 2017). Other generative AI (GenAI) is being used to create images, video, music and other content. Such AI is resource intensive, as Campbell et al. (2024) summarise:

> This new AI is exciting. However, the rising demand for AI is substantially increasing the demand for compute. This is then driving the need for more AI chips and the energy needed to run these. If data is the new oil, AI is the new car. Everyone wants one and soon, everyone will have more than one. But all this AI is costing us environmentally. We won't know how much it costs for certain until hyperscalers publish a detailed analysis of AI services. But if we take Microsoft's sustainability reporting from FY21 to FY22, we can see a 43% increase in electricity use across the entire organisation.[5]

In its most recent sustainability report, Google writes, "our total GHG emissions were 14.3 million tCO2e, representing a 13% year-over-year increase and a 48% increase compared to our 2019 target base year—primarily due to increases in data centre energy consumption and supply chain emissions" (Google 2024).[6]

---

[4] *The Disconnect* blog, July 5, 2024. <https://disconnect.blog/generative-ai-is-a-climate-disaster/>

[5] Peter Campbell, Nikos Karaoulanis, Marc Nevin, Caoimhin Graham, Joe; McGrath, *Digital Sustainability: The Need for Greener Software* (Kainos Software, 2024), p. 67.

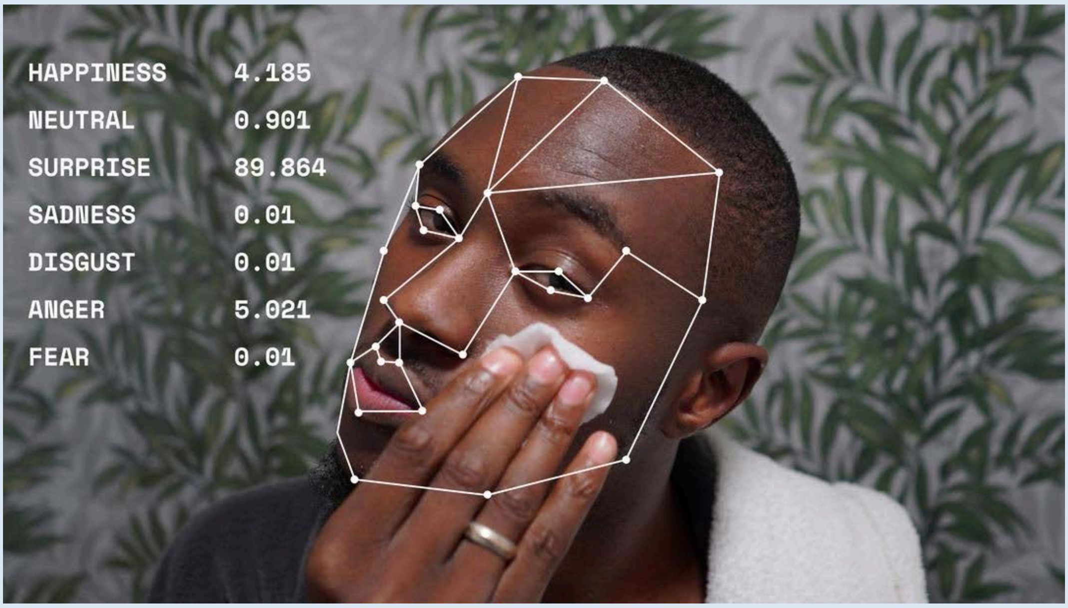
[6] It is a little difficult to make comparisons, as there have been some changes in accounting methodology over the period, including a shift from a spend-based methodology to a Life Cycle Assessment methodology (described in the 2023 Environment Report). The 2024 report puts total GHG emissions in 2019 at 9.7m tCO2e (which aligns with the 48% increase), whereas earlier reports put GHG emissions in 2019 at 12.5m tCO2e (mostly Scope 3). Data centre construction and expansion is an area of special interest for this report; the 2024 report mentions that Scope 3 Category 2 (Capital goods), which includes "manufacturing and assembly of servers and networking equipment used in our technical infrastructure, as well as emissions from materials used in the construction of data centers and offices," made up 11% of total emissions (Google 2024).

Allen (2024) comments:

> With a single generative AI query consuming nearly 10 times the amount of power as a Google search - and Google, as well as other tech giants, integrating the technology into every area of their businesses - the gigantic power spike is unsurprising. It might also be unsustainable, with power grids around the world already struggling to cope with modern levels of demand.

All else being equal, **an AI-centric future is probably a cloud-centric future**. The AI revolution of recent years is closely connected to development in cloud computing infrastructure, particularly from major providers like AWS, Azure, and GCP. These cloud giants offer scalable and flexible infrastructure, supporting the deployment and training of complex AI models. Their platforms provide various AI services, including Machine Learning frameworks, pre-built models, and data storage solutions, which aid in the development and implementation of AI applications. As always, there is nuance—including more lightweight forms of AI in development, and the possibility of shifting more workloads to more nearby, smaller data centres ("the Edge") to increase the potential for heat reuse—but by and large we can say that **more AI means more cloud.**

The **sustainability benefits of AI** are also being widely discussed. Here we have two key messages. First, it appears that the conversation around AI for sustainability is not in a great state—later in this report, we analyse some papers and reports whose claims turn out to be questionable at best. Second, one way of improving these conversations is by making distinctions between different kinds of AI, and between different kinds of climate action. See 'AI for Sustainability?' later in this report.

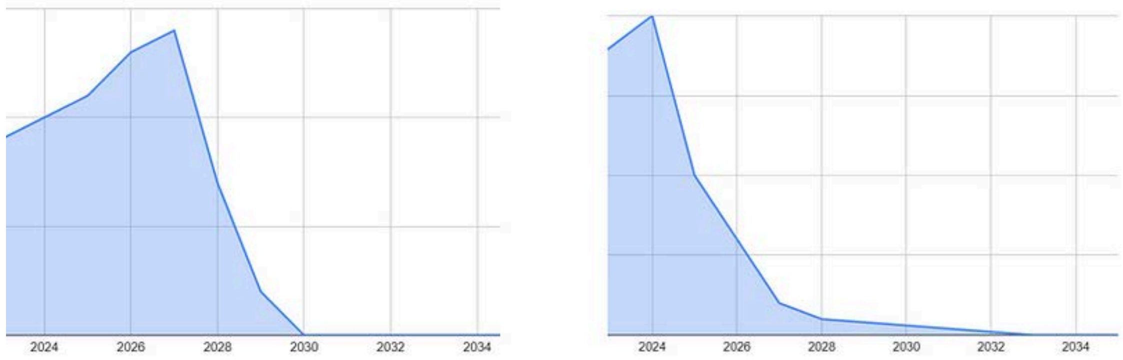


'Mirror D' by Comuzi

Tech’s pathway to net zero

Reading this report, you could be mistaken for thinking that IT is a key driver of climate change. Actually, the biggest impacts come from things like agriculture, heating and powering buildings, industrial manufacture of steel and cement, and road transport.

Nevertheless, it is crucial to recognise that **the tech sector needs to rapidly decarbonise**. Climate transition is not a matter of dealing with the “worst culprits” first: all sectors need to work together to decarbonise their own activities and the global economy. It’s all hands on deck. Global carbon emissions must peak immediately and begin a very swift descent to give us a chance of meeting Paris Agreement—the climate target almost every government in the world has signed up to. Early action is essential, as the pathway to decarbonisation is as important as the particular date by which net zero is reached.



Illustrative decarbonisation pathways: in the scenario on the left, a company meets a 2030 target, but still emits more carbon in total, because the peak is so late. In the scenario on the right, a company misses its 2030 target, but it still emits less carbon in total, because it peaks relatively early.

That is why it is disappointing to see the cloud giants downplaying the carbon impact of AI and the cloud. In 2024, we have heard that despite recent backsliding, they still have every intention of meeting future net zero targets. This ignores the importance of pathways to net zero: we need a steep descent, starting as soon as possible (see the illustrative diagrams above). The words of Google’s Jeff Dean that progress is “not necessarily a linear thing” does not appear to reflect the importance of an immediate steep reduction pathway (quoted in Goldman 2024). (See also Microsoft).

The cloud giants have used analogies to remind us that they aren't the biggest polluters on the planet. But it doesn't really matter if AI and the cloud cause less pollution than buying a hamburger or watching TV. Raising animals for food and using energy at home are also urgent issues for decarbonisation and deep demand reduction, and we shouldn’t think of them as small or unimportant.

What *does* matter is how the carbon impact of AI and the cloud relates to the amount of carbon in the atmosphere, and our options for getting to net zero rapidly, safely, and justly. It matters how it relates to the cloud giants' own decarbonisation pledges, developed and implemented in good faith in alignment with climate science (IPCC, SBTi).

Freitag et al. (2021) estimated that 2020 global emissions from Information Communication Technology (ICT) could be around 2.1%–3.9%. The same paper also estimated that if ICT emissions were to remain stable, and global emissions to decrease in line with 1.5 degrees, by now ICT would be accounting for about 4.4% of the global carbon footprint.[7] However, neither of these assumptions is a good reflection of the past five years: it is probably time for a reassessment, although that's out of the scope of this report. Mytton and Ashtine (2022) note "a wide range of estimates [of data centre energy footprints] with challenging-to-validate calculations that make it difficult to rely on their subsequent estimates."

It has been argued that AI's recently revealed sustainability potentials are so great, that the industry does *not* need to shrink its carbon footprint yet. On this reasoning, rising emissions are acceptable for the time being. However, we should be cautious of the false dilemma, "you're either for AI or against it." Instead, we should always try to **distinguish among different types of AI**, and **distinguish among different types of sustainability benefit**. We should also recognise who is driving this narrative—predominantly big tech companies who are failing the climate pledges they made only a few years ago—and ensure that policy is formed on the best possible evidence, not on vague promises and optimism. See 'AI for Sustainability?' Similar arguments apply to other aspects of digital technology, not just AI-driven applications.

---

[7] The Green Software Foundation suggests that software-related emissions account for 4-5% of global emissions (GSF 2024), although they cite Freitag et al. (2021) in support of this.

First, second, and third-order effects of IT

Sometimes IT's impacts are divided into first-order and second-order (and sometimes third-order). **First-order effects** are the direct environmental impacts of producing, using, and disposing of IT hardware and infrastructure. This includes the energy consumption and carbon emissions associated with manufacturing devices such as computers, servers, and smartphones, as well as their operation. Additionally, it covers the environmental costs related to data centres, which require significant energy to power and cool, and the e-waste generated when IT equipment is discarded. "**Second-order effects**, also known as 'secondary' or 'indirect' effects involve the impact ICT has on the environment resulting from its ability to transform processes" (Charfeddine and Umlai 2023). **Third-order effects** are less well-defined. Sometimes they're essentially everything else: the long-term influence of IT on society and environment, including how it shapes our values, behaviours, and our sense of the future. A distinction is also sometimes used between second-order effects as changes to the behaviour of the system, and third-order effects as changes to the system itself. Sometimes third-order effects are associated more narrowly with rebound effects (see Jevons' Paradox). Although this terminology isn't really used in this report, it's good to be aware of.

Photovoltaic power station

Networks and user devices

This report focuses primarily on data centres, but network infrastructures and user devices (laptops, phones etc.) make up a very large part of IT's environmental impact. Freitag et al. (2021) review surveys suggesting that neworks and user devices could contribute anywhere from 59% to 78% of IT's total carbon emissions. According to the International Telecommunication Union (ITU), there were more than 8.58 billion mobile subscriptions in use worldwide in 2022: more phones than people.

The Cloud and the Climate

Sustainability Potentials of the Cloud

Let's say a company wants to improve the environmental sustainability of its digital operations. What advice might it encounter? Standard advice includes things like monitoring your digital carbon emissions, optimising the code of software you use, minimising the number of devices you use, extending the life-cycle of those devices, cleaning out data you don't really need to store, encouraging everyday responsible behaviours (like plugging into ethernet connections where possible), engaging your IT suppliers about their own carbon emissions … and **migrating to the public cloud**.

Cloud computing involves the delivery of computing services—including servers, storage, databases, networking, software, and analytics—over the internet. The biggest three Cloud Service Providers are Amazon, Microsoft, and Google (the "cloud giants"). Cloud computing allows users to access and store data, run applications, and manage infrastructure remotely, without the need for local hardware and software.

This report tends to focus on businesses, but it is worth mentioning the public sector as well. In 2013, the UK government adopted a Cloud First policy, effectively making the public cloud the default for all central government procurement decisions, and requiring other solutions to be justified according to relatively detailed criteria. The Cloud First policy was also reassessed in 2019 and no major change of direction was announced.

Sustainability is not the only, or even the main reason, why a company might migrate to the cloud. Companies may be looking for cost savings, scalability and flexibility, security, or other benefits. However, sustainability is frequently mentioned as a benefit to cloud migration and, on the flip side, cloud migration is frequently recommended to companies looking to improve their digital carbon. In 2020 the Royal Society suggested:

> Organisations have achieved further energy efficiencies in data infrastructures by moving their data storage and processing from local servers to the public cloud, i.e. in remote data centres, although a large proportion of computing still happens within enterprise servers – smaller, noncentralised servers [...] Data centre management can reduce emissions in several ways [...][8]

Cloud computing can improve carbon efficiency, but there are some big 'buts.' The cloud giants downplay these complexities, and **strongly push the sustainability benefits of cloud migration**. A 2022 article on Amazon's website states, "when compared to surveyed enterprise data centers across several geographic regions, AWS can lower a customer's carbon footprint by nearly 80% today and up to 96% once AWS is powered with 100% renewable energy, by 2025."

---

[8] The Royal Society, 'Digital Technology and the Planet: Harnessing Computing to Achieve Net Zero', 2020, p. 77.

These claims are based on a 2019 Amazon-commissioned report by 451 Research, which surveyed 302 large US-based companies, and used data from AWS' US-based operations and the Uptime Institute (part of the 451 Group), to promote the decarbonisation benefits of cloud migration. The report suggested that AWS infrastructure was "3.6 times more energy efficient than the median of the surveyed US enterprise data centers," attributing this advantage to a combination of more efficient servers, higher server utilisation, as well as an overall higher energy efficiency of AWS data centres "by design."

The methodologies used to generate these figures have shortcomings (see the next section, 'Selling Cloud Migration'). However, the principle of pooling computational resources does offer lots of potential for sustainability gains. One aspect is improving server **utilisation**. What does this mean? As a really simple illustration, imagine two small data centres, owned by two different companies, with each data centre operating at 50% capacity. Why not combine them into one cloud data centre of about the same size, owned by a third party and operating at 100%, to save on cooling and lighting? Also, even when a server isn't storing data or running computational processes, it uses up energy just by being switched on. So improving utilisation can reduce those 'overhead' energy costs.

The same 2020 Royal Society report which strongly recommends cloud migration also continues:

> Moving computing to the cloud has allowed more efficient patterns of server use [...] Centralisation of these servers allows more effective management, with the servers' load being optimised so that they do not consume energy while idle. Illustrating the scale of the issue, a 2017 survey of 16,000 enterprise servers revealed that a quarter of them were entirely idle, consuming energy but performing no useful computing operation [...] However, the best on-premise data centres can be currently as good as the average public cloud provider, with utilisation levels on the cloud reaching only 40% on average versus 30% on-premises – suggesting there is significant room for improvements [...][9]

The principle is clear that improving utilisation can improve efficiency and save resources. But *is* cloud migration, in the form that it exists, actually improving utilisation at the systemic level? This is much less clear. The reality gets complicated—a partial load may help with cooling, for example (see Cooling Data Centres). Moreover, as Russell Macdonald (HPE Chief Technologist) suggests, "Hyperscale cloud platforms are built with redundancy in mind, and are deliberately over-provisioned to provide users with access to scalable and elastic cloud resources, based on what their compute requirements are. The bare metal

---

[9] The Royal Society, 'Digital Technology and the Planet: Harnessing Computing to Achieve Net Zero', 2020, p. 77.

infrastructure that underpins the cloud services we use is therefore poorly utilised – often less than 30% – despite the high levels of automation in cloud datacentres" (quoted in Donnelly (2024)). It might also be argued that a company with an on-prem data centre has some incentive to keep a lid on hardware standing in reserve, whereas a cloud giant has a different incentive: to keep growing and to keep finding new customers to sell to.

Another key advantage of cloud migration could be powering your IT with more **renewable energy**, since the cloud giants are all big purchasers of renewable energy. This can reduce your Scope 3 impact (the 'indirect' greenhouse gas emissions that occur in your value chain).[10] Here again, there are complexities (see 'Green Data Centres'). Despite some disagreement among the cloud giants, it appears the older, discredited, market-based approaches to carbon accounting are losing traction, and a new "24/7 hourly matching" carbon accounting paradigm is emerging. The cloud giants have had a big hand in inventing the new paradigm, too, and it has yet to receive sufficient independent scrutiny and analysis. In terms of renewable energy production, despite rapid growth globally, the US Energy Information Administration predicts higher energy-related CO2e levels in 2050 under almost all scenarios (EIA 2023).

Further advantages to cloud migration might include **scalability** (e.g. dynamically adjusting resources to meet demand — so closely related to utilisation), as well as cost-efficiency (paying only for the services you use, avoiding up-front costs of hardware), accessibility (accessing your data and applications from anywhere, or at least anywhere with an internet connection), and resilience (you can buy all kinds of backup and recovery options). Cloud service providers might also be able to achieve better **hardware refresh rates** (i.e. extending the lifespan of hardware) compared to you. This is important for sustainability because building new hardware means more carbon emissions and resource use. Arguably, the range of services and support provided by the cloud giants might even reduce barriers to entry in various industries, and promote innovation.

Private clouds

In casual conversation, it is not always clear what "the cloud" means. We can make distinctions between the public cloud, private clouds, and hybrid clouds / multicloud. A private cloud is often run in one or more enterprise data centres that are on-prem (i.e. on the premises of the organisation itself), but it could technically be a single-tenant environment managed by a third party in a remote data centre, e.g. on a colocation basis. There are even 'virtual' private clouds deployed within a public cloud infrastructure, such as Amazon VPC: "You can run virtually any type of workload in the AWS Cloud. However, if you require greater control and isolation, you can run a virtual private cloud using Amazon Virtual Private Cloud (Amazon VPC). [...] Amazon VPC is a service that lets you launch AWS resources in a logically isolated virtual network that you define." See also "Types of Data Centres"

---

[10] Carbon emissions are divided into Scope 1, Scope 2, and Scope 3, in line with the Greenhouse Gas Protocol for monitoring and reporting on carbon impacts.

But there are also a lot of hidden complexities. When applied inappropriately or badly managed, **cloud migration may make environmental impacts worse, not better**. Uncritical acceptance of cloud migration does now appear to be shifting. In OpenText's 2023 survey, only around half of companies said that sustainability was moving them to a **cloud first approach**.[11] Careful reflection around such decisions is key, as is exploring the alternatives. For example, as David Heinemeier Hansson, co-owner and Chief Technology Officer of the tech company 37Signals wrote in 2022:

> We've run extensively in both Amazon's cloud and Google's cloud. We've run on bare virtual machines, we've run on Kubernetes. We've seen all the cloud has to offer, and tried most of it. It's finally time to conclude: Renting computers is (mostly) a bad deal for medium-sized companies like ours with stable growth. The savings promised in reduced complexity never materialized. So we're making our plans to leave. (Hansson, 2022)

37Signals is not alone in returning from the cloud (Robinson 2024).

Adopting GreenOps practices is one way that companies who are using the cloud can try to ensure sustainability benefits are actually delivered. GreenOps isn't a magic solution to all these issues, but it does mean a company is thinking about them, and taking action where it can (See "GreenOps").

The term 'public'

The metaphors we use matter. To the casual listener, the term 'cloud' may carry associations of being lightweight, ethereal, intangible. For example, in 2020 the Royal Society was recommending moving operations "to the public cloud, i.e. in remote data centres." The term 'cloud' can encourage the false impression that IT has little or no environmental impact.

What about the term 'public'? It is interesting that we sometimes use the term 'public cloud' to refer to services purchased from cloud service providers, often the three cloud giants. 'Public cloud' carries connotations of 'public utility,' something that historically has often been either state-owned or subject to careful state regulation in the public interest. However, this is not particularly the case with the cloud giants. We might also wonder whether "moving to the public cloud" carries some implications of sharing, openness, even democracy, pluralism, the civic, the commons? In fact, of course, widespread cloud migration doesn't really mean any of these things. It means concentrating digital infrastructure resources which formerly were distributed across many different kinds of entities, and concentrating these in the hands of a small number of large commercial entities. Ironically enough, some of those that are 'moving to the public cloud' are government bodies — in these cases, what's being described is actually *outsourcing to the private sector*, while winding down existing public sector resources or opting not to invest in new ones.

---

[11] Companies were asked, "Is sustainability driving increased consideration of public cloud resources at your company?"

Selling Cloud Migration: Beyond PUE

Few reports on cloud sustainability try to convey the reality — that **cloud migration is complicated** — in ways that are lively and accessible for a non-technical audience. Reports tend to fall into two categories: relatively challenging technical information, or 'thought leadership' characterised by enthusiastic advocacy, sometimes with lots of stats but light citations, or hidden methodologies.

Some of these advocacy documents emphasise **innovation and disruption**. TechUK's Climate Action campaign at COP28 highlights the tech sector's potential to reduce global emissions. "By 2030, digital technology can cut global emissions by 15%. Cloud computing, 5G, AI and IoT have the potential to support dramatic reductions in carbon emissions in sectors such as transport, agriculture, and manufacturing." But details are often vague: see "AI for Sustainability?" below.

Other reports focus on **barriers to cloud migration**, in a way that makes it feel just *obvious* that you should want to migrate to the cloud. For example, UKCloud's 2020 survey revealed that public sector organisations face barriers to cloud adoption due to a lack of clear strategy, technical skills, and cost management. Cloud Industry Forum reported that 50% lack the necessary skills, prompting the industry body techUK to identify areas that could encourage cloud adoption, particularly among smaller enterprises. These include:

- Encouraging interoperability
- Providing skills and guidance for organisations,
- Improving awareness of the sustainability benefits and
- Increasing full fibre broadband provision.

More recently, Flexera's 2024 State of the Cloud report also adopts the framing of barriers to cloud adoption.

The **cloud giants continue to promote the sustainability benefits** of their cloud services. As one might expect from someone with something to sell, they are not entirely transparent about the complexities of cloud migration. An Amazon-commissioned report by 451 Research in 2019 suggested that the average US enterprise migrating to the AWS cloud would be using infrastructure that was 3.6 times more energy efficient, and could enjoy carbon emissions reductions of 88%:

> More than two-thirds of this advantage is attributable to the combination of a more energy efficient server population and much higher server utilization. AWS data centers are also more energy efficient than enterprise sites due to comprehensive efficiency programs that touch every facet of the facility. When we factor in the carbon intensity of consumed electricity and renewable energy purchases, which reduce associated carbon emissions, AWS performs the same task with an 88% lower carbon footprint. [...] 451 Research expects this carbon benefit to grow in the coming years.

However, this report places a lot of emphasis on Power Usage Effectiveness (PUE) to create these impressive statistics. Enterprise data centres do indeed often have much worse PUE scores than colocation centres or hyperscale data centres. PUE is a useful metric among others, but **PUE alone is not a good indicator of climate impact**. All it really tells you is how much energy is being used to power IT equipment versus other aspects of the data centre (cooling, lighting). If a cloud provider had a data centre with a PUE of 1.0, that would be a 'perfect' score. It would mean that 100% of the energy the data centre uses is devoted to running IT equipment, none of it being 'wasted' on supporting activity.

Does that mean that moving your operations to the 1.0 PUE data centre is definitely greener? Not necessarily. Firstly, PUE doesn't reflect the way this energy is generated. Whether the data centre is powered by coal or solar, it makes no difference to PUE. If you have the opportunity to power your own data centre from on-site renewable sources, or if the national energy grid is very green in your region, this might produce less carbon than the cloud option, even if your data centre has a worse PUE score. The 451 Research report does mention "the carbon intensity of consumed electricity and renewable energy purchases," but it turns out there are some serious issues here too (see Green Energy Procurement below).

Secondly, PUE also has very little to do with the actual differences in behaviour that may occur when you switch from running your own data centre to using the cloud. If cloud migration isn't accompanied with the right training and skills (and often code refactoring too) there is a risk of inefficient practices—like accidentally leaving a Virtual Machine running long after you stopped needing it!

Thirdly, the 451 Research report defined data centre carbon emissions in a very narrow way. It focused on AWS data centre Scope 2 emissions only (emissions from electricity purchased), without taking into account Scope 1 emissions (e.g. from cooling system refrigerants or diesel engine emergency power generators) or Scope 3 emissions (including embodied emissions in buildings or hardware).[12] The actual carbon emissions attributable to these data centres is much higher. The authors comment: "Future studies may consider these views [water usage and Scope 1], as well as embodied emissions (Scope 3) in buildings and hardware for a more complete picture" (451 Research 2019). This more complete picture may be a complicated one: the net effect of switching to the public cloud also depends on what happens to the buildings and hardware of your previous on-prem solution.

---

[12] 451 Research suggest that Scope 1 emissions "do not reflect the core operational efficiency of a datacenter" because "virtually all operators need generators that run tests or when the grid fails – there is little room for differentiation" (p.18). While we have not had the opportunity to assess this claim in detail here, it is not very convincing. If Scope 1 emissions were indeed relatively consistent across a variety of different data centre designs, this in itself would not be sufficient grounds for excluding them. Such data could be useful for many purposes, e.g. for reporting purposes, to verify the claim, to identify any variation and outliers if the claim is broadly correct, to foreground potentials for innovation, and to understand the full environmental implications of using cloud compute in the first place (not just choosing between on-prem and cloud solutions, or between different cloud providers).

More recently, **AWS has begun to promise carbon reductions of up to 99%, but many of the same limitations remain**. For example, the article 'AWS can help reduce the carbon footprint of AI workloads by up to 99%. Here's how' (Amazon 2024) features a map of the world highlighting different data centre regions and offering attractive-sounding statistics: "US & Canada: 3.6 times more energy efficient and up to 99% reduction in carbon emissions when optimized." The lowest reduction on offer is 96% (Brazil). The obvious question is, more energy efficient than *what*? What baseline is used to calculate these impressive percentages, alongside the claim itself—is it a typical customer?

Good transparency should mean presenting the methodology on the same page, and not forcing the reader to search for it. It turns out the statistics come from *How moving onto the AWS cloud reduces carbon emissions* (Accenture 2023), a study which AWS commissioned from the consultancy firm Accenture. To its credit, this report does lay out more of the methodology, and it mentions that its figures don't include embodied emissions from the concrete and steel of non-IT infrastructure, or transport and end-of-life of IT components.

The issues here also go deeper. Amazon continues to **attract criticism for how it calculates the carbon emissions of its data centres** in the first place, including lack of clarity on Scope 1 emissions; counting Scope 2 emissions as zero if renewable energy Power Purchase Agreements (PPAs) are in place; omitting Scope 3 emissions; using outdated carbon intensity metrics; and reporting only on very large regions, likely obscuring the poor environmental performance of individual data centres. You can find out more about most of these criticisms later in this report. (Google and Microsoft have attracted criticisms too).

Secondly, the baseline carbon emissions Accenture and Amazon are offering here come from **abstract models of on-premise data centres**—starting with energy consumption multiplied by Emission Factors, and building in factors like carbon-free energy procurement practices, Power Usage Effectiveness, and data centre utilisation. This would be useful, if the details of these models had been included. But either way, a more robust approach would include **empirical data from actual customers**. Don't just tell us about potential sustainability improvements: show proof points where those improvements are actually being made!

For such a comparison, the more data the better. For a headline comparison, quantify the average change in carbon emissions, for companies who have moved from on-prem to AWS, and assess the variability in these changes. Accenture and Amazon could analyse the variance or standard deviation in emissions reduction across different customers, as well as estimate the proportion of customers who experience emissions savings versus those who may see increased emissions. Broad patterns and case studies could allow potential AWS customers to compare themselves to similar companies, and identify risks and opportunities. Although it is challenging to collect and interpret such empirical data, the climate crisis means the stakes are high. It would provide a clearer and more reliable understanding of the carbon savings achievable through AWS cloud services.

Within the EU, anti-greenwashing regulations are being rolled out. These might in theory impact on promises of lowering carbon emissions by up to 99%. For example, where such claims are based narrowly on data centre energy consumption, the draft text of the Green Claims Directive states that companies must

> specify if the claim is related to the whole product, part of a product, part of a life-cycle of a product, or certain aspects of a product, or to all activities of a trader or a certain part or aspect of these activities, as relevant to the claim (European Parliament 2024).

We welcome such regulation, which aims to support more informed consumer decisions. However, it appears unlikely that there is time for it to be finalised and come into force, to be tested in courts, and to reshape the economics of cloud procurement and the practices of the cloud giants, quickly enough to align with international net zero needs. Such regulation at least indicates a direction of travel, which may join other progressive factors in pushing the cloud giants to take action now to improve the transparency and usability of their sustainability data. At time of writing, Amazon and Microsoft have been recently dropped by the Science-Based Targets initiative (SBTi). In the meanwhile, however, cloud customers and other stakeholders can at least do their own homework to develop a nuanced understanding of the costs and benefits, and risks and opportunities, of using the cloud.

Which cloud?

You may have come to this report because you want a snappy answer to a question like: **Which cloud provider is the most sustainable?** Unfortunately, we can't really tell you! We can advise:

1. **Look beyond the usual suspects.** There are many specialist cloud providers (beyond the three cloud giants) who are placing great priority on sustainability.
2. **You don't have to pick just one.** Hybrid solutions are increasingly the norm, and worth exploring.
3. **Third-party tools and data can enhance your sustainability.** Approaches like GreenOps can help to improve the sustainability of your cloud operations.
4. **Your voice counts.** Whichever cloud provider you choose, the key thing is to ask them about sustainability, both at the procurement stage, and on an ongoing basis.
5. **They are trying to be sustainable in different ways.** If you are choosing between AWS, Azure, and GCP, you may wish to consider who has the most plausible philosophy around renewable energy purchase and reporting (see "24/7 Hourly Matching" and "The Cloud Giants and the Greenhouse Gas Protocol").
6. **You may get a greener provider, but fundamental questions about sustainability and social value can never be outsourced**. No matter how green your cloud provider is, storing data and running computations uses resources. Your cloud provider will have a short-term structural incentive to get you doing *more* of that, whereas the planet needs us to figure out how to do *less* of it.

Cloud Nuance

We've mentioned **catches and complexities:** What are they? The popularity of the public cloud as a sustainability solution, which went almost unquestioned only a few short years ago, fits into a wider trend of **convenient techno-fixes.** Climate crisis can be emotionally and intellectually overwhelming, and techno-fixes can gain a lot of traction without due consideration for their underlying scientific uncertainties.

The three cloud giants offer step-by-step decision support for cloud migration. Cloud migration can become an exercise primarily (or exclusively) in collecting technical and financial data, running an analysis, and discovering potential savings. For example:

> Google Cloud Migration Center is a unified migration platform that helps you accelerate your end-to-end cloud migration journey from your current on-premises environment to Google Cloud. With features like cloud spend estimation, asset discovery of your current environment, and a variety of tooling for different migration scenarios, Migration Center provides you with what you need for your migration. Learn more or try the Migration Center now.[13]

Scalability is often mentioned. Although cloud computing is often touted for its cost-efficiency, the "pay-as-you-go" model can sometimes lead to unexpected costs. Companies may find it hard to predict their monthly expenses, especially if they don't have the resources in terms of staff, or don't have effective FinOps / GreenOps strategies (see also "GreenOps" below).

In fact, as is now much more widely understood, **not *every* cloud migration will be greener.** It depends for example on how green the local grid is, as well as capacity for local on-site clean energy generation. When renewable energy is generated locally, or likely to be in the near future, it is certainly worth considering on-prem solutions (having your own data centre), multicloud / hybrid solutions with a considerable private component, smaller cloud providers, and/or innovative sectoral collaboration. If you do decide to rely on public cloud solutions, make sure you keep putting pressure on the cloud giants to improve their sustainability, and explore whether GreenOps and grid-aware computing approaches can help you to manage your sustainability.

The cloud giants may be big purchasers of renewable energy, but the challenge lies in the details (see the sections on 'Green Data Centres'). There are problems around how they purchase this energy, how they report on it, and the quality of data they offer to clients who want to understand the carbon implications of their cloud-based activities. This includes controversies around **Renewable Energy Certificates** for purchasing energy, around **carbon credits** for **offsetting** their carbon emissions, and around proposals to mainstream "project-based" carbon accounting methodologies. These intricacies are explored later in the report.

---

[13] Google Cloud Migration Center. <https://cloud.google.com/migration-center/docs/migration-center-overview>

In terms of the **embodied carbon** of hardware, Eric Zie (2023) writes, "Most CSPs [Cloud Service Providers] are now stating refresh cycles ranging from 3 to 6 years, not the lower refresh rates previously claimed as a result of maximising utilisation. This indicates a lower level of efficiency than originally claimed and a realisation that software applications need to be designed to make the most efficient use of the cloud."[14] Even hardware efficiency improvements need to be carefully contextualised. Aware of concerns about environmental impacts, Nvidia have emphasised energy efficiency as they unveiled their new Blackwell GPU system. But the embodied carbon of these newer systems needs to be considered also. More efficient new servers may sometimes replace older servers. But sometimes they will run side by side with these older servers, expanding energy demand, over the crucial next few years as we attempt to steeply bring down carbon emissions. And if greater energy efficiency leads to cheaper compute, this may lead to increased demand and therefore mitigate or outweigh the savings (a "rebound effect").

The documentaries *Clouded* (2022) and *Clouded II* (2024) delve into some of the complexities of cloud migration via a series of expert interviews. In the first documentary, Scott Robertson, Cloud Architect at Co-operative Group outlines the issues with migrating legacy applications.

> "You're taking a non-cloud native app and you're running it in a cloud platform. That comes with a cost overhead, so you re-engineer it. Well, if it's a compelling application like your e-commerce app, you might want to re-engineer it because that's your IP, right? There's value in doing so. It's specific to you. If it's your ERP [Enterprise Resource Planning] app, do you put that on public cloud? Well, if you do that with no engineering or change to the way that you operate and do stuff, it will be guaranteed more expensive because you can't make use of those other bells and whistles that the cloud provider gives you, with its features and its functions. If you're not making use of the elasticity, then it will be more expensive."

Then there are the shifts in habits, incentives, and affordances within an organisation. As Corey Quinn, an AWS Cloud Economist at DuckBill Group, points out: "**Every engineer you have with access to a cloud account is able to incur cost**." Bill Roth, Cloud Economist says: "You see, for example, a particular type of VM [Virtual Machine] cost eight cents an hour. Well, that seems like hardly anything unless you leave it running for three years and then it becomes a decent amount of money." Thomas Maurer, Senior Program Manager & Chief Evangelist for Azure Hybrid says, "You basically have all these capabilities you get from the cloud, and you put these capabilities on the fingertips of every developer and IT person out there." As Corey Quinn puts it, "**We are all procurement now, whether we know it or not.**" FinOps has emerged as an approach to control cloud overspend. Building in sustainability considerations has led to GreenOps.

---

[14] Eric Zie, *Decarbonise the Digital: Facts. Methods. Action* (2023), p. 74.

There is also the risk of **sustainability complacency**. Adam Turner, former Head of Digital Sustainability at Defra: “The challenge is that the industry as a whole will be saying, ‘don't worry about it, we're getting greener, we're using green energy’. And the reality is that unfortunately, due to lack of transparency from those suppliers, in many cases, the figures aren't necessarily as accurate as they should be” (quoted in Michel, 2023).

Then there is **autonomy**. Relying on an external provider also means that you are at their mercy in terms of service availability and quality. Downtimes, though rare, can significantly affect business operations. There could be restraints on creativity and innovation too, including in the sustainability space. While the cloud giants offer a range of customisation options, you are still using third-party services, which limits your control over configurations and settings compared to an on-prem solution. Different countries have varying regulations regarding data storage and transfer, which can make it challenging for international operations to stay compliant. With data protection regulations like GDPR coming into play, cloud providers are also now focusing more on compliance and offering services that allow data to be stored in specific regions.

Cloud providers also now offer specialised services for Machine Learning and other AI, Internet of Things (IoT), and more. So companies might choose providers based on specific needs rather than general cloud services; in practice of course, it’s often about choosing specific products or packages offered by one of the three cloud giants.

Of course, there are plenty of other factors (not just doing something in the cloud or on-prem) that influence the environmental impacts of a digital activity. Eric Zie (2023) points out that “a poorly designed cloud deployment running on a very energy-efficient public cloud provider can generate more carbon than a fully optimised architecture running within a traditional data centre.”[15]

We also might want to think about how different approaches to the cloud might impact the life cycles and energy use of **network infrastructure** and **user devices**. Here too there are complexities. For example, whereas it might be intuitive to assume that network energy consumption is mainly determined by data transfer levels, there is evidence that "most of the total power consumption is unrelated to usage" (Mytton et al. 2024).

Clearly any company should do due diligence before rushing into a particular cloud solution. In general, there are often significant sustainability benefits to be realised from such a shift. But it is important to contextualise the benefits that an individual company may realise within a more **systemic picture of digital technology and decarbonisation**. Bresnihan and Brodie (2023) write, in an Irish context of proposed public incentives to construct data centres alongside bog reclamation projects for carbon sequestration:

> Like colonial improvement of the bogs, big tech is enrolling new landscapes into an emerging ‘smart’, ‘green’ supply chain, a ‘future’ which is seen to be already here. Within

---

[15] Eric Zie, *Decarbonise the Digital: Facts. Methods. Action* (2023).

this future, data centres appear to be a central and inevitable part while the existing communities, ecologies and practices become ‘waste’, excess life that must be relegated to the past or managed and extracted from to conjure a ‘green’ energy culture centred on global tech. [...]

The direction that regulatory and corporate frameworks around smart, green economies have travelled in the past decade tells us that more big tech intervention will dictate climate action without directed action by activists, scholars and policymakers to not only imagine but also to enact something different.

The Cloud in Context - A Reading List

Au, Y. (2022). ‘Data Centres on the Moon and Other Tales: A Volumetric and Elemental Analysis of the Coloniality of Digital Infrastructures’. *Territory, Politics, Governance*. <https://doi.org/10.1080/21622671.2022.2153160>.

Bassett, C. & Roberts, B. (2023). ‘Automation anxiety: a critical history-the apparently odd recurrence of debates about computation, AI and labour.’ *Handbook of Critical Studies of Artificial Intelligence*. Edward Elgar Publishing.

Bladow, K. & Ladino, J. (2018). *Affective Ecocriticism: Emotion, Embodiment, Environment*. Lincoln: University of Nebraska Press.

Blum, A. (2013). *Tubes: A Journey to the Center of the Internet*. New York: Ecco Press.

Brevini, B., Fubara-Manuel, I., Le Ludec, C., Linaa Jensen, J., Jimenez, A., Bates, J. (2024). 'Critiques of Data Colonialism,' in *Dialogues in Data Power: Shifting Response-abilities in a Datafied World*. Policy Press.

Bresnihan, P. & Brodie, P. (2021). ‘New Extractive Frontiers in Ireland and the Moebius Strip of Wind/Data.’ *Environment and Planning E: Nature and Space* 4(4), 1645–1664. <https://doi.org/10.1177/2514848620970121>.

Bresnihan, P. & Brodie, P. (2023). ‘Data Sinks, Carbon Services: Waste, Storage and Energy Cultures on Ireland’s Peat Bogs.’ *New Media & Society* 25(2), 361–383.

Brodie, P. & Velkova, J. (2021). Cloud Ruins: Ericsson’s Vaudreuil-Dorion Data Center and Infrastructural Abandonment. *Information, Communication and Society* 24(6) #AoIR2020 Special Issue, 869–885. <https://doi.org/10.1080/1369118X.2021.1909099>

Brodie, P. (2020). ‘Climate Extraction and Supply Chains of Data.’ *Media, Culture and Society*, 42(7– 8), 1095–1114. <https://doi.org/10.1177%2F0163443720904601>

Brown, I. (2023). ‘Expert explainer: Allocating accountability in AI supply chains.’ Ada Lovelace Institute, June 29. <https://www.adalovelaceinstitute.org/resource/ai-supply-chains/>

Carpenter, J.R. (2016) *The Gathering Cloud*. <https://luckysoap.com/thegatheringcloud/sources.html>

Couldry, N. & Mejias, U.A. (2019). ‘Data Colonialism: Rethinking Big Data’s Relation to the Contemporary Subject.’ *Television & New Media* 20(4), 336– 349. <https://doi.org/10.1177/1527476418796632>.

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Connecticut: Yale University Press.

Cugurullo, F., Caprotti, F., Cook, M., Karvonen, A., McGuirk, P., & Marvin, S. (2024). The rise of ‘AI urbanism in post-smart cities: A critical commentary on urban artificial intelligence.’ *Urban Studies*, 61(6), 1168-1182. <https://journals.sagepub.com/doi/full/10.1177/00420980231203386>

Dauvergne, P. (2020). *AI in the Wild: Sustainability in the Age of Artificial Intelligence*. Cambridge, MA: The MIT Press.

Gonzalez Monserrate, S. (2022). ‘The Staggering Ecological Impacts of Computation and the Cloud.’ *The MIT Press Reader* (14 February). <https://thereader.mitpress.mit.edu/the-staggering-ecological-impacts-of-computation-and-the-cloud/>.

Greene, D. (2022). ‘Landlords of the Internet: Big Data and Big Real Estate.’ *Social Studies of Science* 52(6), 90427. <https://doi.org/10.1177/03063127221124943>.

Hu, T.H. (2015). *A Prehistory of the Cloud.* Cambridge, MA: The MIT Press.

Juckes, M., Bane, M., Bulpett, J., Cartmell, K., MacFarlane, M., MacRae, M., Owen, A., Pascoe, C., & Townsend, P. (2023). *Sustainability in Digital Research Infrastructure: UKRI Net Zero DRI Scoping Project final technical report*. Zenodo. <https://doi.org/10.5281/zenodo.8199984>

Lehdonvirta, V. (2024). *Cloud Empires: How Digital Platforms Are Overtaking the State and How We Can Regain Control*. Cambridge, MA: The MIT Press.

Messeri, L., Crockett, M.J. ‘Artificial intelligence and illusions of understanding in scientific research.’ *Nature* 627, 49–58 (2024). <https://doi.org/10.1038/s41586-024-07146-0>

McQuillan, D. (2022). *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*. Bristol: Bristol University Press.

Monroe, D. (2022). ‘Seeding the Cloud.’ *Logic(s) Magazine* (27 March). <https://logicmag.io/ clouds/seeding-the-cloud/>

Moro, J. (2021). ‘Air-Conditioning the Internet: Data Center Securitization as Atmospheric Media.’ *Media Fields Journal: Critical Explorations in Media and Space* (26 April).

Conditioning <http://mediafieldsjournal.org/air-conditioning-the-internet/2021/4/26/air-the-internet-data-center-securitization-as.html>

Narayan, D. (2022). ‘Platform Capitalism and Cloud Infrastructure: Theorizing a Hyper-scalable Computing Regime.’ *Environment and Planning A: Economy and Space*. <https://journals.sagepub.com/doi/full/10.1177/0308518X221094028>.

Pasek, A. (2019). ‘Managing Carbon and Data Flows: Fungible Forms of Mediation in the Cloud.’ *Culture Machine* (2 April).
<https://culturemachine.net/vol-18-the-nature-of-data-centers/>

Rodima-Taylor, D., Campbell-Verduyn, M., & Bernards, N. (2024). ‘Repoliticizing the technological turn in sustainability governance: Moralities, power, space.’ *Environment and Planning C: Politics and Space*, 42(5), 699-707.
<https://journals-sagepub-com/full/10.1177/23996544241258858>

Schütze, P. (2024). ‘The Problem of Sustainable AI: A Critical Assessment of an Emerging Phenomenon.’ *Weizenbaum Journal of the Digital Society*, 4(1).
<https://doi.org/10.34669/WI.WJDS/4.1.4>

The World’s Digital Memory Is at Risk. New York Times (21 June)
<https://www.nytimes.com/2023/06/21/opinion/digital-archives-memory.html>

Vallor, S., Vierkant, T. ‘Find the Gap: AI, Responsible Agency and Vulnerability.’ *Minds & Machines* 34, 20 (2024). <https://doi.org/10.1007/s11023-024-09674-0>

Vallor, S. *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*. (2024). Oxford University Press.

Vonderau, A. (2019). ‘Storing Data, Infrastructuring the Air: Thermocultures of the Cloud.’ *Culture Machine* (2 April).
<https://culturemachine.net/vol-18-the-nature-of-data-centers/the-nature-of-data-centers/>

Wang, X. (2020). *Blockchain Chicken Farm: And Other Stories of Tech in China's Countryside*. FSG Originals x Logic.

Wadhwa, B. and Verma, A. (2014). ‘Energy saving approaches for Green Cloud Computing: A review,’ 2014 *Recent Advances in Engineering and Computational Sciences* (RAECS)
.<https://doi.org/10.1109/RAECS.2014.6799608>.

Widder, D.G., West, S. & Whittaker, M. (2023). ‘Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI.’ SSRN Scholarly Paper. Rochester, NY (17 August). <https://doi.org/10.2139/ssrn.4543807>

Cloud Governance

Measuring Environmental Impacts of Cloud Usage

It is sometimes said that what you can't measure, you can't manage. Taken literally, this is obviously not *quite* true: many things can be managed without being measured. When part of your house catches fire, you may not run to fetch a thermometer. Nonetheless, measurement plays a key role in the governance of the cloud. Currently, all three cloud giants provide some **tools to track carbon emissions** associated with their client's cloud operations. As we'll see, these tools have attracted some criticism.

- Azure: Microsoft Emissions Impact dashboard, Sustainability assessment, Azure carbon optimisation service
- AWS: Customer Carbon Footprint tool
- GCP: Carbon footprint dashboard

There are also independent third party tools and services available, both open source and commercial:

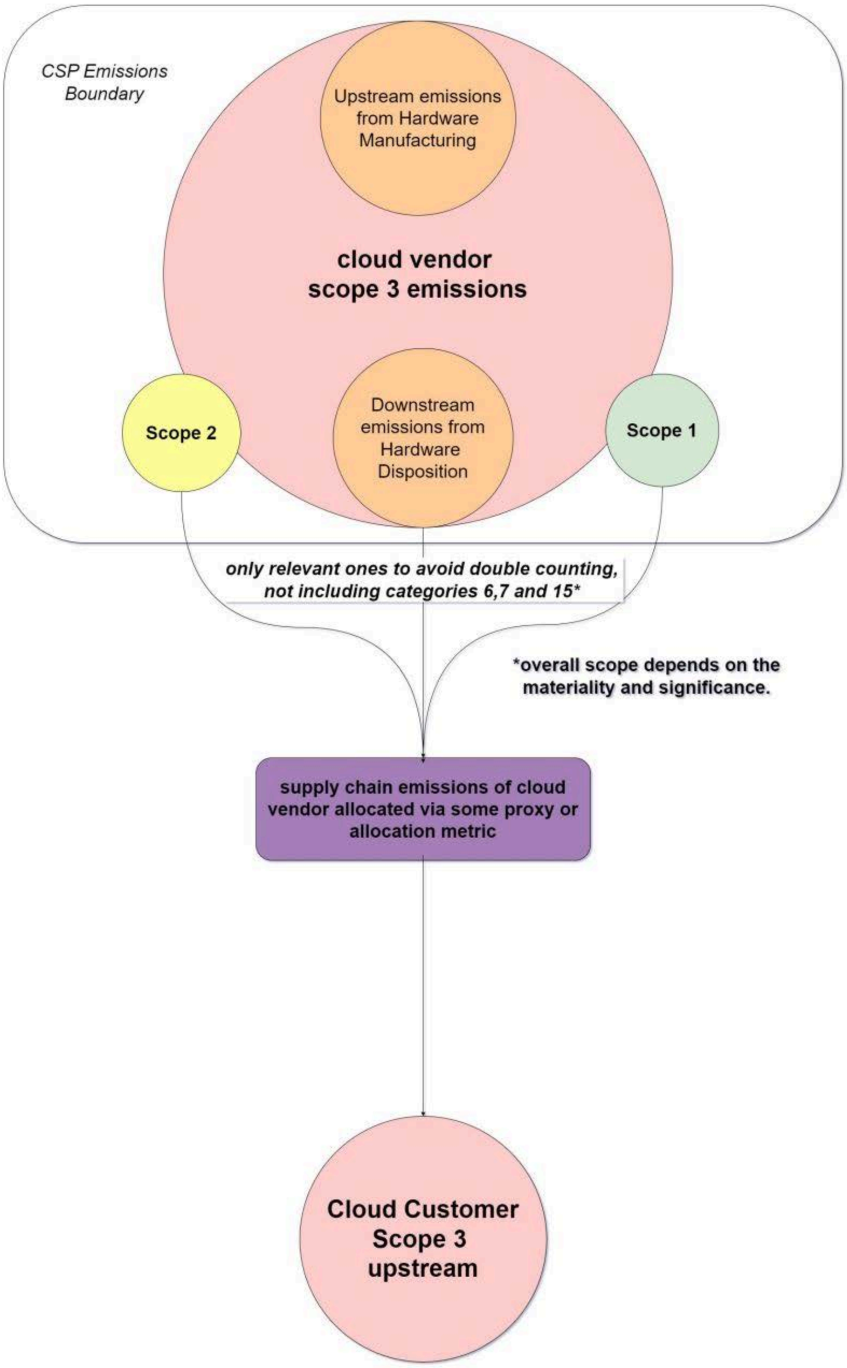
- Boavizta Cloud Scanner
- Commitments.cloud's sustainability calculator for AWS, Azure, and GCP
- Greenpixie
- Thoughtworks's Cloud Carbon Footprint tool

Open source tools like Cardamon.io are also useful in this context.

As reporting tools improve, they can empower cloud users to **voluntarily improve** the sustainability of their cloud operations — they can benchmark their cloud-related carbon emissions, make changes, and see the effects of those changes. For example, they can adopt carbon-aware or grid-aware processes, or more fundamentally change the nature of their offering. Crucially, good reporting tools also make it possible for policymakers to create and apply **legal regulation**. Various third party certifications are a kind of middle ground between voluntary action and legally mandated action.

For cloud customers, quantifying cloud carbon is currently highly challenging. Estimates are highly sensitive to the choices made regarding the assessment boundary and the rules for allocating emissions, as Boavizta (2023) describe. As Jaskaran Singh of Birchlogic comments, “Take whatever approach you are comfortable with but start somewhere. Start with the most material emissions then expand from it.” Singh also recommends addressing the uncertainties head-on, through modelling techniques like sensitivity analyses (Singh 2024).

Cloud emissions by scope



Source: Birchlogic

In 2022 the **AWS Customer Carbon Footprint tool** launched to a lukewarm response. According to Edward Targett, writing for *The Stack*, the new tool only showed "emissions data by extremely high-level geographical groupings such as EMEA (Europe, Middle-East and Africa) and AMER (North, Central and South America) not by AWS Regions themselves; a lack of precision that may frustrate some users hoping to optimise emissions reductions by swapping workloads to the AWS region with lowest emissions" (Targett, 2022). Targett further quotes AWS users who describe the tool as lacking data on embodied emissions as well as details on methodology, with insufficient granularity and transparency.

Some of this relates to the **scope** of emissions included in AWS's tool. AWS's Customer Carbon Footprint tool only reflects Scope 1 and Scope 2. Scope 1 emissions come directly from the company's activities, like running its own buildings and vehicles. Scope 2 emissions are produced by the electricity that the company purchases and uses. It omits Scope 3 emissions, the emissions produced not directly by a company but by activities in its wider supply chain, like the suppliers they work with and the customers they serve (GHG Protocol).

The push for AWS to be more transparent has been growing. Google and Microsoft have been sharing this same type of environmental data with their customers since 2021.

Here too there is room for improvement:

> Beyond the lack of transparency in Scopes 1 and 3, Microsoft's approach lacks part of Scope 3, namely emissions related to IT equipment transportation, buildings, employee commuting and also technical equipment (non-IT) of the data center.
> (Boavizta, 2023)

While AWS has been offering Scope 3 to its biggest clients under non-disclosure agreements (NDAs), the company has not yet made it freely available to all customers. Following criticisms, AWS announced that it will incorporate its Scope 3 greenhouse gas emissions (GHG) data into this tool starting in early 2024. As of August 2024, this update has not yet been implemented. Boavizta (2023) comments in AWS's defence that "AWS puts more emphasis on its Well-Architected Sustainability Pillar (GreenOps good practices) than this calculator, which is much less advanced than that of its competitors."

Although Google and Microsoft's carbon reporting tools are better than Amazon's, they also still have some limitations. Data is only available on a monthly basis, despite billing and usage data available on a daily basis. There are no water metrics. Microsoft's granularity could be improved.

For the Google calculator, Scope 1, 2, and 3 are only visible in exported data, not in the cost and usage dashboard itself. Furthermore:

> the perimeter of the Google calculator is quite transparent on the elements taken into account and those not taken into account. The difficulty in taking into account elements outside the scope may explain why they are not included (lack of methods, data, etc.). Unfortunately there is a lack of transparency on the life cycle analysis data which forms the basis of the impact assessment, meaning we cannot comment on the accuracy of this aspect.
> (Boavizta, 2023)

Cloud Giant Reporting Tools

	Scope 1	Scope 2	Scope 3	ISO-14064 verified	Water metrics	Daily reporting cadence	Reports alongside spend & usage	Resource & instance level data
AWS	✔	(✔)						
Azure	✔	(✔)	✔	(✔)			(✔)	(✔)
GCP	(✔)	✔	(✔)					✔

✅ Included
(✔) Partly included

Source: Adapted from Mark Butcher, GreenIO Presentation (September 2024)

As environmental regulations tighten, companies will need to have a clear understanding of their carbon footprint, which includes emissions from all three scopes. The whole Scope 1, Scope 2, and Scope 3 framework comes from the Greenhouse Gas Protocol, and is currently undergoing a major review. Changes to the framework are likely to appear in 2025. Such frameworks of course pay considerable attention to industry voices; the cloud giants, as very large and well-resourced businesses, play a role in writing the rulebook which they are to follow. It is important that this influence is not allowed to jeopardise a scientifically credible approach to decarbonisation.

However, criticisms of the cloud giant reporting tools go deeper than customers wanting convenient features. It speaks to the underlying material processes: how they build and operate their data centres, and how they power them.

Carbon-Aware Computing and Grid-Aware Computing

The tools offered by the cloud giants are intended to support a **carbon-aware** approach to cloud computing, including:

- **Location-shifting**: Moving your jobs to regions where the energy grid is greener.
- **Time-shifting**: Running jobs (especially jobs that are not particularly time sensitive, such as training an AI model), when demand is low (e.g. certain times of night) and/or when the proportion of renewable energy in the grid is the highest (e.g. because the sun is shining on the solar panels, and the wind is blowing in the wind turbines).
- **Demand-shaping**: This term isn't always used consistently. Demand-shaping may mean actually changing what you do, not just when and where you do it. An example of demand-shaping might be monitoring grid carbon intensiveness in real time, and switching off or paywalling non-essential features of a service when renewable energy is not available.

Demand response and **flexible computing** are closely related terms. In a demand response arrangement, a consumer such as a data centre agrees with its power provider to alter its consumption at required time-scales, receiving financial incentives in return. The power provider offers such programmes to help balance energy generation and demand. Large consumers become "flexible loads," which can also help to integrate more renewables in the grid, rather than relying on coal and gas. See Coskun et al. (2024).

See the Carbon Aware SDK for more on carbon-aware computing. In practice, **sustainability is never the only factor.** Factors such as power availability, potential future performance requirements, proximity to users, access to GPUs, and latency, may influence the choice of where to run workloads. What is the right amount of emphasis to place on sustainability? This is a challenging question. This overall framing also tends to make sustainability sound like a nice bonus if you can do it—rather than our collective effort as humans to preserve the planetary conditions within which we all live.

Moreover, there are questions around the **additionality** and the **scalability** of carbon-aware approaches. Can *we all* do our computational work when demand is lowest?—very unlikely! Even at the current scale, there are limitations to the benefits that can be realised, as Sukprasert et al. (2024) suggest:

> [...] although there is the potential for some significant carbon savings from spatiotemporal workload shifting, the benefits are often limited in practice. For temporal shifting, these limits derive from a lack of variability in carbon intensity at many locations. In addition, the locations with low variability – where temporal shifting is least effective – tend to be those with the highest absolute carbon emissions – where reducing carbon emissions is most important. Likewise, locations with significant variability tend to have low average carbon emissions, and thus adapting to such variations does not yield significant savings.

> For spatial shifting, resource constraints will likely limit much of the, potentially significant, carbon savings in practice by preventing most jobs from migrating to the lowest carbon regions. In addition, migration overheads may also prevent many large jobs that consume significant resources and energy, i.e., from processing large datasets, from benefiting from migration. [...] Of course, as the grid becomes greener our results may change.

Questions around scalability also relate to questions around **additionality**. Google's Region Picker, for example, allows the user to adjust the weighting between three desirable qualities — fast, cheap, and green — to shift workloads to services which match their criteria. In other words, you might say, users who are uninterested in sustainability are rewarded with higher quality and more inexpensive service. Another way of looking at it is that carbon-aware computing, in its current form, allows cloud giants to improve their revenues via indirect price discrimination, "a microeconomic pricing strategy where identical or largely similar goods or services are sold at different prices by the same provider in different market segments" (Wikipedia).

Smith and Velasco (2023) also ask some pressing questions around the additionality of carbon-aware computing:

> Does programming our software to responsively seek periods and locations with lower carbon intensity electricity actually make a tangible difference? Where are the studies that can prove this? If these patterns are implemented at scale, can the tech sector legitimately say it's contributed to actually reducing global carbon dioxide (CO2) emissions?

Smith and Velasco point out plausible scenarios in which scaling up carbon-aware shifting could even increase the overall carbon emissions of the grid over the course of a day. They propose evolving carbon-aware computing into **grid-aware computing**: "the refinement to current carbon-aware time-shifting or location-shifting approaches is to prioritise demand intensity first and carbon intensity second. We need to do this in collaboration with one another and local grid systems" (Smith and Velasco, 2023).

Ultimately, the incentives that individual entities face need to be aligned with the systemic picture of matching energy demand and supply, minimising unplanned spikes or drop-offs. Unplanned spikes are often dealt with by increasing the amount of dirty energy in the grid. Unplanned drop-offs may mean wasting clean energy that could be put to good uses. As Smith (2024) also points out, "the very act of rapidly ramping up or down supply adds extra emissions."

Reporting Environmental Impacts of Cloud Usage

A variety of standards, frameworks, certifications, platforms, and organisations might be relevant to monitoring and reporting the environmental impacts of AI, and/or the environmental impacts of cloud usage more generally. Reporting is important, although it is also important not to fixate on reporting as an end in itself. It is sometimes supposed that so long as accurate and transparent data is generated, the problems will fix themselves—customers will vote with their feet, investors will snap up the greenest financial assets, and regulators will swoop in and tidy up any mistakes that the markets have not fixed. In the real world, such self-correcting mechanisms often are confounded by other factors, or do not deliver change rapidly enough to meet climate goals.

To begin to map the reporting landscape, we might consider:

- **Non-financial reporting within financial reporting.** That sounds strange: let's explain! Companies are legally required to report annually on their financial activities. Some companies are also required to have their reports audited by external auditors, who can in theory independently confirm their accuracy. The financial reporting requirements, including audit requirements, depend on the nature and size of a company, as well as the jurisdiction. Generally speaking, publicly traded companies and large private companies are subject to fairly strict financial reporting requirements. Their financial reports are annually inspected and signed off by an external auditor. In recent years, there have been moves to integrate non-financial disclosures (specifically carbon emissions and climate-related risks) into the financial reporting process. The idea is to make sustainability reporting more mandatory and robust. This is where the TCFD, TNFD, and ISSB come in. Some companies are voluntarily fulfilling these requirements, in order to get used to the process in advance.
    - *Climate risk management*. Financial institutions also include climate and the environment within their broader risk modelling. This tends to overlap somewhat with ESG. E.g. tools like Implied Temperature Rise are designed to show the temperature alignment of an investment portfolio with global climate transition goals.
- **ESG.** ESG stands for Environment, Social, and Governance. ESG can be a somewhat vague term. On the one hand, ESG can be a broad term for everything a business does to try to operate in an ethical and sustainable way — a sort of spiritual successor to Corporate Social Responsibility (although CSR continues to be used too). On the other hand, ESG may refer more specifically to the ESG ratings provided by ratings agencies like MSCI, Sustainalytics, S&P, Refinitiv, and others. These scores are used by investors and other actors in the financial markets, e.g. they may determine whether a company can be included in an ethical investment fund's portfolio. Ratings agencies use

various methodologies to calculate ESG scores, and there are varying degrees of transparency. ESG has had an extremely troubled few years, and there remains some fundamental confusion about its purpose. ESG is about risk management.[16] If an organisation is responsible for a negative impact, but this impact is not actually a risk to the organisation, then it will tend not to show up in ESG ratings. However, ESG is sometimes treated as though it offers a comprehensive account of organisations' social and environmental impacts. Given that different ESG ratings providers pay different degrees of lip service to the full impacts of the companies they rate, this also means that ESG ratings diverge considerably, which can further undermine the credibility and usability of ESG information.[17]

  - *Sustainability reporting*. Closely linked with ESG are the various sustainability reporting frameworks that organisations use. Some of the most important ones are the Global Reporting Initiative (GRI), CDP (formerly known as Carbon Disclosure Project).
  - See Tkachenko (2024) on integrating AI carbon footprints into risk management.

- **Responsible AI.** There are various laws, regulations, and best practice guides for developing and deploying AI in socially responsible ways. These sometimes touch on environmental considerations, although often the focus is more on things like algorithmic bias, transparency, and explainability. Algorithmic Impact Assessments take various forms, often a lengthy questionnaire, and sometimes also stakeholder / expert consultation requirements. The UK has taken a very light touch approach to AI regulation compared to the EU, although there are indications that the new 2024 UK Labour government may change this.

- **Science-Based Targets initiative.** The IPCC is the world scientific authority on climate change. While it is limited in what it can say by its mandate to stay politically neutral, it is very clear about the need to act swiftly and decisively to reach net zero and bring global warming to a halt. Many organisations have therefore pledged to transform what they do to reach net zero by a particular date. The Paris Agreement is a legally binding international treaty on climate change that was adopted by most of the world's countries. Its goal is to limit global warming to well below 2, preferably to 1.5 degrees Celsius, compared to pre-industrial levels. As more and more laws and regulations come into force to implement the Paris Agreement, any organisation that doesn't accomplish net zero should eventually go out of business. The Science-Based Targets initiative is a collaboration between the CDP, the United Nations Global Compact, World Resources Institute (WRI), and the

---

[16] Julia Binder, 'Let's Be Clear: ESG is not 'Woke' and It's Different from Sustainability' (2023). <www.imd.org/ibyimd/magazine/lets-be-clear-esg-is-not-woke-and-its-different-from-sustainability/>

[17] Florian Berg, Julian F Kölbel, Roberto Rigobon, Aggregate Confusion: The Divergence of ESG Ratings, *Review of Finance*, Volume 26, Issue 6, November 2022. <doi-org/10.1093/rof/rfac033>

World Wide Fund for Nature (WWF). It aims to encourage businesses to set ambitious carbon reduction targets in line with the latest climate science to meet the goals of the Paris Agreement. Alignment with SBTi is key for avoiding greenwashing in carbon reporting. SBTi comes under regular pressure from industry to water down its requirements and diverge from its scientific methodology.

- **Technical projects.** Organisations may also be approaching questions of cloud sustainability from a more technical angle, e.g. to improve the robustness of their sustainability reporting / carbon accounting, as part of the evolution of FinOps into GreenOps, because they are invited to do so by their cloud providers or other stakeholders, because they are exploring strategic opportunities and risks and want better data, because it is part of their procurement process, etc.

- **Strategic positioning.** Let's not forget about internal reporting. Organisations may be motivated to work on a sustainability initiative as e.g. a marketing edge within their sector. Incentives are focused on performance relative to competitors using cherry-picked methodologies, and associated reporting reflects this focus.

Do We Need A Business Case For Climate Transition?

"The business forecast is for more and more cloud with a chance of significant savings … financial and carbon," enthuses Accenture in a recent thought leadership piece.[18] Sustainability is very often aligned with cost savings. Many companies have migrated to the cloud or have adopted FinOps practices purely for financial reasons, and inadvertently reduced their carbon footprint. Many sustainability initiatives, not just in the cloud computing space, follow a pattern of increased short term costs, long term reduced costs. This reflects a broader trend observed across various sustainability initiatives, of an initial increase in costs followed by significant long-term savings. As a result, it is common for businesses to consider the return on investment (ROI) when evaluating sustainability efforts. This approach not only makes financial sense but also helps to reduce negative environmental impacts, thereby aligning economic incentives with ecological responsibility. As markets increasingly value sustainability, businesses that proactively embrace these practices may gain a competitive edge. Consumers and investors alike are becoming more environmentally conscious, rewarding companies that demonstrate a commitment to sustainable practices. This shift in market dynamics further encourages businesses to integrate sustainability into their core strategies, not just as a moral or environmental act, but as a foundational element of sound business management. In our conversations with Chief Sustainability Officers, sustainability consultants, sustainability champions, and others, this point recurs again and again. To drive change, speak to senior

---

[18] *The Green Behind the Cloud* (2021).

management, the team responsible for delivery, or to investors or other stakeholders, one needs a business case. The triple bottom line is all very well, but unless you can show value for the single bottom line, don’t waste your breath.

… right?

So the story goes. There are aspects of this story to stay. But some of it is looking very old-fashioned. The fact is, this story has been tested over the past several decades, and we now have the data to reject it. It implies a pace of change which is now considerably slower than what is necessary. Certainly as an overall framing it is intellectually bankrupt.

Commercial actors are unlikely to adopt more sustainable practices, unless there is a business case for it. If there is not a business case, the expectation is that the responsibility lies with policymakers to adjust the incentive environment (for example by reporting requirements, other regulations, taxes and subsidies, etc.). If policymakers are unable to adjust the incentive environment appropriately, the expectation is that democratic processes will replace them with policymakers who can make such adjustments. If democratic processes are not fit for this purpose, then the expectation is that progressive political actors, civil society actors, and the public at large should work to improve them.

However, environmental crisis operates on specific timescales. Climate change is a vivid example of this. The model that many individual business cases will eventually create a sustainable economy is not self-evident — it depends on the assumption that the necessary challenges can be met within a given timescale. Evidence from the past half century strongly indicates that this is not happening.

Actually, there are already some promising signs that such attitudes may be shifting in the climate space. Companies increasingly recognise that climate leadership can be a key factor in attracting and retaining top talent and maintaining competitiveness. Although this does fit into a business case paradigm, it also speaks to something more fundamental — a business is made up of people, and those that survive and thrive in the future will be those for whom environmental sustainability is a basic need.

The nuances may be important too. The nature of a ‘business case’ may be evolving. For example, when we spoke with the Chief Sustainability Officer of one large technology company in early 2024, we noticed a few new twists on these old themes. This CSO emphasised the importance of reporting and compliance not only to drive change, but also to clarify priorities to support sustainability initiatives beyond the mandatory. A strong interest in systems thinking also came through in our conversation, and a sense that the business case probably does exist for every worthwhile sustainability policy, although it may require ambitiously holistic thinking to identify.

Furthermore, as described by Matthew Gitsham, Professor of Sustainable Development and Director of the Ashridge Centre for Business and Sustainability at Hult Ashridge Executive, “There is something around assuming that there is now a widespread view that decarbonisation is a good thing that we're all working towards. I do hear this more and more. Just assuming in the language that we all need to decarbonise, we know we all need to decarbonise, all organisations are moving towards decarbonisation. In the past, these sorts of proposals were written in terms of, 'This is going to save you money in this way.' But I am seeing it more that decarbonisation is just the norm. We assume you are working towards decarbonising, because everyone else is. So here are some things that are going to help you do that. Saving money and regulatory compliance also come into it, but not as top-level considerations."[19]

Meanwhile, leading consulting firms (e.g. KPMG, Deloitte, Accenture, PwC, McKinsey, BCG) have often promoted cloud migration as a sustainability solution, but typically lack expertise or incentive to consider the wider political, ethical, cultural, and economic contexts of the cloud (cf. Mazzucato and Collington 2023). In 2023-2024, several of these firms have enthusiastically promoted the expansion of AI throughout practically all economic sectors. "Boston Consulting Group and McKinsey are expecting 20% and 40% of their consulting revenue respectively to come from advising clients on how to use Generative AI in 2024" (Smith and Adams 2024). An excessively technical focus on "what clients want" could have disastrous consequences for people and planet. For some alternative perspectives, see The Cloud in Context.

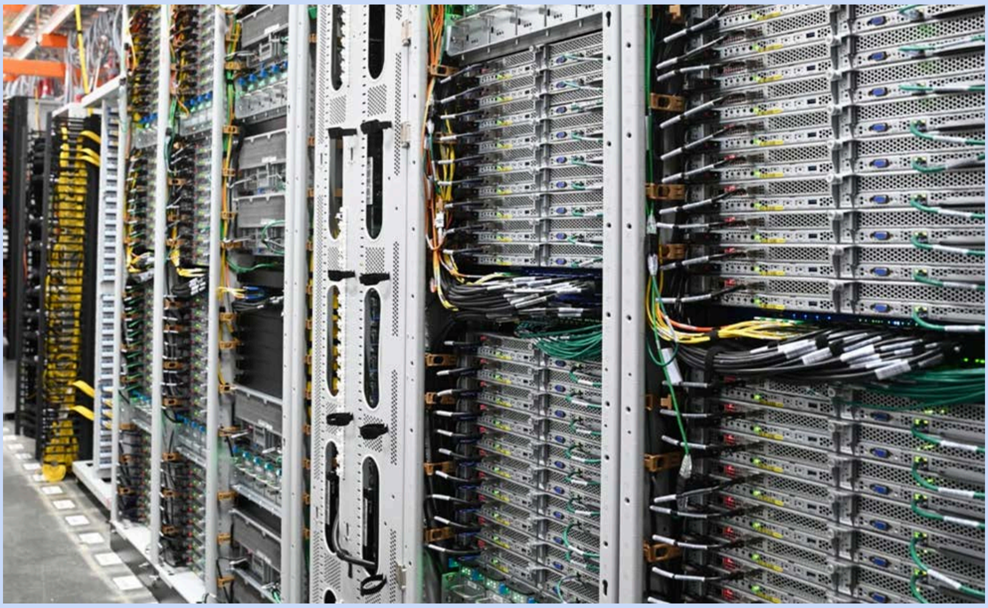
Types of Data Centres

There is quite a bit of variation across different data centres, making generalisations challenging. However, the table below can help identify the main types of data centres. A company may choose to do its data storage and processing itself in its own enterprise data centre, often located wherever it does most of its business (on-prem). This may be as small as a closet or as big as a dedicated building. Or it may want to own and maintain its own IT equipment, but have somebody else maintain the right environment (power, cooling, security) — that's the colocation solution. Or it might go for the public cloud (dominated by the cloud giants), which means its data will probably be in a hyperscale data centre somewhere. Much of this report pertains to hyperscale data centres. Edge data centres are smaller data centres which can improve latency (see Bamforth 2020, Mansouri and Barbar 2021). A company also might decide it wants a Managed Service Provider (MSP). MSPs offer a broad range of IT services beyond cloud computing, including network management, cybersecurity, support and maintenance, and sometimes, management of client IT infrastructure hosted in third-party data centres or on-prem. So an MSP may function like an additional layer between the company and the various options it has for hosting its data and applications. The MSP may offer a fully managed data centre that is implemented in a colocation data centre or a hyperscale data centre.

---

[19] Unpublished interview, 2024.

What is a data centre?



What goes on inside a data centre? A data centre contains lots of servers. A server is basically a computer: a computer that is switched on all the time (or a lot of the time). The servers in a data centre will do things for users who are physically located elsewhere. They will process and manage network requests and to run software applications, "serving" the needs of other computers or clients within a network. This includes:

- Storage: data centres are repositories for vast amounts of data, ranging from website content to business databases.
- Processing: Servers in data centres carry out complex calculations and execute applications.
- Network Management: data centres manage network traffic, connecting up servers with end-user devices (e.g. phones, laptops).

There are various **types of data centres** (see "Types of Data Centres" for more detail). Some are big, some are small. There are **enterprise data centres**, which are owned and operated by the organisation that they serve. These might be located nearby the organisation's other operations (**"on prem"**), or some distance away. Another type are **colocation** data centres, where multiple organisations share the facility and infrastructure. Usually there is a data centre company that is responsible for providing the space (cooling, power, security, etc.), and the various organisations who rent that space provide and maintain their own equipment (but lots of different arrangements exist). **Hyperscale data centres** are often giant data centres owned by cloud service providers like AWS, Google

Cloud Platform, or Microsoft Azure. So whereas colocation data centres tend to offer a space to house privately-owned hardware, with the benefits of professional facilities (e.g. the space, the cooling, power, security), hyperscale cloud data centres are more about providing fully managed IT resources and services over the internet. These also tend to be more associated with the term “public cloud”. Customers typically pay for the services and resources they use, often on a subscription or pay-as-you-go basis. AWS, GCP, and Azure all have big networks of data centres, so the subscriber’s relationship is more with the company than with a specific data centre.

There are some other kinds of data centre too. For instance, **edge data centres**, which refers to smaller data centres located near the users that they serve. **Edge Computing** (a distributed computing paradigm that brings computation and data storage closer to the location where it is needed, to improve response times and save bandwidth) also emerged as a new trend to reduce latency and improve performance. By processing data closer to where it is generated or consumed, edge computing can improve the efficiency and speed of services. Content Delivery Networks (CDNs), for example, help to deliver streaming video to users’ devices and to manage peaks in demand (for example, during major sporting events). Other smaller data centres might have other specialisms, e.g. certain types of compliance, security … and of course sustainability!

The evolution of the internet is tightly linked with the history of data centres. In the early days of the internet, organisations typically hosted their own servers on-site. However, as the internet grew and the demand for online services spiked, so did the need for more specialised and scalable computing environments. This led to the development of data centres.

The concept of **virtualisation** (the act of creating a virtual version of something, including virtual computer hardware platforms, storage devices, and computer network resources) emerged as a way to make more efficient use of physical hardware. Virtualisation allowed a single physical server to be partitioned into multiple Virtual Machines (VMs), each capable of running its own operating system and applications. This innovation greatly increased the efficiency of data centres and made it easier to manage and scale applications. Virtualisation can also help with **consolidation:**

> One problem with datacenter design is that it is hard to predict hot spots or uneven computer temperatures which will vary by the season, tasks undertaken, air circulation, and the power profile of a given task. Consolidation optimizes the maximum number of virtual machines with the minimum number of physical hosts, while also turning off idle hosts (Sovacool et al. 2022).

With the advent of Cloud Computing, the need for on-premises servers was challenged. Companies began to offer Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). These models let businesses rent computing resources or software applications from a cloud service provider, reducing the need for companies to own and maintain their own data centres.

	Enterprise Data Centre	Colocation Data Centre	Hyperscale Data Centre	Edge Data Centre
What is it?	A company’s own data centre, often located on its own premises (“on prem”).	Usually a data centre that provides space, power, cooling, and security. The customers usually provide and maintain their own hardware. Some customers may be Managed Service Providers, who then provide cloud / data centre services to other businesses.	Massive data centres, often owned by AWS, GCP, Microsoft Azure. In principle highly responsive to customers’ computing needs in real time: can scale up or down as required. Some customers may be Managed Service Providers, who then provide cloud / data centre services to other businesses.	Smaller data centre, located near end users. Localised processing to reduce latency, improve speed for specific tasks.
Who uses it?	Organisations who want complete control over their own data and IT infrastructure. And/or organisations with significant investment in on-prem solutions which they wish to continue with.	Organisations seeking a bit of flexibility and scalability without full data centre responsibilities. Also Managed Service Providers.	Companies requiring vast compute/storage, typically large enterprises, or Managed Service Providers. Or any company who wants the scalability and other features of the public cloud.	Many possible use cases. Real-time applications, Internet of Things, Content Delivery Networks, etc.
Who owns it?	Owned and operated by the organisation for private use.	Owned by a provider, but customers usually manage their own equipment. Various arrangements are possible though.	Usually owned and operated by cloud giants or major tech companies.	It depends - cloud giants, smaller cloud providers, telecommunications companies, Content Delivery Networks, enterprises.
Services Offered	Mainly Infrastructure-as-a-Service (IaaS).	IaaS through customer's equipment. Some colos also may offer Platform-as-a-Service (PaaS) or Software-as-a-Service (SaaS).	Broad range, including IaaS, PaaS, and SaaS.	Primarily IaaS and PaaS for local processing needs.
Relevance to MSPs?	Not so much, although an MSP might also maintain on-prem data centres, if that’s the agreement.	MSP may use a colocation company to house and manage clients' hardware in a secure, scalable, and high-availability environment, without owning the data centre infrastructure.	Yes, the public cloud could provide the infrastructure backbone for MSPs' cloud offerings.	Yes, MSPs could use these in delivering edge computing solutions.

What about the term **private cloud**? Is that the same as enterprise data centre and/or "on-prem" data center? Not exactly. A public cloud is used by lots of organisations. A private cloud is used by just one organisation. However, a private cloud can be hosted on-premises, or off-site in a colocation data centre. It can be managed either by the organisation itself, or by third-party providers such as a MSP. You can even sort of get a private cloud in the public cloud — either dedicated equipment reserved for one organisation, or something called a Virtual Private Cloud. A **hybrid cloud** combines public cloud computing with a private cloud or some other kind of on-prem infrastructure. A multicloud approach is similar to a hybrid cloud approach. It tends to mean the integration of multiple public clouds.

In practice, multicloud and hybrid cloud setups are the norm. According to the Flexera State of the Cloud report 2024, 59% of respondent organisations use multiple public clouds, while 73% use a hybrid of private and public clouds. Meanwhile, only 10% used a single public cloud only and 3% used private cloud only (including multiple private clouds). Investment trends for on-prem private cloud and hyperscale public cloud predict further adoption in both areas.

In their article 'The Making of Critical Data Center Studies' (2024), Dustin Edwards, Zane Griffin Talley Cooper, and Mél Hogan describe the history of the data centre, including the changing ways in which we imagine and feel about data centres. From the 1950s onward, the term data centre (or data bank) often referred to governmental statistical offices (and occasionally corporate offices) that kept records and used mainframe computing. To begin with, data centres tended to be defined by the kind of data that they kept.

> These were centers for specific types of statistical information processing. This data may have been partially processed and analyzed by mainframe computers, but, as scholars like Mar Hicks (2018; Mullaney et al., 2021) have shown, a majority of the work on and with computerized information during this time was conducted by women, particularly women of color. As such, data centers in the 1960s and 1970s were profoundly human places in which the politics of gender, race, and technology were brought into sharp relief.

Edwards et al. (2024) continue to describe the situation today:

> Today's data centers are far more depopulated, or at least they seem that way. They present as mammoth structural storehouses of digital information, within which narratives and imaginaries of human labor and maintenance tend to vanish in the face of their alleged technological sophistication and architectural vastness – houses for wires, metal, and electricity. As Taylor (2019) writes: 'The end result is an image of a technological landscape emptied of people and any obvious signs of human presence: a mechanized world of techno-wilderness'. This popular image of the data center as

a place ‘not built for humans’ drives how data centers are understood both in popular culture, and by the powerful actors who build, maintain, and regulate them. As Daniel Greene reminds us, however, ‘The internet’s infrastructure evolved with its landlords’; whether as owners or leasers of the buildings where data centers are housed, landlords strike deals and manage the internet as ‘assets’. Defining a data center in the twenty-first century does not have much to do with the data, or the people working on/with it, but predominantly with the material structure in which the data is held. The data itself is incidental – a given, almost taken for granted.[20]

Blade inspection, Colorado

---

[20] Edwards, D., Cooper, Z. G. T., & Hogan, M. (2024). ‘The making of critical data center studies.’ *Convergence*. <https://doi-org/10.1177/13548565231224157>

Sustainability Transparency of the Cloud Giants

The cloud giants control around two-thirds of cloud infrastructure globally. Other large cloud service providers include Alibaba, IBM Cloud, and DigitalOcean Cloud (CloudZero, 2024). AWS, GCP and Azure are subsidiaries of Amazon, Google (Alphabet), and Microsoft respectively. In this section we look at the broader picture of the sustainability of these three companies, with some focus on transparency.

The cloud giants are among the largest companies in the world, and play a significant role in 'writing the rulebook' around energy transition and net zero. None of these companies is a monolith either: each harbours important tensions, contradictions, and alternative visions of the future.

In some respects, the cloud giants appear to be taking climate change seriously. They are major purchasers of renewable energy, major investors in decarbonisation technology, and they have made (on the face of it) ambitious climate pledges, listed below.

Amazon	Net zero by 2040	Unclear on the role of carbon removal in pledge
Google	Net zero by 2030	Pledge includes substantial carbon removal (about 50% of 2019 levels)
Microsoft	Carbon negative by 2030	Unclear on the role of carbon removal in pledge

Source: CCRM 2024

However, **none of the cloud giants is on track to meet its own pledge**. The pledges themselves are also not entirely adequate, once we drill into the details. Amazon was one of the companies recently dropped by the Science Based Targets initiative. Microsoft has also been removed, although for somewhat different reasons.

Often the criticisms focus on **transparency.** It is true that the cloud giants could be more transparent, although this wouldn't solve everything—reporting more transparently on inadequate climate action doesn't automatically improve that action! We can ask four important questions about each cloud giant's climate pledge. Unfortunately, it is not feasible to answer "yes" to any of these questions:

1. Does the pledge itself adequately align with the Paris Agreement target of limiting global warming to well below 2.0 degrees and to 1.5 degrees if possible?
2. Is the strategy for achieving the pledge adequate?
3. Is progress on the pledge adequate?

4. Are all of the above (the pledge, the strategy, and the progress) being robustly measured and transparently communicated?

We might also add a fifth question:

5. To what extent does the company's services support and/or undermine its own climate pledges—e.g. through support for oil and gas exploration, as examined in a high profile 2020 Greenpeace report, as well as more recent reporting?

Focusing on the fourth question, transparency: according to the Corporate Climate Responsibility Monitor 2024 report, Google and Microsoft fall in the 30-40% integrity bracket, Amazon in the "unsubstantiated" bracket. Of course, these companies are not directly comparable operationally, because Google and Microsoft don't operate vast fulfilment warehouses and logistics networks in quite the same way as Amazon.

All three cloud giants now claim that their electricity is already 100% from renewable sources. What do we find when we look closer? Unsurprisingly, we find an emphasis on technological solutions. In 2021 Microsoft's then Head of Sustainability Lucas Joppa even used an analogy which compared global climate action to creating a computer program: writing the code, debugging the code, and executing it. For Joppa, the 'inputs' of this 'computer program' would be all our activities. A correct 'output' would be zero additional carbon accumulating in the atmosphere in the year 2050.

Joppa added, "Microsoft's contribution to this is both simple and ambitious. By 2030, we've committed to reduce our emissions by half or more, and then physically remove the rest from the atmosphere. And then from 2030 to 2050, to continue not just zeroing out our annual emissions, but to go back in time and remove all the emissions we were associated with since we were founded in 1975."[21]

Why do the parent companies of the cloud giants get such mediocre transparency ratings from independent reviewers like Carbon Market Watch? There are a few reasons. In recent years there has been some debate over **location-based vs. market-based** accounting methods for tracking the carbon impacts of hyperscale data centres operated by the cloud giants. Roughly speaking, location-based accounting tries to approximate the actual physical reality of energy use in a company's operations. Market-based accounting tries to capture what kind of electricity the company has bought. So the short version is, these are **complementary methods,** although if you were stranded on a desert island and could only pick one, go for location-based.

See the next section, on **green data centres**, for more details; this 2024 article from *The Financial Times* also makes clear the difference between location-based and market-based methods.

---

[21] Lucas Joppa TED Talk (2021).

In its 2024 sustainability rankings (which appeared before Google's most recent sustainability report), *Computing* suggested:

> For two years running, Microsoft Azure was judged the most sustainable cloud, but its margin of victory over second-place Google Cloud Platform (GCP) narrowed last year. This year, Google Cloud Platform is the winner, beating Azure into second place. AWS retains third place in our sustainability rankings (Horwood 2024a).

See also the individual sections on **Amazon**, **Google**, and **Microsoft** towards the end of this report. We don't attempt to compare or rank the cloud giants, nor to provide a comprehensive account of their sustainability credentials—but we do collect some useful starting points.

Burning e-waste, Agbogbloshie

Green Data Centres

There are many options for improving the sustainability of most data centres. Interventions can target both direct and onsite emissions from data centre operations as well as indirect and embodied carbon emissions embodied in the data centre lifecycle. Sovacool et al. (2022) identified no less than 40 approaches available to improve data centre sustainability.

	Solution	See also
Optimising and innovating in data centre design and construction	Green or eco-friendly design	Sustainable AI Innovation
	Green manufacturing	
	Green metrics, assessment tools, and methodology	Green Data Centre Metrics and Certifications
	Environment-related risk mitigation	
	Virtualisation and consolidation	What is a data centre?
	Economies of scale	Sustainability Potentials of the Cloud
Optimising and innovating in cooling	Hydrogen fuel cells	24/7 hourly matching
	Free cooling	Cooling Data Centres
	Hot and cold aisle containment	
	Increasing allowable IT temperatures	
	Cooling management	
	Variable air flow	
	Partial load	
	High energy efficiency components	
	Thermal energy storage integration	

Optimising and innovating in power supply	Advanced use of energy storage devices	Sustainable AI Innovation
	Direct current installation	
	Bypass UPS in normal operating conditions for improved energy efficiency	
	Modular UPS for enhanced efficiency	
	PUE enhancements	Sustainability Potentials of the Cloud
Strategies for heat reuse	Utilisation of waste heat and heat recovery	Data Centre Heat Reuse
	Liquid cooling	Cooling Data Centres
	Environmentally Opportunistic Computing	Data Centre Heat Reuse
	District heating	Data Centre Heat Reuse
	Absorption refrigeration	
Integration of renewable energy	Onsite generation from onsite renewables	Green Energy Procurement
	Onsite generation from offsite renewables	
	Offsite generation	
	Renewable energy supply from third parties	
Green disposal and waste	Responsible disposal and recycling	Appendix 1: Actions and Resources
	Eco-labeling of IT products	Eco-labels

Data centre policy and governance	Energy efficiency standards for equipment	Relevant UK policy on sustainable AI and Appendix 1: Actions and Resources
	Tax credits and procurement standards	
	Public funding for R&D	
	Regulatory compliance	
	Impact Benefit Agreements	
	Equitable taxation schemes	
	Restrictions on cryptocurrency mining	
	Require local content or employment	
	Ensuring transparent and reliable data	Sustainability Transparency of the Cloud Giants

Source: Adapted from Sovacool et al. (2022)

Green Energy Procurement

All three cloud giants now claim that their electricity is already 100% generated from renewable sources. This includes data centre electricity. However, **generating and distributing energy is complicated and messy.** The electricity which powers a data centre at a given moment is just whatever comes out of the grid, usually powered by a mix of clean and dirty energy. Renewable energy generators aren't neatly distributed around the world in a way that matches energy demand — they are heavily clustered where the sun shines, where the winds blow, and where the tides pull. That also means the availability of renewable energy varies over time. There are limits to how much energy can be stored. You won't have much luck getting your data centre powered by solar energy in the middle of the night.[22]

There are also different ways of reporting how renewable your energy is. **Market-based accounting**, which allows companies to claim reductions by buying Renewable Energy Certificates (RECs), Renewable Energy Guarantees of Origin

---

[22] Then again, the middle of the night is when overall demand tends to be low. Then again, this may change in the future as fleets of Electric Vehicles are charged overnight. So: it's complicated!

(REGOs or GOs), Green Energy Certificates (GECs), and other instruments. Market-based accounting by itself has been criticised as greenwashing. This method can lead to claims of renewable energy usage that doesn't directly correlate with the real-time energy consumption or local grid mix of a data centre. For example, on their sustainability website, Amazon writes:

> We contract for renewable power from utility scale wind and solar projects that add clean energy to the grid. [...] We also may choose to support these grids through the purchase of environmental attributes, like Renewable Energy Certificates and Guarantees of Origin, in line with our Renewable Energy Methodology.

These 'unbundled' renewable energy credits can be bought and sold multiple times on renewable energy markets before being 'retired.' They have allowed companies to legally claim to have used renewable energy, but they don't ensure any robust connection between the generation and consumption of renewable energy. They also fail to promote the development of additional renewable energy capacity. "Most wind and solar farms were built with the help of government subsidies, not REC dollars. And when the operator of a wind or solar farm earns fresh dollars from REC sales, it is under no obligation to use the money to expand capacity" (Naik 2021; see also Langer et al. 2024).[23] The greenwashing is even more egregious when a company uses renewable energy credits bought in one region to claim reduced or zeroed carbon emissions in another region.

Location-based accounting, which focuses on the average Emissions Factors (EFs) of the grid where energy consumption occurs, offers a more direct measure of a data centre's carbon intensity. It calculates greenhouse gas (GHG) emissions by considering the average emissions intensity of the electrical grid where the energy is actually used. It employs averages from specific regions, countries, or smaller grid areas, to estimate the emissions resulting from electricity use at the site of operation. There are various data sources that can help with these estimates, e.g. the National Grid, Electricity Maps, the US EPA: North American Electric Reliability Corporation, the European Environment Agency, and Country Specific Electricity Factors. Inferences can also be made from data providers like Our World in Data, Ember, and the Energy Institute. The Software Carbon Intensity (SCI) specification requires location-based accounting.

However, the market signals and incentives for renewable energy development provided by RECs and other instruments are still part of the bigger picture, so a transparent approach is to offer **both location-based data and market-based data.**

The idea behind market-based accounting is to emphasise demand signals, since these stimulate investment in green energy. Does it really matter if the renewable megawatt-hour you pay for is generated precisely when and where you were using a megawatt-hour of energy? In

---

[23] There are some counterarguments, i.e. even if additionality is low (the RECs are not funding new renewable plant construction), the expectation of RECs is already priced into the financing of existing capacity which might not otherwise have been built in the first place.

practice, it matters a lot. A company might have a high electricity demand at a time when the contribution of renewables to the grid is already at capacity. Even though this would result in an increase in non-renewable energy production, currently the company can still find RECs to purchase to legally make a 100% renewables-powered claim. There is good evidence that such practices do not lead to real additional investment in renewable energy generation (Langer et al. 2024). At present, under the GHG Protocol, RECs must come from the same region as the power consumed, but they don't have to come from the same grid or from the same time period. Amazon is pushing for these regulations to be loosened even further. RECs have also tended historically to be very inexpensive, raising concerns over their real influence on additional clean energy investment.

The difference between market-based accounting and location-based accounting can be significant. The table below shows all Scope 2 emissions (not just data centres).

	Scope 2 emissions reported in 2022 (Megatonnes of CO2 equivalent)	
	Location-based	**Market-based**
Amazon	[Unavailable]	2.89
Google	8.5	2.5
Microsoft	6.4	0.3

Source: CCRM 2024 + Amazon 2022 Sustainability Report

Power Purchase Agreements (PPAs) are an improvement on RECs and other "unbundled" instruments. PPAs are longer-term agreements to purchase renewable energy at a fixed price for a fixed period. They can roughly be thought of as a lot of RECs "bundled" together. In some instances, usually when producer and purchaser are located within the same grid region, the electricity generated through a PPA is physically delivered to the purchaser. Even virtual PPAs serve to facilitate the development of new projects by providing financial stability and long-term price certainty for both producers and purchasers.

In short, it is misleading to compare companies investing in more environmentally effective procurement strategies like PPAs and/or utility tariffs with hourly matching, with companies using cheaper and less impactful options, like buying standalone RECs and matching energy bought and consumed only on an annual basis. The use of unbundled RECs can lead to exaggerated claims of clean energy, especially if they are purchased in one country to cover energy demand in another.

24/7 Carbon-Free Energy is like a strict version of the market-based approach, or we might call it a hybrid of market-based and location-based approaches. 24/7 Carbon-Free Energy (CFE) aims for every kilowatt-hour of electricity (or even smaller units) to be matched with carbon-free sources at all times, emphasising real-time, local clean energy use. This is an improvement over annual matching. 24/7 CFE aims to incentivise renewable energy expansion, based on core principles: (1) matching consumption with carbon-free generation hourly to align with actual use; (2) buying energy locally to close loopholes that allow exaggerated additionality; (3) defining clean technologies inclusively; (5) focusing on adding new carbon-free energy sources; and (6) specifically targeting the highest fossil fuel usage hours, for maximum impact on decarbonising electricity systems. Of the cloud giants, Google has been most enthusiastic about this approach. Google writes:

> [...] because of differences in the availability of renewable energy sources like solar and wind across the regions where we operate—and because of the variable supply of these resources—we still need to rely on carbon-emitting energy sources that power local grids. [...] That's why, in 2020, we set a goal to run on 24/7 carbon-free energy (CFE) on every grid where we operate by 2030, aiming to procure clean energy to meet our electricity needs, every hour of every day, within every grid where we operate. Achieving this will also increase the impact of our clean energy procurement on the decarbonization of the grids that serve us.

Google and Microsoft are both signatories of the UN-backed 24/7 Carbon-Free Energy Compact. In itself, this is not a major commitment: "Our main request is that signatories dedicate time and attend meetings organised to shape the conversation on 24/7 Carbon-Free Energy" (GoCarbonFree247.com). However, the 24/7 CFE approach hopes to more effectively stimulate investment in green energy, and to incentivise energy purchasers to think more carefully about how they 'shape' their energy demands across time and space. Advocates of 24/7 CFE predict that "through hourly matching of demand with clean electricity, electricity consumers can both completely negate their carbon emissions and contribute to broader system-wide decarbonisation," and that "24/7 CFE matching creates an early market for advanced energy technologies" (Riepin and Brown 2023). As Killian Daly of Energy Tag puts it: "The basic fact is you can be solar powered all night long with today's accounting."[24]

The granular energy purchase approach of 24/7 CFE is somewhat new, and **much more research and scrutiny is required**. A few criticisms have emerged, although many of these are closely tied to Amazon's efforts to reshape the Greenhouse Gas Protocol. Is 24/7 CFE, as we would hope, an ambitious, inclusive and scientifically robust approach to a key aspect of energy transition, one that will work effectively with policy to correct serious greenwashing issues? Or is it a voluntary scheme from self-styled climate leaders with dubious track records and significant conflicts of interest, which will delay or

---

[24] Quoted in Kenza Bryan, Camilla Hodgson, Jana Tauschinski, 'Big Tech's bid to rewrite the rules on net zero,' *The Financial Times*, 14 August 2024.

undermine decisive and timely policy solutions?

The GHG Protocol is currently reviewing its Scope 2 guidance, which is the global gold standard for how companies report on the carbon impact of the energy they buy. There are some promising signs, including efforts to address problems outlined in this section (unbundled RECs, annual matching, lack of additionality, etc). Google has vocally advocated for more granular geographic and temporal accounting (Google 2023). However, Amazon (along with Meta and others) has championed a different approach, rooted in **project-based accounting**, also known as consequential accounting, and the notion of **emissionality**. Sloane (2021) has a nice clear explainer on the difference between the two approaches, complete with highly niche memes. As the *Corporate Climate Responsibility Monitor 2024* explains:

> The proposal to introduce a new accounting method – labelled 'project-based accounting' – would allow companies to claim reductions in their scope 2 emissions based on emissions avoided from renewable energy projects implemented anywhere in the world, whether inside or outside of the local grid region or market. This proposal appears closely aligned with the Emissions First Partnership, initiated by Amazon and with a small group of corporate signatories that includes Meta, Intel, and General Motors. In practice, this would effectively be the same as offsetting with carbon credits, which is a highly contentious proposal for improving the Scope 2 Guidance.

All this is slightly different in the case of data centres built with their own independent power sources, that don't draw (or draw much less) from the grid. However, the issues don't simply go away—even if it is completely independent of the grid, a data centre is still part of the planetary energy system. The energy that is being generated to power the data centre could be generated in a different way, and/or be used to power something else. There is interest in more grid-independent data centres, partly driven by sustainability concerns, and partly grid capacity as such, even in highly developed parts of the world (see e.g. Lee 2024).

Finally, using green energy to deliver services, and stimulating demand for green energy investment, are also only part of the story. As a rule of thumb, of course they are good things. But how available will that green energy be to **socially useful purposes**? How responsive will it be to future changing values about what constitute socially useful purposes? What are the opportunity costs of each green energy project? These questions also need urgently to be addressed in policy and practice.

The Cloud Giants and the Greenhouse Gas Protocol

The GHG Protocol includes a set of rules about how you account for the carbon impacts of the energy you use. The GHG Protocol feeds into other major initiatives like SBTi, as well as climate legislation. It is currently being reviewed and, as recently reported in *The Financial Times*, big tech is "**working behind the scenes to shape a once-in-a-decade rewrite of the rules governing how pollution from power use is disclosed**" (Bryan et al. 2024).

The cloud giants are clashing. Google is advocating that the GHG Protocol mandate a **24/7 hourly matching approach**, as laid out in Google's 2023 submission to the World Resources Institute. 24/7 hourly matching is a step towards 24/7 Carbon Free Energy. Microsoft appears inclined to support this approach as well. The approach would seek to eliminate problems with Renewable Energy Certificates (RECs). Currently, RECs allow clean energy produced at one time and location to offset carbon-intensive energy consumption elsewhere, often for a minimal cost. However, 24-7 hourly matching still requires much more scrutiny and debate. See the sidebar on 24-7 hourly matching for more details.

Meanwhile, Amazon (along with Meta and a few others) opposes 24/7 hourly matching. We have not been able to review their submission to the GHG Protocol consultation, but there is enough public information to indicate the general approach. It is based on **consequential accounting,** also known as **project-based accounting**. This means Amazon wants to double down on the REC markets, refining them to make them even more similar to voluntary carbon credit markets. Companies would subtract 'avoided' emissions from their carbon footprints (Bryan et al. 2024). The principle is to prioritise the total impact of each specific intervention, in an effort to allocate finance more efficiently, rather than worry about matching up consumption and production.

Efficient investment in renewable energy is a worthy goal. However, this approach tends to mean a lot more judgments about counterfactuals—"What *would* have been the case, *if* we hadn't taken this action?"—which creates significant greenwashing opportunities. This is perhaps particularly the case where there are major power imbalances. Increasing the complexity of these markets may further insulate them from public scrutiny. Furthermore, such a system would place very little constraint on decisions about where to build new data centres.

We should also be clear about the circuits of finance involved here, and consider questions of **green colonialism**. To simplify slightly for clarity: investors principally in the Global North would expand their ESG portfolios with renewable energy projects in the Global South. These projects would have been de-risked and made profitable by a top-up of finance from Global North tech companies (the ones buying their RECs, or equivalent). The Global North investors then reap the financial benefit from the Global North tech companies. The Global North tech companies benefit by becoming compliant with net zero legislation, while continuing to emit greenhouse gases and heat up the climate. The Global South may benefit somewhat from the expansion of clean energy infrastructure, of course. But whether this is an arrangement which does justice to the Global North's historic responsibility for climate change, and the Global South's relative vulnerability to climate impacts, demands careful consideration.

Although we can't offer a definitive assessment here of the cloud giants' proposals, the signs are not promising. If we were to put this provocatively, we might even ask if the disagreement between Google and Amazon is a matter of "good cop / bad cop" — not deliberately staged, of course, but nonetheless a kind of false dichotomy, one which distracts society from developing truly effective policy solutions.

Developing **accurate carbon reporting** standards means bearing in mind their **purposes**. Accurate carbon reporting is important for customers and investors to make informed decisions, and perhaps also to allocate climate finance efficiently and justly. But perhaps even more importantly, we need more accurate carbon accounting **to enable more effective climate policy-making**. This includes, for example, incentivising companies to get to net zero through **cap-and-trade schemes**, **carbon taxes,** or other instruments. Perhaps the most promising features of 24-7 hourly matching and the consequential approach can be combined and expanded (informed by sociologically and historically rich accounts of how comparable markets have been found to function in the past, not just how they are intended to function) to support such policy-making.

Just for example, Shafik Hebous and Nate Vernon-Lin, writing for the IMF, point to **reducing or removing data centre tax breaks** as an easy starting point for reining in carbon pollution:

> many data centers and crypto miners enjoy generous tax exemptions and incentives on income, consumption, and property. Considering the environmental damage, the lack of significant employment, and pressures on the electrical grid (possibly raising prices for households and reducing demand for the use of other low emissions goods, such as electric vehicles), the net benefits of these special tax regimes are unclear at best. (Hebous and Vernon-Lin, 2024)

Bold, responsible policy is necessary to bring the tech sector into line with climate goals, but **tech cannot be understood as straightforwardly subordinate to state power**. The cloud giants are all in the top ten biggest companies in the world by market capitalization. These powerful global entities shape research agendas, and influence policymakers' beliefs, attitudes, and regulatory approaches. See also 'Relevant UK policy on sustainable AI' later in this report.

Cooling Data Centres

Data centre cooling is an evolving field. See for example Masanet et al. (2013), McFarlane (2024), Van Geet and Sickinger (2024), EPRI (2024). Cooling often takes up a significant proportion of data centre energy use.

Free cooling, or natural cooling, uses ambient outside air. It can be an energy-efficient alternative to traditional compressor-based systems. Instead of continuously recirculating and cooling the same air, free cooling releases hot air outside and brings in cooler outdoor air. Free cooling is obviously very dependent on location and time of year.

Hot and cold aisle containment is a widely adopted strategy. By physically separating the hot and cold air streams, this method reduces the mixing of hot and cold air, ensuring that servers receive cooler air at a more consistent temperature. This approach not only reduces the overall cooling load but also allows for more precise temperature control. More broadly, **effective air management** in data centres involves all aspects of design and configuration aimed at minimising the mixing of cooling air supplied to equipment and the hot air expelled from it. For instance, careful and continuous cable management can play a role. Modelling tools and thermal imaging can support effective air management.

Raising the allowable IT temperatures in data centres is another approach to improving energy efficiency. Historically, data centres were kept at relatively low temperatures to ensure reliable operation of IT equipment. However, modern servers are designed to operate at higher temperatures.

By adjusting the airflow based on real-time cooling needs, data centres can **avoid overcooling**, which not only conserves energy but also extends the lifespan of equipment by reducing thermal stress. Data centres also sometimes **over-control humidity**, which reduces energy efficiency.

Data centres can also be run at **partial load**. By spreading the workload across multiple servers rather than running a few at full capacity, the heat generated can be more evenly distributed, making it easier to manage and cool. Data centre cooling systems are typically designed to handle the highest possible loads. However, these peak conditions are rare, meaning that the cooling systems often operate well below their maximum efficiency. This does also have implications for **utilisation** and energy proportionality.

Growing interest in heat reuse (see below) is also linked to the **shift from air cooling to liquid cooling**, something which would probably be happening even without the need to become environmentally sustainable. Liquid cooling can operate at higher temperatures, making it more suitable for heat reuse. In general heat reuse is an area characterised by a great variety of approaches and a lot of innovation and experimentation.

Emerging approaches include air-assisted liquid cooling, immersion cooling, microconvective liquid cooling, radiative cooling, and two-phase liquid immersion cooling (EPRI 2024).

Data Centre Heat Reuse

Data centres generate a lot of heat from their operations, and **reusing this waste heat** could help reduce carbon emissions. Heat reuse advocates have even been known to describe data centres as giant electric heaters, that just happen to provide computing as a handy side-effect!

Heat reuse techniques include supplying heating needs on site, district heating supply (Huang et al. 2020), using heat to help to power refrigeration in absorption or adsorption cooling cycles (Gupta and Puri 2021), electricity production using Seebek and piezoelectric effects, contributions to biomass fuel production, and desalination (Yuan et al. 2023).

Heat reuse is relatively well-established in some markets such as Finland. For example, the town of Mäntsälä received 70% of its district heating from a Yandex datacentre in 2022 (YLE 2022). In the UK, Deep Green is working on a model where multiple mini-data centres are located in sites that require hot water (like heated swimming pools). CEO Mark Bjonsgaard says, "If we were sitting here thirty years ago designing what the data centre industry would be we would say, well, we'll just break it all up and put it where the heat's needed and we wouldn't even debate it" (The Engineer, 2023). **Environmentally Opportunistic Computing (EOC)** is a model which "envisions a datacenter as a series of distributed heat suppliers for other buildings, distributing computational loads across a number of distinct nodes based on where heat is needed" (Sovacool et al. 2022).

David Kohnstamm, Chief Sustainability Officer and Founder at Leafcloud, points out, "Heat reuse is the core of the solution, since it's the only way we can actually increase IT power consumption without increasing the overall power needs of society. Putting servers in a field and throwing away the heat using air conditioning does not suddenly become green simply because it's been powered by renewable energy from Norway."[25]

As usual though, things are complicated, especially at the systemic level. As with many proposed climate technologies, many heat reuse opportunities are proposals, pilots, or currently relatively niche, with more research needed into capacity to scale. What are the challenges and complexities involved? Some have to do with **location**. The proximity and availability of infrastructure, like pipes, fluid pumps, and heat exchangers, are crucial for effective heat transfer. Data centres in urban areas can connect to existing heat networks more easily than larger ones in outlying regions. But if a data centre is located in an urban area, then to some extent it is competing with the residents' other energy needs for clean energy generation. It might be argued that it is best to situate data centres, and their energy sources, in areas that otherwise wouldn't be generating and using green energy at all. This puts us in the necessarily vague and contestable territory of counterfactuals and opportunity costs — what would be there if the data centre wasn't there?

Risks and uncertainties are intrinsic to scaling up technologies and adapting them for new contexts, and these need to be defined and where possible quantified. Heat reuse which improves the green

---

[25] Private correspondence, LinkedIn.

credentials of a data centre does not necessarily translate to improvements at the systemic level. Future heat recapture innovations, peeking over the horizon, must not be allowed to distract from decisive policy today, in order to align the impacts of the cloud with the needs of the climate.

Ljungdahl et al. (2022) offer a decision support model for designing heat reuse systems. The main inputs are load profile of district heating need, load profile of IT load, local district heating load, minimum district heating temperature, ambient air temperature profile in local climate, data on available/expected cooling system (heat transfer area, specific fan power, etc.), local electricity prices, local DH price, waste heat tax. The main outputs include yearly energy savings, yearly cost savings and efficiency gains through the Power Usage Efficiency and the Energy Reuse Efficiency. This is a good example of an effort to think more holistically about data centre heat reuse. However, currently, there is a significant lack of comprehensive frameworks for assessing the sustainability impact of heat reuse projects, considering social and economic dimensions.[26] To make sure heat reuse genuinely helps with net-zero goals, there needs to be **interdisciplinary scrutiny**, and a realistic insistence on the timescales and available carbon budget. We have not conducted a comprehensive literature review, but it does appear that most research to date focuses on technical engineering problems, with little attention to the economics and political economy of heat reuse, and very little attention to the wider social and cultural issues explored within critical data centre studies.

District heating is the most researched heat reuse application, but the **low thermal quality of data centre waste heat** often requires upgrading with heat pumps. In areas of the UK with already constrained grid access, adding heat pumps could strain the grid further, challenging project viability. All this relates to questions of **opportunity cost.** Cloud providers need better frameworks to identify and evaluate trade-offs between enhancing overall energy efficiency and heat reuse projects. For energy conservation, it is crucial to ensure that investing in heat export is both economically and environmentally more viable than traditional methods.

The potential failure of external heat network infrastructure might disrupt cooling systems, causing service interruptions and revenue loss. A recent white paper based on industry consultation also revealed concerns about **potential hidden costs**, such as being billed for cooling energy when contributing to a heat network (techUK 2024). Another consideration is the time it takes for a new data centre to reach projected workloads, and the **limited heat production in the early years**. The **ownership and operation** of heat network infrastructure may also pose challenges, particularly concerning security and access agreements with clients (techUK 2024).

---

[26] For comparison, Duboc et al. (2019) suggest that software systems should be designed to maintain the sustainability of the wider socio-technical system to which they belong, and offer the Sustainability Awareness Framework as a way of reflecting on and visualising this, at a high level of abstraction: with economic, environmental, social, technical, and individual dimensions, each subdivided into three orders of effect: immediate, enabling, or structural.

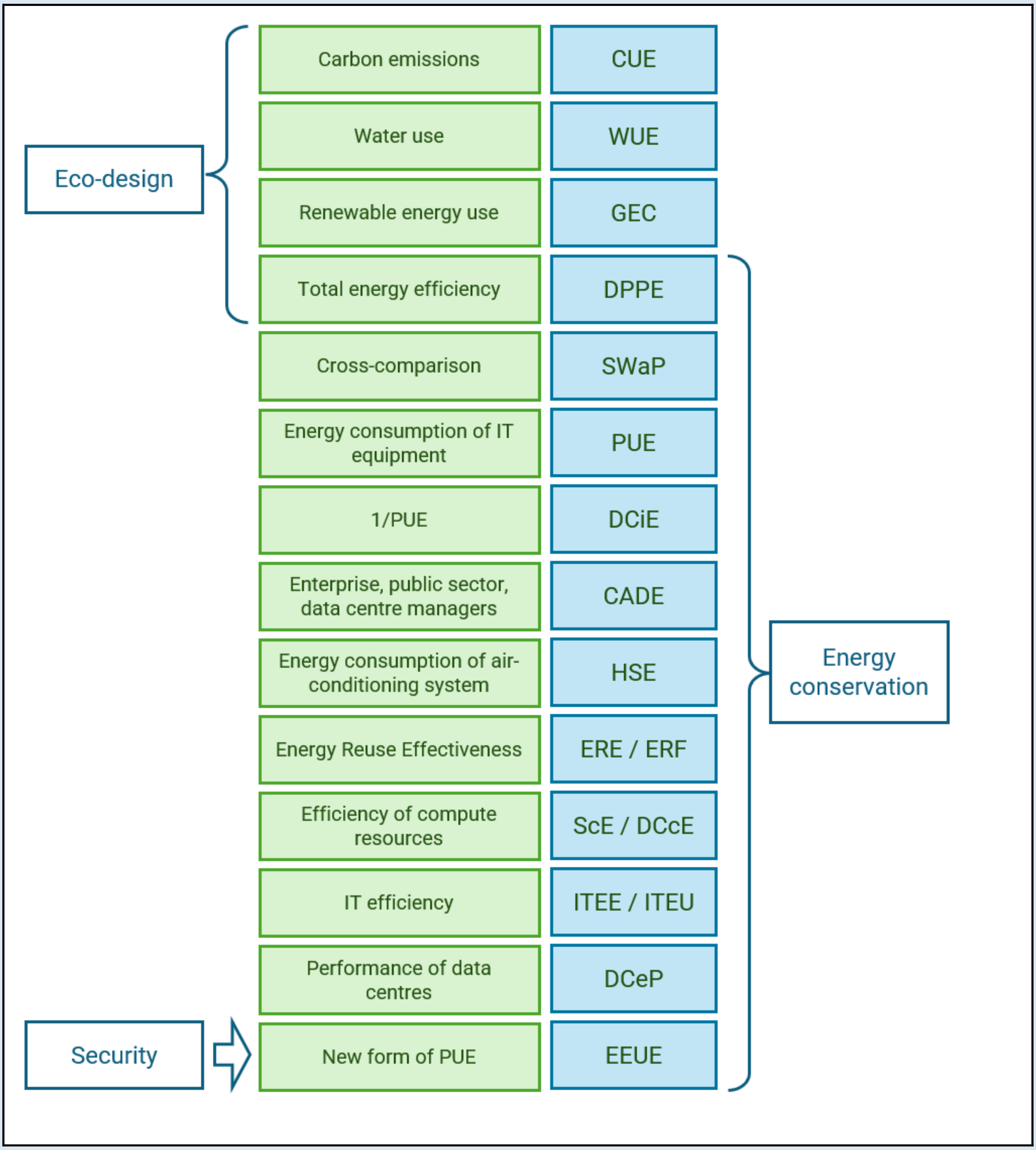
In brief, while data centre heat reuse is established in some contexts, and a promising frontier in others, **we should be cautious of overhyping its potential**. Heat reuse can be a relatively tantalising story, easy to grasp in its essentials, but filled with knotty details. There do not appear to be mature analytic frameworks and tools to assess feasibility in detail, to quantify sustainability costs and benefits, nor to integrate heat reuse into the broader sustainability picture of data centres. Even though data centre heat reuse is far from new, the evidence base to date also appears to be insufficiently interdisciplinary, with social and economic dimensions, as well as potential for unintended adverse impacts, not yet well understood.

Flooding in Sirajganj

Green Data Centre Metrics and Certifications

There are various metrics to assess how green a data centre is. The chart below is adapted from Shao et al. (2022) (which is also a good introduction to data centre sustainability that goes into slightly more technical detail than this report).

Useful Metrics and Certifications



Source: Adapted from Shao et al. (2022)

So some of the more important metrics, from a sustainability perspective, are Carbon Usage Effectiveness, Water Usage Effectiveness, and Renewable Energy Use, as well as DPPE, ERE and ERF. Datacenter Energy Sustainability Score (DESS) has recently been proposed as a complementary metric.

There are also some standards and certifications out there. Greenqloud has proposed Green PUE. BREEAM and LEED relate to the energy efficiency of buildings. The Green Climate Initiative's CGCF Assessment Framework is an eco-label combining a variety of efficiency, power chain, cooling, and air flow metrics, to produce a consolidated rating (with PUE thresholds as an additional requirement). CSN EN 50600-4-3 is the European Standard for defining and calculating the Renewable Energy Factor of a data centre. There is a proposal for an Energy Star for AI eco-label.

None of these metrics look deeper at the social impacts (although the authors of the DESS proposal explore this limitation in commendable detail). There is scope to explore how established methodologies in sustainability reporting, e.g. those involving stakeholder mapping and engagement, might be transferred to the evaluation of data centre sustainability reporting.

While confidentiality and privacy would make it challenging, there is in principle also scope to assess the sustainability of a data centre on the basis of the types of data and operations that are running there. If data were available about the mix of coding languages used by different customers, this could be factored in. More ambitiously, weightings could be assigned by sector. **If a heavily polluting company is the main user of a green data centre, using the data centre to advertise its heavily polluting services, is that really still a *green* data centre?**

The Global Electronics Council created a purchaser guide for sustainable cloud procurement. It is now a few years old, but still contains plenty of useful information.

Planting in the Kubuqi desert

Carbon offsets

If you want to learn about carbon offsetting, Carbon Market Watch and SourceMaterial are good places to begin.

When you are faced with a dashboard of **carbon offsets**, you tend to take their claims at face value. Usually, it's not the buyer's job to check if these offsets really deliver the benefits they say they do – that's supposed to be handled by the offset providers and regulators. In reality, offset programs may often overstate their impact on reducing emissions. They may include projects that would have happened without the extra funding from offsets (the “additionality” problem). They also tend to let projects use methods that embellish their actual impact.

The carbon credits which get issued are usually verified to some standard, e.g. the Gold Standard, and so they may look credible to the purchaser. However, there are currently a great many carbon credit standards; Bano (2024) counted 32 such standards.

Corporate Climate Responsibility Monitor 2023 heavily criticised Google’s purchase of carbon credits which probably did not benefit the climate, but still allowed Google to make carbon neutral claims:

> “**Some projects are unlikely to meet even historical Kyoto-era definitions of additionality.** Most of the offset credits that **Google** has procured stem from projects in the United States that capture and utilise methane from landfill sites to avoid its release into the atmosphere. The installation of methane capture technology is mandated by local or national governments in several industrialised countries. Analysis for other countries where there is no policy mandate for the technologies shows that there is a high economic incentive to implement such projects without support, if the biogas can be used for electricity generation (Warnecke et al., 2017). Accordingly, the additionality of the offset credits from the initial investment on this type of project is contentious. Google notes in a footnote of its carbon offsetting whitepaper that the credibility of offsets from these projects is contended, but that the company prefers to support projects that utilise captured gas (Google, 2011). Making use of the gas is indeed environmentally and economically attractive, and therefore good practice, but it is also the reason in this case why the credit revenue from Google may not lead to any additional climate action”
> (*Corporate Climate Responsibility Monitor* 2023, p. 62).

In theory Voluntary Carbon Markets (VCMs) are meant to enable the flow of climate finance to the most **efficient climate change mitigation opportunities**, which are often located in the **Global South**. The reality has all too often been something else. Wealthy companies release carbon into the atmosphere, heating the climate. They pay other companies a fee to point to some forests, and their carbon emissions are magically cancelled. That is, they are legally entitled to subtract the carbon credits from their emissions figures. Counterfactual deforestation is intrinsically difficult to assess — would trees *really* have been logged without this specific payment? In a world without VCMs, might they not be protected by some other means? — and companies like Finite Carbon have been shown to use gerrymandering, bundling inaccessible patches of forest ignored by loggers, and using them as the basis for issuing themselves carbon credits to sell. Such practices appear to be legal and difficult to legislate against. The pointing-at-forests industry is itself lucrative and expanding (see e.g. SourceMaterial 2023, 2024). VCMs are off course prone like any financial markets to volatility and instability.

In May 2024, in an effort to restore credibility to the troubled VCMs, the US government issued a statement of VCM principles. Of course, various laws, policies, principles, and standards are already in place. To an extent, it is impossible to imagine the VCMs without them. This is how markets are created in asset classes which did not previously exist. The **Integrity Council for the Voluntary Carbon Markets** (ICVCM) is establishing through its **Core Carbon Principles** a benchmark to enable carbon credit buyers and sellers to claim high integrity. The CDP (formerly Carbon Disclosure Project) provides a broad environmental disclosure platform, where businesses report and track environmental impacts, including carbon credit purchases. In addition to VCMs there are also what are effectively the mandatory carbon markets of cap-and-trade schemes. Under such a scheme, companies can sell the unused portion of their carbon emissions rights; such markets are also sometimes integrated with VCMs.

The Voluntary Carbon Market Integrity initiative (VCMI) was established in 2021 in preparation for COP26, to collaborate with various stakeholders, including civil society, the private sector, Indigenous peoples, local communities, and governments, seeking to improve the voluntary carbon markets. The VCMI Claims Code of Practice sets guidelines for companies using carbon credits in credible, science-based net-zero strategies. On the supply-side, the VCM Access Strategy Toolkit advises countries on how to foster high-integrity VCMs, in order to advance national climate goals and economic prosperity. Carbon credit companies can get their credits assured by a third party in alignment with VCMI standards and promote their products with the rating they achieve. Slightly ominously, the available levels are Silver, Gold and Platinum. In November 2023, VCMI launched a Scope 3 flexibility claim in beta version (to be refined during 2024), intended to incentivise companies to return to progressive Scope 3 emissions reductions targets by 2030 (VCMI, 2023). While Scope 3 emissions constitute a major part in overall corporate GHG emissions (on average over 70%), reducing them is considered a particular challenge for many companies (Science Based Target Initiative, 2023). However, the VCMI Scope 3 flexibility claim has also been viewed with some concern (Corporate Climate Responsibility Monitor, 2024), as set out in the Scope 3 sidebar.

Note that the purchase of carbon credits for offsetting is different to the purchase of unbundled Renewable Energy Certificates or similar instruments (see 'Green Data Centres'), although both practices share some of the same logic and some of the same problems.

The University of California (UC) got a chance to dig deeper into offsets thanks to funding from the Carbon Neutrality Initiative (CNI). They decided to stop using them altogether.

> With a $47 billion annual budget, almost 25,000 faculty members, ten campuses and the Lawrence Berkeley National Laboratory, and a public-interest mandate, the UC system decided in 2018 to invest time and resources into developing a quality offset procurement strategy. That decision created a three-year, all-campus, cross-discipline effort that generated an extensive analysis of existing offsets, methods for doing that analysis, and a bold approach to building offsets.

The UC project ran from 2018 to 2021. The goal was twofold: first, to address growing concerns about whether carbon offsets were truly effective and trustworthy, and second, to make sure that any carbon offsets they chose were in line with the UC's mission. The project set basic requirements for choosing offsets: they must not likely overestimate their impact, should not create social risks, especially vulnerable communities, and should use scalable technologies aligned with the goal of achieving global net zero emissions by mid-century, as recommended in the IPCC report on 1.5C global warming. Priority was also given to projects that align with the University's goals, like promoting research, involving students, and providing health and social justice benefits. Projects ideally should also help the UC and local communities, and have climate benefits beyond just the credited reductions. Researchers developed a methodology to check the quality of credits from each project, understanding that some methods might both overestimate and underestimate impacts.

Would the emissions reductions have happened without the offset program or the University's policy? What would have likely happened without the offset program or the University's policy? What about leakage: might the projects cause emissions outside their scope? Are any potential increases in emissions properly managed? Are the methods for estimating reductions accurate and based on the latest science? Is there a risk that stored carbon might be released again, and is this risk properly managed and accounted for? Additional criteria were also assessed for each project.

> As of July 2023, the University system replaced its 2025 carbon neutrality goal with goals for direct decarbonization of campus greenhouse gas emissions (see UC's press release describing its new goals, and section III.C of UC's Policy on Sustainable Practices for the formal policy). [...] Under its current policy goals, the University would no longer rely on voluntary carbon offsets to reach its carbon reduction targets.

See more: Offset Program Development for the University of California.

Planting in Montana

24/7 hourly matching

The **24/7 carbon free energy** (24/7 CFE) concept is somewhat new. It probably deserves a cautious welcome. 24/7 CFE aims to address greenwashing in energy procurement and market-based clean energy claims. Under existing approaches, it is possible to legally claim that a facility is powered by 100% clean energy, while releasing emissions into the atmosphere and heating the climate. EnergyTag, who are developing granular energy procurement standards and services, describe the issue:

> It's a cruel irony that the more successful we are in deploying renewables, the harder they are to integrate it into the grid. Current methods for procuring clean energy match supply and demand over a year and large geographical boundaries, resulting in a disconnect between carbon accounting and grid realities and failing to address the hardest hours of decarbonization. Even when claiming to be "100% renewable" an organization may still be reliant on fossil fuels for a significant portion of the day, enabling substantial underreporting of real-world emissions, as demonstrated by leading research in Nature. The lack of data about the time of electricity production on Energy Attribute Certificates is a key root cause of this issue. Given that corporates purchase over 50% of the world's electricity, **fixing these carbon accounting issues is crucial for achieving deep grid decarbonization.**

What is not to like? Something clearly needs to be done. But the question is not just whether granular accounting and 24/7 CFE are an **improvement** over existing clean energy procurement and accounting, but whether it is a **credible response** to the climate crisis, within the **timeframe** available. Might this be an improvement that is not enough of an improvement? Might it be taking up space that really requires more ambitious action? Here are a few brief observations about 24/7 hourly matching, to give a sense of areas where greater scrutiny and debate is required.

Above all, there is **not a lot of independent research in this area,** proportional to how high the stakes are. Some criticisms of 24/7 hourly matching have emerged from Amazon, Meta and other tech companies, including a recent Amazon-commissioned report advocating for globalising the clean energy markets (Turner et al. 2024; see also Bryan et al. 2024). The report argues for placing greater emphasis on avoided emissions, and proposes that energy consumption in one part of the world should be offsetable by investment in other parts of the world where the impact is most efficient. Such criticisms can be valuable and should be considered carefully. However, Amazon and Meta are likely to share many of Google and Microsoft's assumptions, which deserve to be examined from perspectives outside of big tech. Furthermore, Amazon has a poor track record on providing transparent, comprehensive, and actionable information on its carbon emissions. It is possible that its criticisms are shaped by wanting to undermine initiatives that would fundamentally challenge Amazon's reporting practices. In line with this, the *2024 Corporate*

Climate Responsibility Monitor report is critical of the alternative approach favoured by Amazon.

Overall, there is a somewhat urgent need for **much more interdisciplinary, peer-reviewed research, into 24-7 hourly matching**, independent of the cloud giants and big tech. This should include more research into the alternatives proposed by Amazon, Microsoft and others.

Another issue might simply be that **24/7 CFE is a voluntary scheme**. Well, you have to start somewhere, right? Environmental policymaking is sometimes described on a "conveyor belt" model: ambitious companies show leadership by developing best practice, which is then mainstreamed through standards and certifications, and finally mandated through policy. For example, Google is currently voluntarily pursuing 24/7 matching, and supplementing its sustainability reports with some detail on its Carbon Free Energy metrics and methodologies. Google is also advocating for 24/7 hourly matching, or something like it, to be included in the next revision of the GHG Protocol, which would count as some progress along the conveyor belt. If this happens, there could then be future phases which "give more teeth" to 24/7 CFE. RE100 indicates it "does not currently plan to introduce tighter than annual temporal matching or market boundary requirements tighter than those in the 2022 technical criteria" but notes that" granular matching is a topic that has received significant attention in updates to the GHG Protocol, which may influence the technical criteria in the future" (FAQs 2024).

But voluntary schemes can **either be a step toward better regulation, or a way of deferring it.** They can also be a mix of both. The conveyor belt model places a lot of onus on industry to create the 'prototypes' or first drafts of the policy which will eventually become mandatory. After all, they know their business best and are well-incentivised to spot unduly disruptive side-effects. At the same time, there is usually a certain degree of conflict of interests involved. So the conveyor belt model is not always appropriate—for example, when there are serious ongoing environmental harms which require swift action. It is also not appropriate when innovative regulatory approaches are not required, because there is 'low hanging fruit' that could be implemented following much more minimal stakeholder consultation, and/or when there is doubt as to the private sector's capacity to prototype effective regulation.

What other issues might there be? There is potential for **confusion between 24/7 hourly matching, the 24/7 Carbon Free Energy Compact, and actually achieving 24/7 carbon free energy.** Might we see companies proudly claiming 100% renewable energy and 24/7 hourly matching, with a low percentage of hourly matched carbon free energy buried in the small print?

The 24/7 carbon free energy goal also essentially creates a **new green metric on which companies can compete.** *We have 80% hourly matched carbon free energy; our competitors have only 70%*. There are advantages to this, of course: a company can seek to improve year on year. But there are disadvantages too. **Improving year on year is not in itself an**

adequate goal: what is required is emissions reduction consistent with limiting global warming to 1.5 degrees. Comparisons between companies may also give a false impression of transparency and precision. In themselves, they don't tell us which company is doing better: Company A may have 80% hourly matched carbon free energy, but extremely high emissions making up the last 20%, meaning they are polluting more than Company B with its 70% hourly matched carbon free energy. Of course, customers who are interested in sustainability are also likely to do their own more comprehensive analysis (rather than being misled by just one metric), but where time and capacity is limited, there may still be a risk. One solution might be some kind of aggregated metric, although as far as we have been able to determine, none has been suggested.

The 24/7 hourly matching approach also enacts a sort of unofficial **amnesty on previous misleading clean energy claims**. That is, framing 24/7 hourly matching as the "next step" has the additional consequence of softening criticisms of unjustifiable and deceptive (albeit widespread) energy accounting practices, as merely being immature. Should we continue to accept companies' claims that they have been powered by 100% renewable energy from such-and-such a date, when the first few years of this period used misleading accounting methodology? Will unbundled energy accounting approaches continue to be perceived as an acceptable "first step"? "Some buyers have turned to this approach as the next step to drive decarbonization after having achieved 100% renewable energy purchasing goals on an annual basis" (Hausman and Bird 2023).

Green hydrogen — somewhat mysteriously, green hydrogen is a prominent part of the 24/7 Carbon Free Energy Compact web presence (gocarbonfree247.com/). Green hydrogen is proposed as a transition fuel as we move to net zero, but is currently relatively inefficient.

Hydrogen fuel cells, specifically solid oxide fuel cells (SOFCs), promise to power data centres improving waste heat. Hydrogen fuel cells generate electricity by extracting electrons from hydrogen molecules, offering a low-emission power solution that holds promise for reducing the carbon footprint of data centres. Hydrogen fuel cells also have the potential to produce cooling as a byproduct of their operation. Compared to other energy sources, hydrogen has several advantages, which for some analysts imply its importance as a transition fuel. Hydrogen produces minimal emissions, unlike fossil fuels. Its availability is stable, unlike wind or solar energy. It can be used in any location, unlike geothermal. It doesn't pose the long-term environmental risks associated with nuclear power.

However, hydrogen comes with limitations and controversies. Producing hydrogen, particularly green hydrogen, is energy-intensive. Moreover, the infrastructure for hydrogen production, storage, and distribution has not scaled. These factors contribute to ongoing debates about the role of hydrogen in energy transition. Critics argue that green hydrogen is being used to perpetuate a catastrophic reliance on fossil fuels. Without getting into the details of the controversy here, we can say that the 24/7 Carbon Free Energy Compact might appear more credible if it created more distance from any one particular proposed clean energy solution. More broadly, there is always the risk of an **"omnibus" approach**, where strong corporate interests are able to demand concessions that are not strictly related to the improvements to corporate carbon accounting being developed.

Finally, it is worth exploring in greater detail whether hourly matching is **actually granular enough**, or whether within-hour variations might cause issues (the Taiwanese REC system uses a 15 minute granularity, for example).

Treating coati after wildfire, Brazil

Cloud Computing

What Is Cloud Computing?

"You could have an entire, I don't know, Cloud Anonymous. You could have meetings. My name is ... I have been three weeks since my last bad cloud decision."

— Bill Roth, Cloud Economist

Sociotechnical imaginaries change over time. Over the past twenty years or so, one paradigm shift in IT has been to do with where computational processes take place, and who owns the machines that perform these processes. This shift has been about increasing efficiency, productivity, and resilience. But it has also been about *defining* efficiency, productivity, and resilience in particular ways. These definitions are not inevitable. In fact, they come with baggage that may jeopardise a rapid and just transition to net zero.

We're talking of course about **cloud computing**. Instead of owning and maintaining all their own physical hardware and software, organisations and individuals can now rent or lease them from a Cloud Service Provider (CSP). A Cloud Service Provider is a company that offers a range of cloud computing services including compute resource, data storage and networking capabilities. As of 2024, the largest global players in this area are Azure, AWS, and GCP, the 'cloud giants'.

Migrating to the cloud essentially means that instead of running your IT processes on site ("on prem"), you are doing it in one or more remote **data centres**. A data centre is a location for the storage, management, and dissemination of data and information. Tech historian Nathan Ensmenger describes data centres as *factories* (Ensmenger 2021). It's a provocative characterisation, chosen deliberately to remind us of the physicality of data centres. A word like cloud *might* make some people imagine something ethereal, something weightless, something without much impact on the world. A hyperscale data centre might house hundreds of thousands of servers in one giant location. Obviously cloud migration doesn't necessarily mean doing *all* your IT processes within some distant data centre — there are all kinds of complex and supple configurations possible. Certain things can be done here, other things there.

For example, **Amazon Elastic Compute Cloud** (EC2) is a service offered by AWS. It allows you to rent virtual computers to run your applications. The word "**elastic**" in EC2 stands for its ability to easily scale up or down based on your needs. The idea is, you can quickly start up or shut down what are essentially virtual computers, or **Virtual Private Servers (VPS),** and only pay for the time you use them. You can choose the type and number of virtual computers, known as "**instances,**" you want to use. You can even select where these virtual computers are located around the world, which can help with latency issues and potentially sustainability too (see 'Carbon-Aware Computing and Grid-Aware Computing'). This is done through an Amazon Machine Image (AMI). An AMI can be configured with a wide range of software according to your requirements. You get the flexibility to create, launch, and shut down these server instances as per their needs, with billing based on the actual time the servers are active, billed by the second. There is also a **serverless** paradigm, which pushes the approach even further.

Back in the 1960s the DARPA-funded Project MAC, a system for sharing computer time, could be considered a precursor to cloud computing. The use of the cloud metaphor can be traced to General Magic in the 1990s. The first wave of modern cloud service providers really emerged in the mid-2000s, with Amazon Web Services launching cloud services in 2006. AWS offered a suite of services that included storage, compute, and other services. For a bit of historic context, CouchSurfing was founded in 2004, AirBnB and TaskRabbit in 2008, Uber in 2009 — the words "sharing economy" seemed to be on everybody's lips, although of course this was a kind of sharing you paid for. Shortly after, other major players like Microsoft and Google entered the market with Azure and GCP. As the cloud market has matured, other significant players like IBM and Oracle also entered the space, each with its own set of specialised services. And there are also plenty of small and mid-sized cloud service providers out there. However, the big three cloud giants have largely remained dominant.

Virtual Machines and Autoscaling

AWS, Azure, and GCP are branches of Amazon, Microsoft, and Google respectively, specialising in the delivery of on-demand cloud computing solutions and APIs to individuals and organisations. They provide a comprehensive range of cloud-related services, encompassing various aspects like networking, computing power, data storage, middleware, and Internet of Things (IoT) capabilities, among others. All these are accessible via the cloud giants' networks of hyperscale data centres. The idea is that this frees clients from the responsibilities of hardware and software management, including scaling and security patching. These services are billed according to usage, offering flexible, pay-as-you-go models. A key feature is **autoscaling**, which enables clients to automatically adjust computing resources based on application demand—scaling up during high-traffic periods and scaling down to save costs during low-traffic times. For example, a cornerstone offering of AWS is the Amazon Elastic Compute Cloud (EC2). This service gives users the ability to control a virtual cluster of computers through multiple interfaces such as REST APIs, a Command-Line Interface (CLI), or the AWS management console. These virtual machines are designed to mimic the functions of a physical computer down to the minutiae, including hardware features like Central Processing Units (CPUs) and Graphics Processing Units (GPUs), as well as software aspects like local and RAM memory, disk or SSD storage options, a range of operating systems, and pre-installed application software such as web servers and databases. Likewise, Azure and Google Cloud Platform's Virtual Machines provide on-demand, scalable computing resources that you can use to run workloads in the cloud.

Serverless Computing

The **serverless** paradigm goes even further in the direction of flexibility and scalability. In a serverless environment, you don't manage servers or instances directly; instead, you focus purely on writing code that runs in response to specific events. This approach abstracts away the infrastructure layer, in theory allowing for even greater elasticity and fine-grained control over resource consumption. AWS Lambda is an example of a serverless compute service, that charges you for the exact time your code is executing, rather requiring you to provision an entire VPS upfront to accommodate peak capacity.

As cloud computing matured in the 2000s and 2010s, the need for on-prem servers started to decline (in relative terms, at least). Different models have emerged: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). You can think of these as a spectrum, with IaaS giving the customer a lot of control, and SaaS giving the customer a lot of convenience (so long as they don't want to do anything weird or fancy).

- **IaaS:** Provides virtualised computing resources over the internet.
- **PaaS:** Provides a platform allowing customers to develop, run, and manage applications.
- **SaaS:** Software is provided over the internet, eliminating the need for installations or maintenance.

These models let clients rent computing resources or software applications from a cloud service provider. The advantage of this, you'll often hear, is that they don't need to maintain their own data centres. Of course, we could *also* imagine an alternative cloud, where most companies don't own and maintain their own data centres, but the cloud giants are not so giant, and clients can choose from a very wide range of smaller and medium sized cloud providers. A number of successful smaller players already exist: Alibaba, DigitalOcean, IBM Cloud, Krystal, Linode, Oracle, Salesforce, Scaleway, Tencent, UpCloud, to name just a few.

The point is: let's be cautious of the framing that the only options are cloud computing as it exists today, or every company having the hassle of maintaining its own on-prem data centre — this is really just marketing speak. There are better ways of specialising, sharing, and collaborating.

AI and Cloud Computing Glossary

See also the Cloud Governance Glossary.

AI (Artificial Intelligence): There is no good consensus definition of AI, but nowadays AI tends to refer to software developed using Deep Learning techniques. Contemporary AI is characterised by its reliance on large datasets, complex neural networks, and high computational power.

API (Application Programming Interface): A set of protocols and tools that allow different software applications to communicate with each other, often used for accessing web-based services.

Application Monitoring: Tools and processes used to oversee the performance and availability of software applications, helping to identify issues and improve user experience.

Architecture: The overall design and structure of some computer system, network, model, or application. It typically encompasses the hardware, software, protocols, and services that are used to build and operate a computing system, including how these components interact to support the system's objectives and requirements.

Bare-metal: The physical hardware of computers. Working with "bare metal" usually means you're dealing directly with the physical servers rather than virtualised environments or cloud-based solutions.

Carbon: Carbon dioxide is the main greenhouse gas, responsible for most of global warming. Carbon is constantly moving into the atmosphere and out of the atmosphere (into forests, oceans, soil, etc.). The problem is that this balance has been upset by anthropogenic (human-caused) emissions, especially from burning fossil fuels and developing land. In this report we are sometimes a little sloppy, saying “carbon” when we probably should say “CO2e” (carbon dioxide equivalent) or “greenhouse gases.” Methane is another very important greenhouse gas (see e.g. CarbonBrief 2024), and there are others too.

Carbon Intensity: A measure of how much carbon dioxide emissions are produced per unit of another variable, such as energy produced, economic output, or population. Carbon intensity could refer for example to how “clean” or “dirty” the energy grid is at a given time (according to how much renewable energy is powering it), or it could refer to the amount of carbon emitted per product manufactured.

Carbon Neutral: See *Carbon Neutral* in the other glossary.

Carbon-Aware Computing: To date, carbon aware computing has mostly comprised time shifting, location shifting, and demand shaping in an effort to perform workloads in the least carbon intensive way possible. Existing paradigms of carbon-aware computing have come under pressure, however. Is there a risk that an individual organisation can reduce its footprint in carbon accounting terms, while creating negligible or even counterproductive impact at the systemic level? One recent study concludes, "the potential for some significant carbon savings from spatiotemporal workload shifting, the benefits are often limited in practice" (Sukprasert et al. 2024). Grid-aware computing has been suggested as an alternative, more holistic paradigm.

CDN: See *Content Delivery Network*.

Chargeback: A FinOps (or GreenOps) term. Billing departments for their cloud service usage, fostering financial accountability and encouraging cost-effective behavior. See also *Showback*.

The Cloud: The cloud may mean slightly different things in different contexts. It generally refers to a model of computing where data and applications are stored and accessed over the internet rather than on an organisation's local physical hardware. In this sense, the cloud relies on remote servers hosted by third-party providers like AWS, Azure and GCP, who are responsible for ensuring the data is secure and available. See also *Public/private clouds*, and *Colocation*.

Cloud Giants: In this report, we use this term to refer to Amazon Web Services (AWS), Azure, and Google Cloud Platform, or to their parent companies, respectively Amazon, Microsoft, and Google / Alphabet.

Cloud Native: "Cloud native" refers to the paradigms, perspectives, and practices that typically go along with creating software with and for the cloud. A cloud native application is usually very modular, a bit like Lego. It's made up of multiple small pieces of software called microservices. According to the Cloud Native Computing Foundation (2023), "Cloud native technologies empower organisations to build and run scalable applications in modern, dynamic environments such as public, private, and hybrid clouds. Containers, service meshes, microservices, immutable infrastructure, and declarative APIs exemplify this approach." Eric Zie, in *Decarbonise Digital* (2023), points out that "not all workloads operating in the public cloud are cloud-native, and cloud-native workloads can be run outside of the cloud (the term 'cloud-native' is misleading in this regard)."[27] Critics also suggest that cloud native approaches can lead to vendor lock-in, which may make moving to a greener provider more challenging.

Cloudflare: Cloudflare is an example of a cloud service provider specializing in content delivery network (CDN) services, internet security, and DDoS (Distributed Denial of Service) mitigation, alongside DNS services and performance optimization tools for websites. Unlike the cloud giants, which offer a broad spectrum of services including computing power, storage, databases, and advanced machine learning and analytics, Cloudflare primarily focuses on improving web performance and security. Cloudflare has green

---

[27] Eric Zie, *Decarbonise the Digital: Facts. Methods. Action* (2023), p. 68.

pledges similar to those of Google and Microsoft, and including offsetting historic emissions. See also *Content Delivery Network*.

Clustering: Multiple servers (physical or virtual) being connected together so that they act as a single system, providing high availability and fault tolerance.

Colocation: A service where businesses rent physical space, power, and cooling within a data centre to house their own servers and networking equipment. The data centre is owned and managed by a third-party provider, but businesses maintain control over their own hardware and software. Colocation allows companies to benefit from advanced data centre infrastructure without having to manage the facilities themselves.

Compute Efficiency: Usually refers to how effectively a machine learning model uses computational resources, such as processing power and memory, to perform its tasks. Higher efficiency typically means faster training and inference with less resource consumption.

Compute: Refers to the processing power required to run applications on a computer or in a data centre.

Container: A container is essentially an application plus everything it needs to run, helping to eliminate the old problem of “but it works just fine on *my* computer!” See *Containerisation*.

Containerisation: A lightweight form of virtualisation that packages an application and its dependencies into a virtual 'container'. This enables lightweight and portable application deployment, thus improving efficiency and scalability. Traditional virtual machines include not just the application but also an entire guest operating system. This setup can be resource-intensive. One benefit of containerisation is to include just the bits you actually need.

Content Delivery Network (CDN): A widely distributed network of servers designed to optimise the delivery of data to users. By caching content closer to end-users, CDNs can enhance load times and improve overall user experience. CDNs can help websites stay up and function adequately during peak traffic times, and can also provide defence mechanisms against cyber threats such as Distributed Denial of Service (DDoS) attacks. They are widely used to deliver streaming content with minimal buffering, to facilitate software distribution by hosting and delivering software updates and patches, and for e-commerce to enhance the speed and responsiveness of online stores. CDNs also face challenges such as high setup and maintenance costs, the risk of local content going out of date, and complex security and configuration requirements. In terms of environmental sustainability, CDNs offer some interesting trade-offs and opportunities. They can reduce energy consumption and emissions by caching content closer to users, and decreasing long-haul data transmission. However, maintaining a CDN requires extensive infrastructure, including numerous global data centres and servers. The energy efficiency of these data centres varies, influenced by location, energy sources, and cooling technologies. Balancing the efficiency gains from optimised delivery with the environmental impact of maintaining this extensive infrastructure presents a complex challenge for CDN providers aiming to minimise their carbon footprint. The cloud giants operate CDNs: Amazon CloudFront, Azure CDN, Google Cloud CDN. See also *CloudFlare* for another example of a CDN provider.

Cross-Validation: In the context of AI, cross-validation is a technique for assessing how the results of a statistical analysis will generalise to an independent dataset, often used to prevent overfitting and ensure the model's performance is robust.

CUE: Carbon Usage Effectiveness. A metric used to assess the greenhouse gas (GHG) emissions produced for each unit of IT energy consumed in a data centre. This metric complements another well-known Green Grid metric, Power Usage Effectiveness (PUE), which gauges the energy efficiency of a data centre. The formula for is: CUE = Total CO2 emissions / Total IT Energy (kWh). To determine the CO2 emissions, different energy sources (like coal, gas, oil, biomass, nuclear, or renewables) have varying Emission Factors, which represent the CO2 emitted per unit of energy produced. These are typically available in public databases or provided by utility companies. Total IT energy refers to the energy used by the IT equipment within the data centre, such as servers, storage systems, and networking hardware. A lower CUE signifies a smaller carbon footprint, indicating greater carbon efficiency in the data centre. The minimum possible CUE value is 0, representing a data centre powered entirely by carbon-free energy sources. See also *PUE*.

Data Centre: A facility used to house computer systems and related components, such as telecommunications and storage systems, where data is stored, processed, and managed. See 'Types of Data Centres' section in this report.

DCIM: Data Centre Infrastructure Management (DCIM). Software such as Cormant-CS, EkkoSense, FNT Software, Nlyte, EcoStruxure IT, and Sunbird, used for managing and optimising data centres.

Demand Shaping: This term is not always used consistently. It can mean providing a more basic service during periods when the grid is green, e.g. inviting users to opt in to extra features or to pay a premium for full service during these times.

Disaster Recovery: Strategies and processes designed to restore and protect a business IT infrastructure in the event of a disaster.

Docker: Docker is a popular platform for creating and managing containers, offering an ecosystem for developing, shipping, and running containerised applications. Docker can package an application and its dependencies in a virtual 'container' that can run on most computers, which also enables it to run on the cloud. See *Containerisation*.

Edge: Edge computing involves deploying applications on devices or servers closer to end users or data sources, rather than relying solely on remote centralised data centres. While these applications can still communicate with remote data centres, the primary processing and data handling occur locally, reducing reliance on centralised infrastructure. This local processing minimises latency, enhances bandwidth efficiency, improves reliability and security, and supports scalability. Edge computing often employs a hybrid approach, where some data is processed locally, and other data is sent to remote data centres for further analysis or storage. This integration allows for tasks requiring more computational power or large-scale data storage to be efficiently handled by the cloud, while the edge infrastructure manages tasks closer to the source, optimising performance and resource utilisation. Edge computing is to some extent an existing reality, but is also an aspirational paradigm which is not fully matched by the infrastructure we have today. See also *Content Delivery Network*.

Elasticity: The ability of a cloud service or infrastructure to dynamically scale resources up or down based on the changing demands of an application or workload. It is one of the key characteristics of cloud computing and ideally should play a fundamental role in optimising resource utilisation and ensuring that applications can perform efficiently, reliably, and cost-effectively. Note that “elasticity” can also have another, unrelated meaning in economics, that may sometimes be relevant, e.g. in Jevons’ Paradox. Elasticity in this second sense usually measures how much the quantity demanded of something responds to changes in price. High elasticity means a significant change in quantity with a small change in price (maybe the customer can easily switch to and from alternatives), while low elasticity indicates that quantity is relatively unresponsive to such changes (maybe it’s something like insulin, that the customer has to somehow pay for at any price). By combining the concept of elasticity with the concept of carbon intensivity, we can gain a more nuanced understanding of how a change in an entity’s income or costs might influence that entity’s carbon footprint.

Emission Factor: Also sometimes called Conversion Factors. An Emission Factor is a coefficient that quantifies the emissions (or removals) of CO2e per unit of activity (e.g. miles driven). Used in carbon accounting: by multiplying how much you’ve done something by that something’s Emission Factor, you have estimated the emissions produced. Emissions Factors are available in various databases (e.g. maintained by UK government). Extra calculations would need to be done to reflect economies or diseconomies of scale (in other words, sometimes the amount of carbon associated with the 1,000th time you do an activity is different from the first time you do the same activity, and the most basic Emissions Factor methodologies won’t reflect this).

Energy Proportionality: A measure of how efficiently a system uses energy across different levels of utilisation. A perfectly energy proportional system would be equally efficient whatever the level of utilisation. This contrasts with most real-world systems, where energy per operation usually increases at lower workloads due to fixed energy overheads. Energy Proportionality is related to, but distinct from, Power Usage Effectiveness, which is about the overall ratio of energy used by IT equipment in a data centre as opposed to cooling and other overheads. See *Utilisation*, *PUE*.

Epoch: A single pass through the entire training dataset. Training a model usually involves multiple epochs to progressively improve its performance.

Feature Engineering: The process of using domain knowledge to extract features (variables) from raw data that make machine learning algorithms work more effectively.

Fine-Tuning: Fine-tuning is the process of taking a pre-trained model and adjusting its parameters on a specific, smaller dataset to optimise its performance for a particular task. For example, you could fine-tune GPT-4 on Shakespeare's plays to make it sound more Shakespearean.

FinOps: The practice of financial operations related to cloud computing, aimed at aligning cloud technology expenditures with business goals. In some respects, an outgrowth of the DevOps paradigm, which sought to integrate software development and deployment/management. In a somewhat similar fashion, FinOps sought to integrate technical decision-making about cloud usage with financial decision-making. See also *GreenOps*.

FLOPs (Floating Point Operations per Second): FLOPs measure the computational complexity of an algorithm, indicating the number of floating-point operations a model performs per second. It is often used to assess the efficiency and performance of deep learning models.

Foundation Model: A foundation model is a large AI model trained on vast amounts of data, designed to be versatile and adaptable for various downstream tasks. These models may comprise billions of parameters, which makes them computationally expensive to train and deploy. Due to their size and the resources required, they are costly to develop, with training often demanding substantial computational power. Foundation models are highly adaptable (e.g. through fine-tuning) for a wide range of specific tasks

GenAI (Generative AI): In AI, generative AI refers to models that can generate new, synthetic data similar to the input data they were trained on, such as text, images, or music.

GreenOps: Integrating sustainability into the management of IT, and often specifically cloud computing. Evolves from FinOps. Similar terms include DevSusOps, GreenDevOps, and cloud sustainability. See also *FinOps*, and the section on GreenOps.

Grid-Aware Computing: This has been suggested as an evolution of carbon-aware computing (which emphasises efficiency, time and location shifting, and demand shaping) to a more holistic and systemic understanding of computing practices consistent with a rapid and just transition to net zero.

Hallucination: In the context of AI, hallucination refers to a situation where a generative model, such as a language model, produces outputs that are nonsensical or factually incorrect, diverging significantly from the training data.

High-Performance Storage Systems: These are specialised data storage solutions designed to swiftly store and retrieve data. They also include mechanisms for data recovery to protect against data loss due to unforeseen disasters.

Hub: A simple network device that connects multiple computers in a network in a star configuration. It broadcasts data packets to all ports irrespective of the destination, making it less efficient than switches and routers.

Hyperparameter: In AI development, a hyperparameter is a parameter whose value is set before the learning process begins, controlling the behavior of the learning algorithm.

Hyperscale Data Centre (HDC or 'Hyperscaler'): A really big data centre, typically consisting of tens of thousands of servers or more. These are significantly larger facilities than a typical enterprise data centre, and in principle can quickly and efficiently scale in response to increased demand. However, there is a lack of transparency around their actual utilisation levels.

Image: A snapshot of a computing environment, often used to quickly deploy identical virtual machines or software setups. In containerisation, an image is a static file that includes everything needed to run an application: the application's code, libraries, runtime environment, and other configuration files. When you start a container, you're essentially creating a running instance of this image.

Inference: In AI, inference is basically using the model, for example, asking ChatGPT a question. It is the process of making "predictions" or decisions using a trained machine learning model on new, unseen data.

Kernel: The core part of an operating system, responsible for resource allocation, low-level hardware interfaces, and security. A traditional virtual machine would include a full copy of an operating system. A container, on the other hand, is a lighter form of virtualisation; it packages an application along with its dependencies, but it shares the host system's kernel, rather than including its own OS.

Kubernetes: An open-source platform for automating the deployment, scaling, and management of containerised applications. It enables efficient container orchestration, supporting elasticity and high availability in cloud-native applications. See also Amaral et al. (2023) and Currie (2024).

LAN (Local Area Network): A network that connects computers and devices in a specific geographic area, such as a home or office, allowing them to communicate and share resources like printers or internet access.

Latency: In AI development, latency can refer to the time it takes for a model to process input data and produce an output. Lower latency is crucial for real-time applications where quick responses are essential.

LLM (Large Language Model): A type of neural network model trained on vast amounts of text data to ‘understand’ and to generate human-like text.

Low-Latency Networks: These are network configurations specifically designed with enterprise-level components to minimise data transmission delays, commonly known as "latency."

Managed Service Providers (MSPs): Third-party companies that provide IT outsourcing solutions. MSPs remotely manage a customer's IT infrastructure and end-user systems. MSPs provide a range of services, including network management, security, data backup, and cloud services, against a defined Service Level Agreement. MSPs often offer their own cloud services or resell services from other cloud providers (including the cloud giants) while adding value through management, support, and customisation.

Model Compression: In the context of AI, model compression involves techniques to reduce the size of a model, such as pruning, quantization, and knowledge distillation, making it more efficient in terms of storage and computation.

Model: An AI. A model is a mathematical representation or algorithm trained on data to perform tasks without being explicitly programmed for the task.

Multicloud: Multicloud refers to the use of multiple cloud computing and storage services in a single heterogeneous architecture. This strategy involves deploying and managing an organisation's assets, software, applications, and more across several cloud environments, which can include a mix of public, private, and hybrid cloud services. The primary purpose of multicloud is to avoid dependency on any single cloud provider and to benefit from the best features and cost efficiencies of each selected cloud service. This approach also enhances resilience and flexibility, as it allows businesses to tailor their cloud usage to specific needs and avoid vendor lock-in.

Network Infrastructure: Physical and virtual components that are used to build, manage, and operate a network, including hardware and software.

Net Zero: See *Net Zero* in the other glossary.

Neural Network: In AI, a neural network is a series of algorithms that attempt to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates.

On-prem (On-premises): Refers to the deployment of software, hardware, or services within an organisation's own physical location rather than in a third-party cloud or data centre. Companies are responsible for the maintenance, management, and security of on-premises solutions, giving them more direct oversight and control, but often at the cost of greater operational complexity and financial investment.

Orchestration: Automated management of the interactions between workloads on public and/or private cloud infrastructure. Orchestration is about how containers and/or virtual machines are deployed, scaled, and managed. Kubernetes and Docker Swarm are popular tools for container orchestration.

Overfitting: Overfitting occurs when a model learns the training data too well, capturing noise and details that do not generalise well to new, unseen data.

Packet: A unit of data sent over a network, encapsulating the data and information required for its transmission.

Parameter: In AI, a parameter is a variable in the model that is learned from the training data. These parameters define the model's predictions and are adjusted during the training process.

Predictions: AI developers often refer to the outputs of AI models as predictions. This doesn't mean that the model is trying to predict the future (although sometimes it might); it's more like the input you've given the model (the prompt you've fed to the AI) is interpreted as a fragment, and the model tries to "predict" what the missing bits of the fragment are (the answer or other output).

Private / Public Cloud: Standard commercially provided cloud services are sometimes described as the public cloud. By contrast, a private cloud often means when cloud computing principles are used but the organisation has its own data centre located on its premises. It is not quite so simple, however, because a private cloud might also refer to a colocation type set-up, where a company has its own servers but located in a data centre managed by a third party. Or a private cloud may even refer to 'virtual' private cloud services offered by some cloud service providers, although perhaps this stretches the definition a bit far. Many organisations also use hybrid cloud and/or multicloud set-ups.

PUE (Power Usage Effectiveness): A ratio that describes how efficiently a data centre uses energy; specifically, how much energy is used by the computing equipment in contrast to cooling and other overhead. A very popular metric for assessing data centre sustainability, although there is a lot it doesn't tell you. Complementary metrics such as Carbon Usage Effectiveness and Water Usage Effectiveness also exist.

Reinforcement Learning: In the context of AI, reinforcement learning is a type of learning where an agent learns to make decisions by taking actions in an environment to maximise some notion of cumulative reward.

Router: A network device that directs data packets between different networks, typically connecting a local network to the internet.

Scopes 1-3: See *Scopes 1-3* in the other glossary.

Secure Infrastructures: These consist of systems and protocols that regulate access to information and ensure data availability. They provide robust defences against unauthorised access, breaches, and cyberattacks, thereby maintaining customer trust.

Security Systems: Hardware and software solutions designed to protect data centre resources from various threats.

Server: A high-performance computer that provides services, data, and resources to other computers over a network.

Serverless Computing: A cloud-computing model where the cloud provider automatically manages the infrastructure (including things like fault tolerance, elastic scaling of computing, storage, etc.), allowing developers to focus solely on building and deploying applications. If a Virtual Machine is akin to a rental car, then serverless computing is more like taking an Uber. Linthicum (2024) suggests, “The meaning of serverless computing became diluted over time. Originally coined to describe a model where developers could run code without provisioning or managing servers, it has since been applied to a wide range of services that do not fit its original definition.”

Showback: A FinOps (or GreenOps) term. Tracks and reports cloud service usage and costs to departments to promote awareness and responsible usage, without actual billing. See also *Chargeback*.

Spin up: Basically slang, meaning to boot up, start, create, e.g. spinning up a container might mean running an instance of a Docker image.

Static Power Draw: The amount of electricity consumed by a device when it's in idle state, not performing any active computations or tasks. See also *Energy Proportionality*.

Storage Systems: Devices like hard drives and cloud storage services where data is stored.

Storage: Refers to digital spaces, such as hard drives or cloud storage services, where data is stored.

Supervised Learning: A type of machine learning where the model is trained on labelled data, meaning each training example is paired with an output label.

Switch: A network device that filters and forwards data packets between different devices on a local area network (LAN), operating more efficiently than a hub by sending data only to specific devices rather than all ports on the network.

Training Data: In the context of AI, training data refers to the dataset used to train a machine learning model, enabling it to learn patterns. It can then make predictions based on partial data. As a simple example, imagine a dataset of pictures of apples and oranges, each of which has been labelled “apple” or “orange” by a human who is not easily fooled by fruit. This can be used to train a model, which can then output (“predict”) the word “apple” or “orange” when shown an image.

Underfitting: In AI, underfitting happens when a model is too simple to capture the underlying patterns in the data, resulting in poor performance on both the training data and unseen data.

Unsupervised Learning: In AI, “unsupervised” learning involves training models on data without human-labelled responses, aiming to find hidden patterns or intrinsic structures in the data. It is not entirely unsupervised.

Utilisation: The extent to which the computing resources (e.g. servers or processors) are actively being used compared to their total capacity. High utilisation indicates efficient use of resources, while low utilisation suggests underuse. There are complexities, e.g. a CPU with 50% utilisation doesn't imply an even load across half its cores. Depending on the workload, some cores may run at higher frequencies while others run lower or remain idle. See *Energy Proportionality*.

Vendor Lock-In: Vendor lock-in, also called lock-in, customer lock-in, or proprietary lock-in, occurs when a customer becomes integrated with a specific provider's services, such that it is difficult or costly to switch to another provider. Cloud-native applications are typically designed and optimised to run in specific cloud environments, taking full advantage of the unique features provided by a particular cloud provider, so a cloud-native approach can run the risk of lock-in.

Virtual Machine (VM): A software-based simulation of a physical computer, running an operating system and applications just like a physical computer.

Virtualisation: This is a technology that enables the creation of virtual instances of physical hardware or resources, such as servers or storage devices. Well-managed virtualisation can improve server deployment speeds, increase system uptime, enhance disaster recovery processes, and contribute to energy efficiency. See also *Containerisation*.

WANs (Wide Area Networks): These are expansive networks that can cover large geographic areas. WANs have built-in capabilities to manage network traffic by allocating varying levels of bandwidth to different applications based on their priority.

Workload: Basically a computer doing something. A workload just means some amount of processing that a computer system, network, and/or application has to perform.

Zero Downtime: This is an operational goal aiming to eliminate interruptions in business activities by ensuring continuous system availability. Achieving zero downtime reduces both operational costs and the risk of lost revenue.

GreenOps

The Cloud vs. On-Prem vs. Hybrid / Multicloud

GreenOps is a new paradigm that **joins up sustainability and IT operations**, bridging the gap between environmental responsibility and digital transformation. By integrating GreenOps practices, companies aim to optimise cloud usage, reduce carbon emissions, and implement sustainability throughout the entire lifecycle of IT services. "What we also want to do is bring carbon emissions data right next to the cost data. Putting them side by side," comments Mike Jaco of Mastercard (FinOps Foundation 2024). GreenOps includes energy efficiency and waste reduction as well as continuous monitoring, automation, and innovation to align cloud operations with broader sustainability goals. There are actually a few different terms floating around at the moment: GreenOps, DevGreenOps, Sustainable FinOps, etc. GreenOps is continuing to evolve—and in this report we try to map some of the remaining challenges.

There are many reasons why cloud migration may be attractive, including cost, scalability, security, among others (see "Selling Cloud Migration: Beyond PUE" and "Cloud Nuance" above). Among these benefits, sustainability may also be mentioned.

What makes the public cloud a potentially more sustainable option? Well, there are those 100% renewable energy claims (as we describe in another section, these are a little complicated). Cloud data centres can often be located near to energy-generating facilities. Compared to most on-prem data centres, hyperscale data centres also have potential to improve both PUE (how much energy actually goes to IT) and utilisation (whether servers are running at full capacity). There are potential economies of scale when it comes to constructing an efficiently powered and cooled warehouse full of high-performing and well-maintained servers. Fundamentally, the cloud giants have deep pockets. They can in theory take action on climate in ways that resemble coalitions, networks, or other large actors such as states.

But for the individual company choosing a cloud solution, it is important to remember that **these are potential benefits only**. GreenOps seeks to make sure that cloud sustainability is actually monitored and optimised. Cloud migration comes with sustainability risks as well as opportunities. Data centres represent a lot of embodied carbon, and consume significant amounts of energy and water. Crucially, if a cloud provider, and/or a particular data centre, appears to have great green credentials, we need to understand where these come from. If renewable energy is used, where and when is it generated—at the same time and place as the data centre using the energy, or someplace else at some other time? What about the embedded emissions in the materials used to build the data centre and associated infrastructure? Is an accounting methodology used that allows green claims by sleight of hand? (See "The Cloud Giants and the GHG Protocol").

The cloud giants offer sustainability-focused tools, but these don't yet give us the nuance we need. The data provided by tools such as AWS's Customer Carbon Footprint Tool is currently **market-based** rather than **location-based**, meaning that it permits Renewable Energy Certificates, and doesn't offer any analysis based on actual carbon intensity of the grid. Current best practice is reporting that combines both market-based and location-based metrics, as well as **24/7 hourly matching**.

Eric Zie (2023) writes:

> A cloud provider may point to a data centre linked to a wind or solar farm, but your cloud services, applications, and data may not be using that data centre. The data centre your company uses may be in an area, state, or country that uses a coal-fired power facility as its source of energy.[28]

There are bigger questions too. What might the renewable energy be used for if it were *not* used for the data centre? Investment in renewable energy is good, but is it being done in an equitable way? What would that land be used for if it wasn't being used for renewable energy? Is offsetting used — the purchase of **carbon credits**, supposed to represent carbon emissions removed from the atmosphere, or "avoided" somewhere in the world? — and if so, can we unpick the actual impacts, the risks and uncertainties, and the presuppositions of this offsetting?

And it must be remembered that carbon emissions are only a piece of this puzzle. **Water shortages** and energy crises are widespread in many countries around the world and both of these issues are exacerbated by a surge in data centre construction. James Hall, Head of GreenOps (see below) at Greenpixie — an organisation providing GreenOps practitioners with independent cloud sustainability data — points out that in West London, "data centres have been blamed for halting housing projects by sucking up all the available electricity." This real world impact of data centre's resource consumption can also be found in Virginia, USA where residents have long suffered the effects of the "Data Center Alley"'s enormous water consumption. An effective GreenOps mindset must consider these factors alongside carbon.

There are also bigger issues that are probably not yet well-reflected within GreenOps, although there is great potential for evolving the paradigm. For example, cloud computing is also, simply put, a rentier model, subject to standard criticisms around exacerbating economic inequalities. There might be objections from certain perspectives on the right (could there ever be robust competition in a market with such economies of scale?), as well as criticisms from the left (capital accumulation). The big three corporations that own the physical resources (and especially AWS and Azure) clearly dominate the market. Could there be more equitable ways of improving utilisation and benefiting from economies of scale?

ICT represents a significant fraction of global carbon emissions. Of course, **ICT's impacts are nowhere near as large as other activities** like producing food, or

---

[28] Eric Zie, *Decarbonise the Digital: Facts. Methods. Action* (2023), p. 73.

heating and lighting our homes and buildings, or industrial processes like cement manufacture. ICT also enables resource efficiencies in these other domains. But there are at least two big problems here. One is that there is **a lot of uncertainty** around how much ICT contributes to sustainability, and how much it displaces or locks out other ways of achieving sustainability (see AI for Climate and Sustainability).

The other problem is that **ICT is growing rapidly.** This growth partly represents **empowering communities whose digital connectivity was previously limited**. Mignamissi and Jijjo T. (2022) point out that "between 2005 and 2019, the proportion of individuals using the Internet increased from 16.8% to 53.6%" but that "developing regions in general and Africa in particular, remain highly marginalised in terms of digital penetration." Is getting online *always* straightforwardly good? It's not a bad rule of thumb, although Heels (2022) cautions against adverse digital inclusion, when "inclusion in a digital system that enables a more-advantaged group to extract disproportionate value from the work or resources of another, less-advantaged group."

More broadly, the growth in ICT also represents the endless energetic (and energy intensive) churn of **obsolescence, innovation, and creative destruction**. For example, GenAI is fun and exciting, but it also has complex social risks, even before you begin to consider the environmental costs. Our economies don't appear to be very good at evaluating this. There is very little about the characteristic socio-technical imaginaries of Amazon, Google and Microsoft that suggests this is going to change any time soon.

By understanding these criticisms and drawbacks, organisations can make more informed decisions and take proactive steps to mitigate these risks, and to identify opportunities to push the cloud giants to improve, and to inform more decisive, effective policy making.

The History GreenOps: DevOps and FinOps

The rise of **GreenOps** comes after the rise of **FinOps:** Financial Operations, the practice of bringing financial accountability to the variable spending model of cloud computing. **GreenOps** combines sustainability data with financial data. These approaches are generally used by larger companies, and by companies with significant digital operations (tech, finance, e-commerce, etc.). However, there are aspects of FinOps and GreenOps that might be relevant to any company that uses the cloud.

When public cloud computing first took off, it created opportunities for cost savings … and risks of spending way too much. FinOps was born as a way of keeping tabs on cloud usage. The FinOps Foundation, a programme of the Linux Foundation, has a complete FinOps framework. Other variations exist. The cloud giants all provide tooling to support FinOps, e.g. Azure Advisor and Azure Cost Management.

GreenOps is now getting attention. Sánchez and García (2024) write:

> FinOps and GreenOps are strongly related to each other. When we work in cost optimization initiatives, such as powering off virtual machines during off-hours, or rightsizing resources to maximize usage, we are effectively also reducing the carbon emissions that our cloud resources generate. This is a win-win situation for organizations, as they both benefit from the cost optimization side and the improvement to their sustainability.[29]

They further add:

> GreenOps also can benefit from FinOps in other ways: we can, for example, use FinOps practices on dashboards, reports, and the design and definition of KPIs, including carbon footprint metrics and sustainability KPIs in the picture and fostering both FinOps and GreenOps at the same time by creating common work areas and assets between these two interconnected methodologies. FinOps and GreenOps also seek to increase organizational awareness and visibility of both cloud costs and carbon emissions, respectively. By increasing awareness, we will also increase accountability and give everyone a common goal.[30]

GreenOps involves capturing good granular data about the sustainability impacts of your cloud usage, and ensuring that this data is acted on to improve performance. Vik Saluja of Mastercard comments, “Cost is everybody's responsibility, right? So your goal would be sustainability is everybody’s responsibility too. But it seems that engineers are more passionate about sustainability. So their reaction to this, with both kinds of data aligned together, goes a long way to giving you more efficiency across the whole ecosystem” (FinOps Foundation 2024). Carbon KPIs (and other environmental KPIs) can be used to drive more sustainable choices, build greener software, and deploy it in greener ways.

Cost savings and sustainability are certainly not always correlated at a granular level; however, one recent survey found that, “If you want to save money in Enterprise IT, it turns out that sustainability as a KPI is more important than cost” (Butcher 2024). Of course, lots of data is required, from monitoring users’ behaviours to optimise resource allocation, to building in spatial and temporal carbon intensity variability. Such data is not easy to collect. There can be downsides to these data-driven approaches, if that data collection is also understood as data surveillance (see e.g. Zuboff 2020).

The concept of **minimising waste** can extend even further. What does the company need to be doing in the first place? Might demand shaping go beyond offering different versions of the same service, to offering different services? Might we even begin to think about different business models, different visions? What do we consider is necessary in the first place? Where does that

---

[29] Alfonso San Miguel Sánchez Danny Obando García, *Efficient Cloud FinOps: A practical guide to cloud financial management and optimization with AWS, Azure, and GCP* (Packt 2024).

[30] Alfonso San Miguel Sánchez Danny Obando García, *Efficient Cloud FinOps: A practical guide to cloud financial management and optimization with AWS, Azure, and GCP* (Packt 2024).

necessity come from? How are our systems and roles set up to decide what is necessary? To the extent that it combines highly practical tools and workflows with an ambition to see the full picture of sustainability, GreenOps is really about the business's **social licence to operat**e.

To understand FinOps and GreenOps, it's helpful to go back even further, to **DevOps**. DevOps emerged in the 2000s (although you can find earlier precendents). DevOps represented a **shift from traditional software development and delivery approaches**, where the people making the software (Dev) and the people deploying and managing the software (Ops) were often quite siloed. ("Ops" is usually understood to stand for "operations," although sometimes it's "optimisation.")

DevOps, by reconfiguring various traditional roles, responsibilities, skills, and workflows, aimed for **faster delivery of features, more stable operating environments, quicker problem resolution, and more time devoted to innovation** as opposed to bug-hunting. Continuous Integration / Continuous Delivery (CI/CD) improves efficiency, and allows teams to quickly identify and address issues, enhancing overall productivity and reliability. (However, a recent state of DevOps report suggests that DevOps may have become a victim of its own success, enabling more complex projects at the expense of reduced debuggability).

DevOps was and is about **collaboration**. There are many different ideas about what good collaboration looks like, and DevOps is definitely the tech industry version. It goes in hard on **automation**, **continuous integration**, and **quick feedback cycles**. DevOps is about cultivating mutual understandings, but also removing the need for mutual understandings through technological tools and platforms. In other words, it is a model of collaboration where the conversations you *don't* have to have are as important as the conversations you do have.

DevOps fits very well with software built out of **modular microservices**: small bits of code which are loosely coupled with one another, and communicate via lightweight protocols. You can update or tinker with one component without worrying too much about unintended impacts to the entire system. In this respect, DevOps has a soupçon of **open source culture** about it.

As the popularity of cloud computing grew, enter **FinOps**. FinOps is embedded in the same general ethos as DevOps. Cloud computing means, very crudely and reductively, outsourcing your computing hardware requirements to specialists like the cloud giants. The idea is, you only rent the resources you need, when you need them, so your applications become very flexible and scalable.

As the popularity of the so-called public cloud grew, a problem emerged. All that flexibility could have the opposite effect from what was intended. **Some companies found they were renting what they didn't need.** The procurement function was effectively distributed across the organisation. The staff authorised to buy these services included personnel who were much more interested in solving specific technical problems than in weighing up the business case for their investigations and solutions. There is also something of a perverse incentive at play: it's hard for a cloud provider to really want to remind its customer to stop buying something.

The “Fin” in “FinOps” stands for Finance. If DevOps integrated software development with deployment/management, FinOps further integrated financial controls. FinOps is an approach designed to bring **financial accountability** to the variable spending model of the cloud, hopefully enabling teams to balance speed, cost, and quality. The FinOps Foundation says:

> At its core, FinOps is a cultural practice. It’s the way for teams to manage their cloud costs, where everyone takes ownership of their cloud usage supported by a central best-practices group. Cross-functional teams in Engineering, Finance, Product, etc work together to enable faster product delivery, while at the same time gaining more financial control and predictability.

DevOps and FinOps are related in that both aim to optimise processes within companies, but they focus on different aspects. The relationship should be complementary. FinOps can be seen as an extension of the DevOps culture, integrating financial metrics and considerations into the continuous cycle of software delivery. FinOps focuses on the financial aspect, aiming to maximise the value of cloud spend. It involves practices like monitoring, documenting and controlling decisions about resource allocation, and aligning spending with business outcomes.

Then there’s MLOps, ModelOps, etc. (Is -Ops turning into a buzzsuffix?) So what, finally, is **GreenOps**? GreenOps builds sustainability considerations into FinOps. The cloud is material, and storing, processing and moving data has implications in terms of carbon emissions, water use, and demand on renewable energy that might otherwise be used for something else. GreenOps ensures that decisions are informed by data on carbon emissions, water use, and other environmental impacts. A really flourishing GreenOps function will have cloud teams and sustainability teams working side-by-side, complementing each other’s skillsets and learning from each other, backed by real understanding and ownership at a senior level.

Jevons’ Paradox and Rebound Effects

William Stanley Jevons wrote in *The Coal Question* (1865), “It is a confusion of ideas to suppose that the economical use of fuel is equivalent to diminished consumption. The very contrary is the truth.” Jevons’ Paradox (occasionally misleadingly called Jevons’ Law) does not apply to all things, and when it does apply to something, it doesn’t apply to it all the time under any circumstances. It is something that *can* happen.

Freitag et al. (2021) write:

> ICT has seen rapid and continuous efficiency gains. Yet increases in demand for computation and the number of ICT-enabled devices per person have outpaced these energy efficiency improvements, resulting in growth in ICT’s energy consumption and carbon footprint year-on-year. This pattern fits with the rebound effect described by Jevons Paradox whereby an efficiency improvement leads to an even greater proportionate increase in total demand, meaning total resource requirements rise rather than fall, as is often assumed. While Jevons Paradox has not been proved to apply within the ICT industry, it is risky to assume it does not apply given historical evidence of ICT emissions consistently rising despite significant improvements in efficiency (ICT’s carbon footprint).

When we talk about improving efficiency, we’re usually talking about increasing outputs per input. For instance, imagine you have a car that becomes more fuel-efficient, meaning it can travel further on the same amount of fuel. If your car used to consume 10 litres of fuel to travel 100 miles, but after becoming more efficient, it is providing more miles of travel per litre. Since inputs have costs, this usually also means that you can travel more miles for every pound you spend.

So if your car gets more efficient overnight, does that mean you’ll use less fuel overall?

Not necessarily, because you might drive more! When the cost of something goes down, people often want more of it. That’s the classic downward-sloping demand curve you’ll see in most introductory economics textbooks. In our example, if travelling becomes cheaper because your car is more fuel-efficient, you might decide to drive more often or take longer trips, since it’s now less expensive to do so.

This increased usage due to the lower cost is an example of a **rebound effect**. The savings from using less fuel per mile might be offset by the fact that you’re now driving more miles overall.

The rebound effect can vary in size. If it’s small, you still save fuel overall, even though you’re travelling more. But if the rebound effect is large, you might end up using the same amount of fuel, or even more. Jevons’ Paradox refers to the last possibility: when the rebound effect is so large that it completely cancels out the efficiency gains or even results in more resource use than before.

The likelihood of Jevons' paradox occurring depends partly on how much the amount demanded changes in response to a change in price. This is known as **price elasticity of demand**. In fact, fuel tends to be fairly price inelastic—people need to get to where they're going, they've already invested in a car, and they may have few alternatives (depending on the state of the trains and buses). A life-saving medication may also be very price inelastic.

If demand is price elastic, meaning a small drop in price leads to a large increase in demand, then a reduction in the cost of travel could lead to a much larger increase in how much people travel, making the Jevons paradox more likely. On the other hand, if demand is price inelastic, meaning that a drop in price doesn't cause much of a change in demand, then people won't travel significantly more just because it's cheaper, reducing the likelihood of Jevons' paradox occurring.

Suppose a company builds servers that are more energy efficient. Initially, this might seem like a clear win for reducing energy consumption. However, if this translates to lower compute cost for customers, it might encourage more widespread use. More users might run workloads they otherwise wouldn't, or might keep instances running continuously. Companies might deploy software on a larger scale across more devices. As a result, the overall energy consumption could increase, despite the hardware being more efficient on a per-use basis.

In the real world, things are often more complex because outputs typically rely on more than just one input. For example, in addition to fuel, travel depends on other factors like time, labour (e.g. the driver's time), the cost of vehicles, the road infrastructure, etc. Costs and benefits can be distributed in complicated ways. In the case of more energy-efficient hardware, another potential example of a rebound effect would be developers creating more demanding software to run on it.

Implementing GreenOps

Actually, FinOps may well have been contributing to cloud sustainability, almost by accident. Cost can sometimes (although definitely not always) be a good proxy for sustainability. So in some cases, by controlling costs, FinOps helped to keep a lid on carbon impacts. As such, many FinOps practices dovetail with GreenOps practices, and GreenOps borrows a lot from FinOps.

However, as James Hall, Head of GreenOps at Greenpixie, points out:

> True GreenOps does not rely on a cost proxy. There are certain FinOps practices which reduce cost but can actually increase usage and therefore emissions so this can be a dangerous equivalence. GreenOps relies on proper sustainability data

which facilitates informed-decision making in the cloud.

Hall highlights that cloud engineers will sometimes need to pick between the cheaper option and the more environmentally friendly option.

The following FinOps practices do reduce cloud usage and therefore can be considered as GreenOps practices too — as they will have the added effect of reducing water and/or carbon. This list has been loosely based on recommendations from Greenpixie's *GreenOps for Cloud Computing* white paper (2024).[31]

- **Right-size underutilised instances:** Adjust the size of cloud computing resources to match what's actually needed, to avoid paying for capacity you aren't using. (An 'instance' is essentially a virtual server rented from third-party cloud services, allowing you access to compute resources to run applications, databases, microservices, etc.).

- **Select energy-efficient instance types:** Choose computing resources that use less energy for the same work, similar to preferring an energy-efficient refrigerator that uses less electricity. Consider processors and servers with less embodied carbon.

- **Review aged instances:** Check and update or replace cloud resources that have been running for over a year. Newer options might be more efficient or sustainable. (We are referring to the services here, not the underlying hardware).

- **Terminate unused instances:** Stop and remove cloud resources that are no longer in use. Like switching off the light when you leave a room!

- **Terminate zombie instances:** Identify and remove cloud resources that are running, but not for any good reason.

- **Increase usage of auto-scaling and scheduling:** Automate the adjustment of resources based on need, ensuring that you have more when demand is high and less when it's low.

- **Optimise logging (e.g. AWS CloudTrail):** Improve the way that records of computer events are kept. For instance, AWS CloudTrail captures logs from a wide range of AWS services, e.g. Amazon S3, EC2, IAM, Lambda.

- **Optimise http(s) API traffic:** Use best practice around APIs such as caching to lower the strain on servers, compression to move less data around, HTTP/2 for a smoother online experience than the older HTTP/1.1, sensible rate limits, query optimisation for data lookups quicker, etc.

GreenOps is also related to well-architected frameworks, to carbon-aware computing, and to the emerging grid-aware computing paradigm. The same compute can be done with a big carbon footprint or a small one, depending where and when it occurs. Sometimes, this can literally come down to whether the sun is shining and the wind is blowing, supplying clean energy to the power grid. As such, another GreenOps strategy is

---

[31] Draft copy, personal correspondence (2024).

scheduling workloads for times when the data centre's local grid is low-carbon. This is an example of how important cloud sustainability data is over spend-based proxies. Scheduling a workload for a period of lower carbon intensity will not necessarily be cheaper, but can save an enormous percentage of emissions.

Another example of this kind of informed carbon avoidance in cloud GreenOps is **selecting low carbon regions** for your reserved instances or new applications. If the trade-off with performance, latency and cost is feasible, it can be a simple solution to select a region which operates on a lower carbon grid.

This guide from Greenspector goes in more depth into DevOps and GreenOps (DevGreenOps), including priority metrics, detecting green bugs via shift left and shift right tests, feature flipping and A/B testing, and specific tools such as Référentiel NR, Greenspector Studio, Scaphandre, Power API, Carbonifer, Easyvirt, and Ecocode.

There are also **demand shaping** approaches which involve changing the nature of what you do depending on the carbon intensity of the power grid over time. For example, you could switch off certain features, or put them behind a paywall (with proceeds going directly to offsets) during high intensity periods.

GreenOps is also part of the larger opportunity to rethink **where sustainability expertise and responsibilities sit within the organisation**. Truly transformative sustainability is still hampered by its historic links to marketing and PR, and to some extent HR and Health and Safety. Things are changing, but we still don't really know the best model for reconfiguring sustainability, so that key decision-makers don't encounter sustainability as an exogenous constraint that doesn't even have a credible owner. Talk of embedding sustainability throughout an organisation is promising, but there is the phenomenon whereby something that is everybody's responsibility becomes nobody's responsibility.

Well-Architected Framework

A well-architected framework is a set of guiding design principles and best practices developed by the cloud giants to help users understand how to develop and operate reliable, secure, efficient, and cost-effective systems in the cloud. This framework provides an approach to evaluating architectures and implementing designs that will scale over time. The cloud giants each have their own versions of this framework, but they generally cover similar concepts divided into several core pillars:

- Operational Excellence: Focuses on running and monitoring systems to deliver business value and continually improving processes and procedures. Key practices include automation, routine performance reviews, and incident response.
- Security: Prioritises protecting information and systems. Guidelines typically include identity and access management, data encryption, and security monitoring.
- Reliability: Ensures that a system can recover from failures and continue to function. This involves setting up failover mechanisms, backup and restore procedures, and carefully planned incident management strategies.
- Performance Efficiency: Involves using IT and computing resources efficiently to meet system requirements. This pillar often emphasises the right choice of resource types and sizes based on workload requirements, monitoring performance, and making informed decisions to maintain efficiency as technology evolves.
- Cost: Focuses on avoiding unnecessary costs. Understand and control where money is being spent, select the most appropriate and right number of resources, analyse spending over time, and scale to meet business needs without overspending.
- Sustainability: Currently present only in AWS's framework. AWS proposes a shared responsibility framework, in which AWS is responsible for the sustainability *of* the cloud and clients are responsible for sustainability *in* the cloud.

Cloud Governance Glossary

See also Cloud Computing and AI Glossary.

1.5 Degrees: The Paris Agreement is an international treaty adopted in 2015 by almost all countries in the world. It aims to limit global warming to well below 2 degrees Celsius above pre-industrial levels, and ideally to 1.5 degrees Celsius. ‘Well below 2 degrees’ is nowadays sometimes interpreted as 1.7 degrees. The 1.5-degree target is considered crucial to mitigate the most catastrophic impacts of climate change. It is not an ‘all or nothing’ type scenario: climate change is already having effects, which may already feel catastrophic to those most impacted by them. The 1.5 degree target has been urgently pursued by scientists, NGOs, concerned citizens, and some policymakers, for a decade, but overall progress has been too slow and at time of writing it appears that 1.5 degrees has just slipped away. It remains to be seen how the discourse will shift: perhaps 1.6 degrees will become the new ambitious target. Climate science involves some intrinsic uncertainties, which sometimes get left out in the corporate sustainability world, because they are too complex. For example, technically speaking, keeping within our global carbon budget and reaching net zero in a timely fashion wouldn’t guarantee limiting warming to 1.5 degrees, it would just give a good chance at it. Also there is some variation in average temperatures to do with other factors. That’s why, even though we have already exceeded 1.5 degrees in 2023, you are not yet seeing lots of unequivocal news items about breaching the Paris Agreement’s 1.5-degree target. For this to be declared, the global average temperature will need to consistently exceed 1.5 degrees above pre-industrial levels over multiple years. See also *Net Zero*, *Paris Agreement*.

2015 Paris Agreement: See *Paris Agreement*.

Adaptation: Action to address climate change is often divided into “adaptation” and “mitigation.” Adaptation refers to all the things we are doing to adapt to climate impacts, such as more resilient agriculture, flood defences, improving water management systems, building infrastructure to withstand extreme weather, improving disaster preparedness and management, protecting public health against the spread of climate-related diseases, and so on. See *Mitigation*.

Additionality: Additionality refers to the principle that a carbon credit must result in emissions avoidance or reductions that would not have occurred without the purchase. In other words, the finance must provide an environmental benefit that is “additional” to what would have happened under a business-as-usual scenario, and the purchaser should not be credited with avoiding or reducing more emissions than they actually have. See *Emissionality*.

Algorithmic Impact Assessment: An Algorithmic Impact Assessment (or AI Impact Assessment) is a systematic process used to evaluate the potential social, ethical, and legal impacts of developing and/or deploying an AI or other algorithmic system, aiming to identify and mitigate negative consequences.

Avoided emissions: Reduction or prevention of greenhouse gas emissions that would supposedly have otherwise occurred, typically through actions like energy efficiency improvements, shifting to renewable energy, or paying owners of forests or wetlands to preserve them. In contrast, carbon removals involve actively extracting CO2 from the atmosphere, such as through reforestation or Negative Emissions Technologies, rather than just preventing hypothetical emissions. See also *Additionality*, *Emissionality*, *Removals*.

Cap-and-trade: An approach to decarbonisation where organisations are issued with a certain number of emissions credits; those that emit carbon in excess of their allotted amount must buy credits from those who have spare credits. See *EU Emissions Trading System*.

Carbon: See *Greenhouse Gases*.

Carbon credit: See *Offset*.

Carbon elasticity of income: The responsiveness of carbon emissions to changes in income for a household, individual, country, etc. If income goes up by 10%, will CO2e emissions rise by 10% as well (a carbon elasticity of 1.0)? The same basic concept is applicable to other variables as well, e.g. carbon elasticity of total expenditure.

Carbon Neutral: When the carbon released into the atmosphere is fully balanced by carbon removed and/or carbon emissions avoided/reduced, either directly or by purchasing carbon offsets. Carbon neutrality is generally a lower bar than net zero. Watch out! There are some different definitions of 'carbon neutral' and 'net zero' floating around. If you search "What is the difference between carbon neutral and net zero?" you can quickly find several explainers, all confidently telling you different things. The key thing is to examine what the particular company or individual using the term means by it. Read the small print. See also *Offsets* and *Net zero*.

CDP (formerly Carbon Disclosure Project): An important entity in the ESG ecosystem. CDP is a not-for-profit organisation that runs a major global disclosure system for companies, cities, and other entities. They provide scores based on the disclosed information. CDP also has a framework for SMEs. The S&P Global Corporate Sustainability Assessment is influential in ESG ratings.

CO2e: CO2 means carbon dioxide. CO2 is a greenhouse gas, so releasing it into the atmosphere heats up the climate. CO2e (or CO2eq) means "CO2 equivalent." It simply means an amount of greenhouse gases (carbon, methane, etc.) translated into carbon terms, for convenience of comparison. Carbon can exist in various forms and compounds, depending on the elements it bonds with. For example, carbon can form organic compounds like carbohydrates, proteins, and fats, or it can be found in pure forms like diamond and graphite. The reaction between carbon and oxygen, which forms carbon dioxide (CO2), occurs during combustion, such as burning fossil fuels.

Corporate Social Responsibility (CSR): A business model in which companies integrate social and environmental concerns in their business operations and interactions with their stakeholders. See also *ESG*.

CSRD (Corporate Sustainability Reporting Directive): The EU CSRD expands the scope of sustainability reporting requirements for companies in the EU, demanding more detailed reporting on environmental and social impacts, governance, and activities, enhancing the transparency and comparability of sustainability information.

Eco-label: An eco-label is a mark or a label given to products that are deemed to meet certain environmental performance criteria, signifying their lower impact on the environment compared to other similar products. At the time of writing, there are no widely recognised eco-labels for AI (and it is not clear that existing eco-label assumptions and methods would be fit-for-purpose), but there are many projects in motion to investigate / develop such labels.

Energy Efficiency Directive: The EU's Energy Efficiency Directive is wide-ranging legislation to encourage efforts to use energy more efficiently. Recent amendments in 2024 relate to data centre transparency specifically.

Embedded Emissions: See *Embodied Carbon.*

Embodied Carbon: Embodied carbon refers to the total emissions generated during the entire lifecycle of a product, including extraction, manufacturing, transportation, and end-of-life. In the context of cloud computing, for example, embodied carbon will refer to things like the emissions associated with data centre hardware (as opposed to the electricity that the hardware uses). Also known as *Embedded Emissions.* See also *Life Cycle Analysis*.

Emissionality: The effectiveness of renewable energy projects in reducing greenhouse gas (GHG) emissions, focusing on the impact these projects have on displacing carbon-intensive energy from the grid. The Emissions First principles, which Amazon supports, build on this concept by encouraging companies to prioritise investments in the decarbonisation of global electricity grids, particularly in areas that have not traditionally received corporate clean energy investment. These principles take a global perspective, acknowledging that all GHG emissions contribute to atmospheric impacts, and they aim to value clean energy procurement based on the specific emissions reductions achieved on the affected grid. See *Avoided emissions.*

ESG: ESG, which stands for Environmental, Social, and Governance, refers to a set of criteria used to evaluate a company's ethical and sustainability practices. The term ESG can be used loosely to mean something like CSR, Corporate Citizenship, Sustainability, ethical investment, or impact investment. Technically speaking ESG is not rooted in ambitions to make a positive impact; it is a set of risk factors, which include the risk that customers, regulators, litigators, or other stakeholders will hold a company accountable for its unethical behaviours. Environmental criteria consider how a company manages its impact on natural resources and climate, Social criteria assess how it manages relationships with employees, customers, and communities, and Governance criteria examine the company's leadership, executive pay, audits, and shareholder rights. ESG factors are used by investors to assess the long-term risks and opportunities associated with a company. ESG might, somewhat provocatively, be described as what happened when finance and investment took over the CSR agenda. See *CSR*, *ESG Ratings Agency*.

ESG Ratings Agency: A company that rates agencies (especially publicly traded companies) on their environmental, social, and governance dimensions. The ESG ratings industry sprang up out of the financial ratings industry, and the information they generate is primarily aimed at investors. See *ESG*.

EU Emissions Trading Scheme: The EU Emissions Trading Scheme is a cornerstone of the EU's policy to reduce greenhouse gas emissions. It was the world's first major carbon market and remains the largest. It is a cap-and-trade scheme. There are many other Emissions Trading Schemes worldwide. The International Carbon Action Partnership maps these schemes worldwide. See *Cap and trade*.

FinOps: A financial management practice for the cloud enabling teams to collaborate on data-driven spending decisions.

Flexible Computing: Optimizing computational tasks by adjusting the intensity of operations based on the availability, cost, and environmental impact of electricity. The core idea is to increase computing during periods when electricity is abundant, inexpensive, and generated from renewable sources, and to decrease computing during times when energy is more expensive, scarce, or derived from polluting sources. It is therefore closely related to carbon-aware computing.

Greenhouse Gas (GHG): Greenhouse gases are gases that heat our world's climate. They do this by trapping heat from the sun within the atmosphere. Carbon dioxide (CO2) is a very important GHG, which is why we often just talk about carbon emissions, or translate all GHG emissions into carbon dioxide equivalent (CO2e). There are other greenhouse gases, such as methane. While these gases are essential for keeping the Earth warm enough to support life, too much of them leads to global warming and climate change.

Greenhouse Gas Protocol (GHG Protocol): A comprehensive set of tools and standards for measuring and managing greenhouse gas (GHG) emissions. The GHG is where the terms Scope 1, Scope 2, and Scope 3 come from. The GHG Protocol is widely used by businesses, governments, and NGOs to measure and report their GHG emissions, and to develop their strategies for achieving net zero. In this report, the GHG Protocol is often shorthand for the GHG Corporate Standard, which is aimed at organisation. It categorises emissions into three 'scopes': Scope 1 covers direct emissions from owned or controlled sources, Scope 2 includes indirect emissions from the generation of purchased electricity, and Scope 3 encompasses all other indirect emissions in a company's value chain (emissions associated with suppliers, customers using and disposing of a product, and so on). The GHG Protocol has been developed through a partnership between the World Resources Institute (WRI) and the World Business Council for Sustainable Development (WBCSD).

GeSI (The Global Enabling Sustainability Initiative): A global membership organisation dedicated to enabling the ICT industry to meet environmental and social challenges. GeSI has published sectoral guidance for ICT companies seeking to align with the UN-backed SBTi (Science Based Targets initiative).

Global Reporting Initiative (GRI): An international independent standards organisation that helps businesses, governments, and other organisations understand and communicate their impacts on issues such as climate change, human rights, and corruption.

Gold Standard: A standard for climate and sustainable development initiatives, best known as a standard for carbon credits.

Green Claims Directive: A Directive currently being developed in the European Union to crack down on explicit forms of greenwashing, and help consumers to access more reliable information about the environmental impacts of their consumption choices.

GreenOps: An approach to incorporating environmental considerations into operational strategies, often by using data analytics and sustainability metrics. An evolution of DevOps / *FinOps*.

Implied Temperature Rise: A metric used within a finance and investing context to estimate how the global temperature would change if all investments mirrored the carbon emissions intensity of a particular investment portfolio.

The Integrity Council for the Voluntary Carbon Market (ICVCM): An independent governing body that oversees the voluntary carbon market, with the goal of ensuring it effectively contributes to a just transition toward the 1.5°C climate target. It maintains a threshold standard, which is built on the Council's ten Core Carbon Principles (CCPs), and enforced through an Assessment Framework. In this way it seeks to provide an easy and reliable way to identify high-quality carbon credits. See also *VCMI*.

IPCC (Intergovernmental Panel on Climate Change): The United Nations body for assessing the science related to climate change. The global authority on climate change — based on extensive processes of peer review.

ISSB (International Sustainability Standards Board): A board under the IFRS Foundation that develops and approves IFRS Sustainability Disclosure Standards. The Financial Stability Board has declared that the TCFD's efforts have reached their conclusion, with the ISSB's Standards representing the 'final phase of the TCFD's endeavors.' By adhering to the

IFRS S1 General Requirements for Disclosure of Sustainability-related Financial Information and IFRS S2 Climate-related Disclosures, companies will align with the TCFD recommendations since these recommendations are fully integrated into the ISSB's Standards.

While companies have the option to continue utilising the TCFD recommendations, and some may still be obligated to do so, adopting the recommendations serves as a solid starting point for companies transitioning to the ISSB's Standards. The IFRS Foundation has released a document comparing the IFRS S2 requirements with the TCFD recommendations, showing that IFRS S2 aligns with the TCFD's four principal recommendations and eleven suggested disclosures. The comparison makes clear that companies employing the ISSB Standards will inherently comply with the TCFD recommendations, eliminating the need to apply both the TCFD recommendations and the ISSB's Standards concurrently.

Life Cycle Analysis (LCA): A systematic assessment of the environmental impact of a digital product or service throughout its entire life cycle, from raw material extraction to disposal. It evaluates factors such as energy consumption, carbon emissions, and resource use at each stage of the product's life. See also *Embodied Emissions*.

Mitigation: Action to address climate change is often divided into “adaptation” and “mitigation.” Mitigation basically refers to getting to net zero (and ideally net negative), so that the climate stops heating up. “Mitigating climate change” is similar to “stopping climate change,” although we tend to avoid saying “stopping climate change,” because that implies it is something that might happen in the future, rather than something which has already happened and continues to worsen. We have changed, and continue to change, the climate. Mitigation is roughly synonymous with decarbonisation, and has two main aspects: reducing the amount of greenhouse gases we are putting into the atmosphere, and increasing the amount of greenhouse gases that are taken out of the atmosphere. See also *Adaptation*.

MSCI: A provider of investment decision support tools, including indices, portfolio risk and performance analytics. In this context, a provider of ESG ratings.

Net Zero: A state in which a balance is achieved between the amount of greenhouse gas emissions produced and the amount removed from the atmosphere over a given time period (e.g. an annual basis). Achieving net zero for the planet as a whole should halt global warming. Net zero can also refer to a smaller scale, e.g. getting to net zero as a country, a sector, or an individual company. For companies, the Science Based Targets initiative has been the gold standard for defining net zero. The SBTi requires a company to reduce its greenhouse gas emissions to as close to zero as possible, and to remove any residual emissions from the atmosphere by purchasing carbon credits representing removals (not avoided emissions), and also requires a decarbonisation pathway consistent with 1.5 degrees. The SBTi is trying to ensure that ‘net zero’ at the corporate level aligns with ‘net zero’ at the planetary level. Carbon neutral and net zero are easily confused, but they are accounted for in different ways. Net zero has the more stringent criteria. See also *Carbon neutral*.

Offset: The role of offsetting within corporate carbon neutrality and/or net zero is complex and justly controversial. SBTi (2020) advises, “Companies should follow a mitigation hierarchy that prioritizes eliminating sources of emissions within the value chain of the company [abatement] over compensation or neutralization measures.” The “value chain” here includes scope 1, 2, and 3 emissions. Beyond this, then there is offsetting. The Science Based Targets initiative (SBTi) makes a distinction between “neutralising” emissions and “compensating” emissions. These are both ways of offsetting carbon emissions. **Neutralising** emissions means spending money in ways that remove emissions from the atmosphere (for example, by planting forests or restoring wetlands to absorb more carbon). There are issues around the permanence of such removals. **Compensating** means spending money in order (supposedly) to ensure that some carbon emissions that *would* have been released are *not* released. For example, you might buy a carbon credit that provides grants for some third party to install solar panels, or to switch to a more sustainable agricultural or industrial technique, or pays landowners not to deforest their lands, or funds the distribution of cleaner cookstoves. Clearly the additionality of all offsetting (i.e. would it really not have happened, if the offsetting company had not bought the carbon credit) is questionable, and the additionality of many “compensating” approaches is particularly questionable. See also the sidebar on Offsetting.

Paris Agreement: The international treaty on climate change, aiming to limit global warming to well below 2 degrees Celsius above pre-industrial levels, and ideally 1.5 degrees. Almost every government in the world has signed up to the agreement. Note that there is a big difference between *well below 2 degrees* and *just below 2 degrees*. See also *1.5 Degrees*, *Net Zero*.

PAS 2050: PAS 2050 is a publicly available specification for product life cycle CO2e emissions.

RE100: An initiative bringing together businesses committed to 100% renewable electricity in their operations. Companies joining RE100 pledge to achieve this goal by a specific target year. The initiative also provides technical guidance useful to both members and non-members to avoid greenwashing in their renewable energy claims. However, at time of writing, this may involve contentious renewable energy procurement strategies. Recent messaging from RE100 seems to suggest that RE100 intends to follow the GHG Protocol’s lead on 24/7 hourly matching, rather than come to its own view.

Refinitiv: A global provider of financial market data and infrastructure. In this context, a provider of ESG ratings.

Removals: Greenhouse Gas Removals (GGR) are processes that actively remove greenhouse gases, particularly carbon dioxide, from the atmosphere, helping to reduce the overall concentration of these gases and mitigate climate change. These removals can occur through nature-based solutions like afforestation and soil carbon sequestration, or through technological approaches such as Direct Air Capture. On almost all credible models, removals make up a fairly tiny proportion of transition to net zero, but they are still considered important.

Renewable Energy: Energy generated from natural resources such as sunlight, wind, rain, tides, and geothermal heat. Unlike fossil fuels, renewable energy produces little to no greenhouse gas emissions, making it crucial for stopping climate change. The name "renewable energy" can be a bit misleading, since it dates from a time when the big worry was that we were going to run out of these natural resources. As is now more widely understood, we have already discovered more of these resources than we ever should burn (hence the environmentalist slogan, "Leave It In the Ground"). Nuclear power is a special case. The environmental problems with nuclear power are not the same as with fossil fuels. They relate rather to the limited and uneven global distribution of uranium, along with the complex infrastructure needed for its extraction and refinement, and the safe long-term storage of radioactive waste, which will pose significant environmental and security concerns for centuries to come. Biomass is also a somewhat special case, since it does produce greenhouse gas emissions, but not at the same level as fossil fuels, and because the carbon released during combustion is roughly equal to the amount absorbed by the plants during their growth.

Scopes 1-3: Categories used to define the direct and indirect emissions of a company. Scope 1 covers direct emissions, Scope 2 covers indirect emissions from purchased electricity, and Scope 3 covers other indirect emissions from the supply chain. The scopes are maintained by the GHG Protocol. There is an update in the works. See also *Greenhouse Gas Protocol* and the sidebar on Scope 3.

'Scope 4': A new and confusing term, probably better avoided. It is not part of the Greenhouse Gas Protocol. It is very different in meaning from Scopes 1 to 3, referring to activities that supposedly reduce emissions. There is a risk of counting activities that are actually within an organisation's value chain against its emissions. Here is an extreme example of how it could be misused: I decide to drive to my destination, instead of taking a plane. Obviously I've emitted some carbon, but then I subtract the carbon I did not emit from the plane flight: *voila*, I've actually *reduced* the carbon in the atmosphere! Perhaps Scope 4 can be rigorously and responsibly used? But it feels to us like an invitation to greenwashing. Nonetheless some companies will be using it and each case should be considered on its merits.

S&P (Standard & Poor's): A financial services company known for its stock market indices such as the S&P 500, and for providing credit ratings. In this context, also a provider of ESG ratings.

Science-Based Targets initiative (SBTi): Currently the best independent standard for corporate net zero. A partnership between CDP, the UN Global Compact, World Resources Institute, and the WWF that champions science-based target setting as a way to boost companies' competitive advantage in the transition to a low-carbon economy. SBTi has been plunged into controversy in 2024. *The Guardian* reports: "Staff at one of the world's leading climate-certification organisations have called for the CEO and board members to resign after they announced plans to allow companies to meet their climate targets with carbon offsets." Recently, several hundred companies failed to meet a 24 month deadline to submit their science-based targets for validation after making an initial commitment to align with SBTi. This led to their commitment status being changed to 'commitment removed' on the SBTi's company targets dashboard. Amazon's status change to 'commitment removed' took effect in mid-2023 and for Microsoft in early 2024. Alphabet (Google) are still included. SBTI's current direction of travel on corporate net zero is indicated in recent technical documents. See also *Net zero*.

SEC Climate-Disclosure Rule: This rule proposed by the U.S. Securities and Exchange Commission requires publicly traded companies to disclose comprehensive information about their climate-related risks and greenhouse gas emissions, aiming to provide investors with consistent and reliable data for decision-making.

SECR (Streamlined Energy and Carbon Reporting): This UK regulation mandates that certain large companies report their energy use, greenhouse gas emissions, and energy efficiency actions in their annual reports, aiming to encourage businesses to reduce energy consumption and carbon emissions.

Shared Socioeconomic Pathway (SSP): In the last reporting cycle, the IPCC's Shared Socioeconomic Pathways (SSPs) were five distinct scenarios that describe different trajectories of global societal development, based on varying assumptions about economic growth, technological advancement, demographic changes, and policy decisions. These scenarios are reference points, used to assess the potential impacts of climate change and the effectiveness of mitigation and adaptation strategies under different future conditions. Each SSP outlines a unique combination of challenges to mitigation and adaptation, ranging from sustainable development (SSP1) to a world with high inequality and fossil fuel dependence (SSP5).

Sustainalytics: A company that provides environmental, social, and governance (ESG) research and ratings to help investors make more informed decisions. In this context, a provider of ESG ratings.

TCFD (Task Force on Climate-related Financial Disclosures): An organisation established by the Financial Stability Board to develop climate-related financial risk disclosures for use by companies in providing information to stakeholders.

TNFD (Taskforce on Nature-related Financial Disclosures): An initiative aimed at providing a framework for organisations to report and act on evolving nature-related risks.

United Nations Global Compact: A voluntary initiative based on CEO commitments to implement universal sustainability principles and to undertake partnerships in support of UN goals.

Verified Carbon Standard (VCS): A certification programme for carbon credits, from the non-profit Verra. See also *Gold Standard*.

The Voluntary Carbon Market Integrity initiative (VCMI): Established in 2021 in preparation for COP26, to collaborate with various stakeholders seeking to improve the voluntary carbon markets. The VCMI Claims Code of Practice sets guidelines for companies using carbon credits in credible, science-based net-zero strategies. See also *ICVCM*.

World Resources Institute (WRI): A global research organisation that spans more than 60 countries, focusing on six critical goals that must be achieved to reduce environmental impact and improve economic opportunities.

World Wide Fund for Nature (WWF): An international non-governmental organisation working in the field of wilderness preservation, and the reduction of human impact on the environment.

Prototype artificial glacier in Ladakh

Scope 3

Current methods for accounting for carbon date to the 1990s, when the World Resources Institute and other non-profits founded the GHG Protocol. The ways that Scope 1, 2 and 3 are defined and reported on is under review. This is a contentious process, even within the cloud giants (with Amazon opposing 24/7 hourly matching, and Google advocating for it). Watch this space in 2025-2026.

Scope 1 is all the greenhouse gases you actually directly emit through your operations. **Scope 2** is the greenhouse gases you are responsible for through the energy you purchase.

Then there is **Scope 3**, which is essentially 'everything else.' So it is more difficult to estimate (although sometimes this difficulty can be exaggerated for convenience). In a survey of 239 companies, 119 (53.6%) cited Scope 3 as a particular area of challenge in relation to the SBTi's target-setting (SBTi 2024). Recent analysis suggests that the emissions gap between companies' Scope 3 emissions reduction targets and their current emissions amounts to around 1.4 gigatonnes of CO2e, and is projected to rise to over 7 gigatonnes by 2030 (VCMI 2023). A review of 200+ digital companies found that that Scope 3 emissions were, on average, over six times greater than their combined Scope 1 and 2 emissions (WBA / ITU 2024).

In 2023, the EU's CSRD guidance confirmed purchased **public cloud services fall within the Scope 3 category**.[32] If you have an on-prem data centre, and you decide to move to the cloud, that will typically shift some of your carbon emissions from Scope 2 (the electricity you used to power your data centre) to Scope 3 (emissions you are indirectly responsible for).

Scope 3 can be confusing. On the one hand, you might think: 'Why should a company have to decarbonise its value chain? Is that not by definition somebody else's responsibility? Can't everybody just do their own Scope 1 and 2, and forget about Scope 3 altogether? Isn't my Scope 3 somebody else's Scope 1 and 2?'

But there are **good reasons for companies to take responsibility for Scope 3 emissions**, even if this puts them in the tricky position of trying to influence suppliers over whom they have no direct control.

First, **not all companies are required to report their Scope 1 and 2 emissions** or to reduce them. Laws are different in different countries. Even in Europe where regulation is fairly advanced, reporting requirements only apply to larger companies. So some of your Scope 3 emissions might not be reported by anyone, unless you report them (even if, in theory, they would be somebody else's Scope 1 or Scope 2 emissions).

---

[32] 'AR 51. If it is material for the undertaking's Scope 3 emissions, it shall disclose the GHG emissions from purchased cloud computing and data centre services as a subset of the overarching Scope 3 category "upstream purchased goods and services."'

Second, **not all companies are well-incentivised to reduce carbon emissions**. For example, consumer pressure is one mechanism.

Third, if there is overlap in reporting, this can be a good thing. **Different stakeholders having an interest in the same set of emissions can increase scrutiny and data quality**. You might assume it's a bad thing — isn't there a risk of double-counting? Might we mess up and decarbonise too much? But this would be a misunderstanding of what corporate carbon reporting is used for. We are not 'adding up' all these corporate reports to figure out overall emissions levels — carbon concentration in the atmosphere can very easily be directly measured.

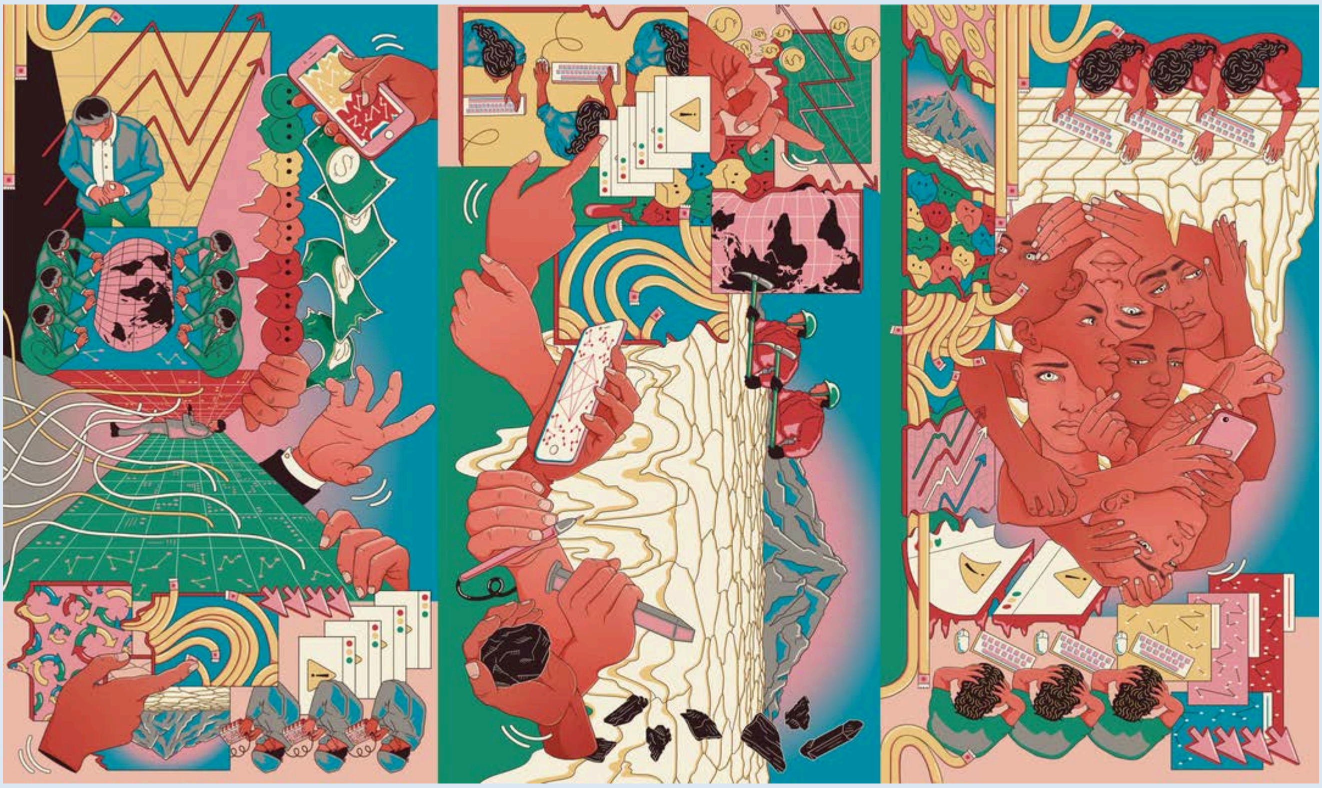
Fourth, **customers don't report on carbon emissions associated with use and end-of-life.**

The **Greenhouse Gas Protocol Corporate Value Chain (Scope 3) Standard** helps companies report on Scope 3 emissions from their value chain. Its main aim is to standardise how companies identify and tackle their largest GHG emission sources throughout their value chain, promoting sustainability in their operations and products. At time of writing in early August 2024, the GHG Protocol Scope 3 Standard identifies 15 categories of Scope 3 emissions, which cover various indirect emission sources across a company's value chain. These categories include emissions from both upstream and downstream activities, such as purchased goods and services, business travel, and use of sold products. Not all categories will be relevant to every company; relevance depends on the company's specific activities, industry, and the products or services it provides. Companies are encouraged to focus on the categories that are most significant to their own value chain emissions.

In 2023, the UK government issued a Call For Evidence on the costs, benefits, and practicalities of Scope 3 emissions reporting, an indirect emission in a company's value chain, which often represents a significant portion of up to 80-95% of total company emissions in many cases. In the SECR framework, Scope 3-emission was made voluntary for all except for the large companies required to disclose them in line with the Task Force on Climate-Related Financial Disclosures (TCFD) recommendations. This call acknowledges the need to ramp up action in this area and seeks to evaluate whether the UK should endorse the newly published standards, particularly IFRS S1 and S2 issued by the International Sustainability Standards Board (ISSB), which require disclosure of Scope 1, 2, and 3 emissions. The insight from stakeholders will help the government assess its stance on adopting the ISSB.

The 2024 *Corporate Climate Responsibility Report* warns that Scope 3 emission reduction targets continue to be of limited depth, presenting "a key limitation for the integrity of most companies' 2030 climate pledges," even though Scope 3 in many sectors are the most significant emissions category and thus crucial for transforming practices that would allow alignment with a 1.5°C pathway. Despite gradual improvements in addressing these value chain emissions, there are particular concerns about proposed flexibility mechanisms under the Voluntary Carbon Markets Integrity Initiative (VCMI) Claims Code of Practice, launched in November 2023. Under the VCMI Scope 3 flexibility claim, currently in beta version pending refinement over the course of 2024, companies would be able to buy carbon credits for up to 50% of their annual Scope 3 emissions as part of attempts to reach their 2030 targets. Corporate Climate Responsibility Monitor (2024) highlights concerns that rather than a progressive mechanism to incentivise commitment the claim could be used by companies as an "alternative, rather than a complement, to cutting their own emissions, during the "critical decade" for action". This could "...effectively nullify the Scope 3 targets of most of the companies we have analysed for the period up to 2030, leaving them accountable only to their scope 1 and 2 emission targets". Because the 50% flexibility threshold is set according to actual emissions in a given year as opposed to 50% of the emission reductions that would be necessary to reach a company's 2030 target

> "… companies could still qualify to make some (currently undefined) form of claim regarding their scope 3 targets, even if their emissions are double the levels implied by the target trajectory in any given year. In most cases, companies could significantly increase their emissions between 2019 and 2030 and still remain eligible for the claim" (Corporate Climate Responsibility Monitor, 2024).



'AI Mural' by Clarote

Artificial Intelligence and the Cloud

When we think of 'the digital' or 'data' or 'the cloud' we may picture something ethereal, intangible, perhaps even immortal. That goes for AI as well. AI never needs to sleep or eat. It's always there when you open your laptop, indefatigable, ready to chat or translate or paint, ready to do whatever it does. “Hello! How can I assist you today?”

Really, we all know that **AI has a physical basis**. It runs on servers, servers built out of copper, steel, gold, silver, palladium, cobalt and other stuff. It took energy to extract those materials and twist and twirl them into servers. When those servers run, they run on electricity, powered by burning coal or gas, or by wind turbines turning, or sun shimmering on solar panels. They heat up and need cooling down. As Nathan Ensmerger puts it, “there is at least one sense in which the Cloud is more than a metaphor. Cooling a typical data center requires roughly 400,000 gallons of fresh water [about 1.5m litres] daily. A very large center might require as much as 1.7 million gallons [about 6.4m litres].” What counts as a ‘typical’ data centre does vary from to place and over time — many data centres don’t use the kind of cooling Ensmerger is referring to — but temperature is a consideration for all data centres (Zhang et al. 2021; Silva-Llanca 2023; Chang et al. 2024)).

Then there is the **human labour**. There's a lot of human work involved in making and maintaining the machinery of the cloud and organising the data to train AI. This includes building and maintaining the tech, going through training data and hand-labelling it, and the psychological tolls of digital neocolonialism. The humans who do these things need to be housed and clothed and fuelled, and even deserve a little treat now and then. AI systems rely heavily on human labour, from the development and training of models to the downstream use cases that affect how professionals interact with technologies like ChatGPT. For example, academics may be saving time in some ways, while losing time in others (like all those meetings with students about misusing ChatGPT). In the history of automation, this has often happened: human work doesn’t simply get replaced, it gets *transformed*. Sometimes that means the destruction of jobs and the creation of new jobs. But it’s also about the myriad small changes that don’t really add up to a whole job.

How do AI models impact the environment? Let’s focus on the first-order or “direct” impacts. We can start by dividing the carbon emissions into three types. First there is the energy cost of training the model in the first place (essentially “making the AI”). Then there is the energy cost every time the AI responds to a query (“inference”). Amazon Web Services (AWS) recently estimated that the Machine Learning it supports breaks down as 10% for training, 90% for inference. Here we can also include the energy cost associated with storing the model. Finally, there is "embodied carbon," which means the carbon emissions associated with manufacturing, servicing and disposing of the hardware. We could also add “other” as a fourth catch-all category — depending on your methodology, you might want to include various other upstream and downstream carbon emissions. Those are just the carbon emissions impacts: there are also other direct impacts such as water usage, and demand for tech metals and rare earth minerals.

How much carbon pollution do LLMs like ChatGPT cause? We often still don't know, is the short answer—AI developers are not always transparent about how much compute time they used to train their models, or where and when their models were trained. There are similar challenges around estimating the carbon cost of inference, storage, and embodied carbon. However, there has been a lot of independent research recently, and we can certainly get **a general sense of the scale**. One recent study estimated that GPT-3 (the model on which ChatGPT was initially based) consumed 1,287 megawatt hours of electricity and generated 552 tonnes CO2e in training, or "the equivalent of 123 gasoline-powered passenger vehicles driven for one year" (EuroNews.com). The LLMCarbon model estimates GPT-3 produced 553.87 tonnes CO2e in training (Faiz et al. 2024). The *AI Index Report* provides a slightly lower estimate of 500 tonnes of CO2e (*AI Index* 2024). Another estimate suggested training GPT-3 might have consumed at least 700,000 litres of water for cooling (Li et al. 2023). In terms of usage (the carbon cost of inference), one study has suggested that a ChatGPT query burns up four to five times more carbon than an old-fashioned internet search. These figures all include assumptions about where and when these computational processes take place (see Carbon-Aware Computing and Grid-Aware Computing).

LLMCarbon

LLMCarbon is a tool designed to estimate the carbon footprint of large AI models, taking into account both operational and embodied carbon. In a preprint article, Faiz et al. (2024) summarise their end-to-end carbon footprint model for LLMs. 'LLMCarbon encompass the LLM's architectural description, data centre specification, and hardware configuration. To output the LLM's carbon footprint, LLMCarbon employs a series of models, each processing specific input details. LLMCarbon can use the parameter model to determine the LLM's parameter count based on its architectural attributes, or directly accept the LLM's parameter count as input. With the LLM's parameter count and training token count, LLMCarbon calculates the test loss by the neural scaling law (Kaplan et al., 2020), and employs the FLOP model to estimate the volume of FLOPs required for LLM processing. Through the parameter count, LLMCarbon generates the optimal data, tensor, pipeline, and expert parallelism setting. Taking into account the parallelism setting and hardware configuration, LLMCarbon's hardware efficiency model computes the hardware efficiency, representing the real computing throughput divided by the peak computing throughput. Utilising data centre details, hardware efficiency, and FLOP count, LLMCarbon applies the operational carbon model to derive the LLM's operational carbon footprint. Similarly, by considering the hardware configuration, LLMCarbon's embodied carbon model yields the LLAMA's embodied carbon footprint. The overall carbon footprint of the LLM is then computed by summing both the operational and embodied carbon footprints.' The source code is available on GitHub.

We may say, but perhaps this training cost is not *that* large? Relative to the social and cultural impact of GPT-3 and other LLMs that have come after it? Besides, aren't cloud giants like Google and Microsoft leading the way on net zero, using 100% renewable energy and so on? Well, these companies' green reputations have taken a distinct hit in the past couple years due to the significant energy demands of AI and greater public scrutiny of carbon accounting methodology. As we explore elsewhere in this report, perhaps these green reputations were not entirely well deserved in the first place.

Today, a lot seems to hinge on future efficiencies which AI is predicted to unlock—**the indirect positive impacts of AI on the environment**, which advocates say will easily compensate for all of AI's negative environmental impacts. But in the conversation around AI for sustainability, it can be difficult to sort fact from fiction, and to clearly identify the risks and uncertainties.

Currently most AI developers lack the incentives and the infrastructure to manage and disclose environmental impacts. Growing pressure for corporate accountability might push them to disclose more details in the future. Work continues on tools to estimate, measure, and optimise the carbon impact of LLMs; Faiz et al. (2024) offer the tool *LLMCarbon*, an end-to-end carbon footprint projection model. Green Coding AI, from Green Coding Solutions, is another interesting and useful tool.

Understanding that **not all AI is the same** can help in estimating environmental impacts. Just as different vehicles have different fuel requirements, AI models vary significantly in their energy consumption and carbon footprints. As we emphasise in several places in this report, a useful step would be to get used to distinguishing between different types of AI. In their book *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*, Arvind Narayanan and Sayash Kapoor offer this analogy for current conversations about AI:

> Imagine an alternate universe in which people don't have words for different forms of transportation — only the collective noun "vehicle." They use that word to refer to cars, buses, bikes, spacecraft, and all other ways of getting from place A to place B. Conversations in this world are confusing.
> (Narayanan and Kapoor 2024)

The Current Carbon Footprint of AI: A Microsoft Case Study

Within the next year, Microsoft is likely to exceed the carbon budget it committed to in 2020, which was supposed to last until 2030 and take the company to net zero. AI has been offered as a reason for this missed target, and Microsoft is confident that the benefits far outweigh the costs.

But does Microsoft really have good data about the direct climate impacts of AI? This is a complicated question. One indication might be found in a recent article in *Nature* (Luers et al., 2024). This article was authored by a mixture of Microsoft employees, researchers funded by Microsoft, and independent researchers. *Nature* is one of the most prestigious science journals in the world. So it is concerning that the article's claim, that AI today represents "about 0.01%" of global greenhouse gas emissions, appears to have quite shaky foundations. *Nature* is a journal where researchers, policymakers, industry, and societal stakeholders can usually turn for a solid stat that they can quote with confidence.

However, to give the article its due, its main purpose is not to offer a robust estimate. It is focused on making an intriguing pitch for more **scenario based AI governance**, underpinned by collaboration between the AI community and the climate modelling community. We explore that aspect the next section. Nonetheless, such a proposal should still be informed by robust evidence of AI's environmental impacts to date. So in this section, we **examine that 0.01% figure**. How was it calculated? Is it as reliable as its appearance in *Nature* might lead many readers to expect?

> AI processors installed in 2023 consume 7–11 terawatt hours (TWh) of electricity annually, which is about 0.04% of global electricity use […] That is less than for cryptocurrency mining (100–150 TWh) and conventional data centres plus data-transmission networks (500–700 TWh), which together accounted for 2.4–3.3% of global electricity demand in 2022, according to the International Energy Agency (IEA). […] in terms of total global greenhouse-gas emissions, we calculate that AI today is responsible for about 0.01%, on the basis of IEA assessments showing that data centres and transmission networks together account for about 0.6% (see go.nature.com/3q7e6pv). (Luers et al. 2024)

If we are interpreting this correctly, 0.01% comes from 7-11 TWh as a proportion of 500-700 TWh, which is 1% to 2.2%. Then if we multiply a number near the middle of that range, let's say 1.7%, by IEA's estimate of 0.6%—that is, how much data centres and network transmissions contribute to global GHG emissions overall—we get a result that rounds to 0.01%. The 7-11 TWh hours figure appears to relate to an estimate of how many AI servers NVIDIA shipped in 2023, as described in de Vries (2023). If that interpretation is correct, then there are five potential problems.[1]

(1) De Vries (2023) **actually gives slightly different figures**. The article mentions 5.7-8.9 TWh, not 7–11 TWh. This probably represents the authors of the Microsoft-led *Nature* article deciding to revise the range upward (5.7 to 7 and 8.9 to 11 are both

increases of about 125%), perhaps to compensate for some of the factors mentioned below. However, no assumptions or methodology is shown. Here is the relevant section from de Vries (2023):

> Given its estimated 95% market share in 2023, NVIDIA leads the AI servers market. The company is expected to deliver 100,000 of its AI servers in 2023. […] If operating at full capacity (i.e., 6.5 kW for NVIDIA's DGX A100 servers and 10.2 kW for DGX H100 servers), these servers would have a combined power demand of 650–1,020 MW. On an annual basis, these servers could consume up to 5.7–8.9 TWh of electricity.

(2) Importantly, de Vries (2023) also **suggests steep growth to 85.4–134.0 TWh in 2027**. The Microsoft-led *Nature* article does include this point, but does not see it as too troubling: "AI should not lead directly to large, near-term increases in greenhouse-gas emissions." In the past we have seen some inflated data centre energy consumption projections (cf. Mytton 2024; Masanet et al. 2020), so Luers et al. might potentially have made a case for de Vries's projection being unduly pessimistic. However, it is unclear if they actually are rejecting de Vries's projection, or if they do not consider growth to 85.4–134.0 TWh in 2027 to constitute "large, near-term increases in greenhouse-gas emissions."

(3) Second, de Vries only **offers one non-peer-reviewed source (*MarketWatch*) for the figure of 100,000 AI servers.** The range of 5.7-8.9 TWh is calculated based on a statement by a stock market analyst, Vijay Rakesh, that he expected Nvidia to deliver 100,000 AI servers in 2023. The lower end of 5.7 TWh is based on the power consumption of 100,000 DGX A100 servers, and the upper end of 8.9 TWh is based on the power consumption of 100,000 DGX H100 servers. We were unable to find any record of the methodology which led to the figure of 100,000 servers, which was reported in several articles in the financial press, one of which is cited by de Vries. 100,000 is also a striking, round number: without further explanation, it seems reasonable to question whether it was ever intended as a precise estimate, or merely to suggest an order of magnitude.

Rakesh is of course mostly concerned with Nvidia's stock price — that is his remit — and his evidence and methodology is not readily publicly available. The context of a science journal may well mislead readers into supposing a more robust and transparent methodology has been used. Moreover, as the *Nature* article appeared in 2024, presumably Nvidia could also have confirmed or revised that 100,000 figure — this information may be publicly available, or Nvidia may be willing to share it with Microsoft researchers. It is also worth noting that for Rakesh, this projected growth is positive, a reason to invest in Nvidia.

In general, whether or not you agree with his assumptions and results, De Vries is admirable for using different plausible methodologies, and for showing the working clearly. Readers can then assess the strengths and weaknesses of his claims. But this nuance is not reflected in the use of his figures in the Microsoft-led *Nature* article.

(4) Still more significantly, whether we use 7–11 or 5.7–8.9 TWh, **energy usage of NVIDIA servers installed in 2023 is not global AI energy usage**. It excludes:

- AI trained or deployed on non-Nvidia servers. Nvidia dominates by a long way, but there are also AI-focused offerings from Google, AMD, Intel, more generic GPUs, and even CPUs.
- AI trained or deployed on any Nvidia servers installed before 2023. Just for example, de Vries also suggests that Google AI alone used around 2TWh in 2021.

De Vries explicitly mentions broader AI-related electricity:

> [...] growth in AI-related electricity consumption will originate not only from new high-performance GPUs such as NVIDIA's A100 and H100 GPUs but also from more generic GPUs. It is already the case that former cryptocurrency miners using such GPUs have started to repurpose their computing power for AI-related tasks [...]

An alarming possibility now presents itself. Has there been **fundamental confusion about the phrase "AI processors installed in 2023"**? The phrase can be interpreted in two very different ways. If used as the basis for AI's global greenhouse gas footprint, then it should mean **total electricity consumption during 2023 by all AI processors**, irrespective of their installation date. The Microsoft-led *Nature* article seems to intend this sense. However, the same phrase could also mean **the annual electricity consumption of only those AI processors installed in the year 2023**. The De Vries paper supports only this second sense. We can assume there will be a substantial difference between these two amounts.

(5) Finally, the de Vries figures **exclude energy used for cooling**, although the Microsoft-led *Nature* article implies cooling is included.

> The direct impacts of AI on climate so far are relatively small. AI operations for large models require millions of specialized processors in dedicated data centres with powerful cooling systems. AI processors installed in 2023 consume 7–11 terawatt hours (TWh) of electricity annually, which is about 0.04% of global electricity use
> (Luers et al. 2024)

In addition to these problems, the Microsoft-led article omits more nebulous aspects of energy use, as well as factors that go beyond energy use.

- **Utilisation.** This can cut both ways: servers are not likely to be operating at full capacity, which will **mitigate the carbon impact**. On the other hand, low utilisation is inefficient—a server that is not running anything is still using up energy. The sustainability case for cloud migration often relies on high utilisation (see The Cloud and the Climate).
- **Embodied carbon of AI infrastructure.** Not just the electricity directly used by the servers, but the emissions associated with manufacturing them, manufacturing the data centre infrastructure, and so on. For example, Faiz et al. (2024) demonstrate that embodied carbon is a significant component of LLMs' carbon impacts.

- **Capacity replacement vs. capacity growth.** More energy-efficient servers can replace older less efficient models, and/or can run alongside them. It may also change *the way* that existing servers are used. The impacts of efficiency gains from new hardware on overall capacity growth, hardware refresh rates, utilisation, and embodied carbon (including disposal) is tricky to estimate, but important.

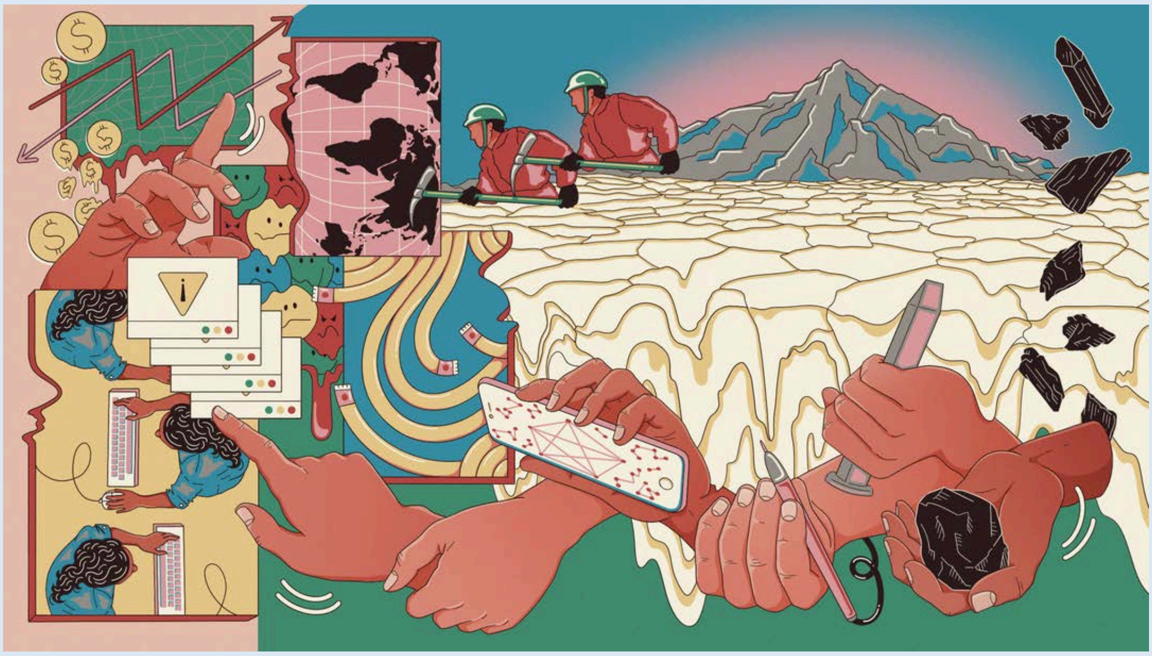
- **The embodied carbon of user devices attributable to AI.** This is a very tricky one. How can you reliably attribute some proportion of the manufacture, transport, and disposal of new phones, laptops and other devices to the AI software they run or access? However, we *do* need to try, because it is clear that the touted AI revolution is very much batting for planned obsolescence ("AI is here! Time to get a new phone!"), not for regenerative design and right to repair, or even just circular economy.

- **Alternative approaches to the IEA assessment of data centre GHG emissions**—this is tricky too. Renewable energy procurement is complicated (RECs, PPAs, 24/7 hourly matching, etc.: see the section on Green Data Centres).

The IEA figure of 0.6% is serviceable but not perfect. It is from 2020—so not reflective of rapid recent AI-driven growth.

The IEA also offers the alternative figure of 0.9% for energy-related GHG emissions specifically. Choosing the 0.9% figure rather than the 0.6% figure would have produced something like 0.015% rather than 0.01%, and would mean you could clearly say, "excluding embodied carbon" (as it stands, embodied carbon is included in one part of the calculation but not the other).

While the figure of 0.01% is offered as an estimate, **one would expect a more careful estimate**. There is enough transparency to be able to have identified the problems and queries mentioned above, although even greater transparency would certainly have been possible.



'Labour / Resources' by Clarote

Modelling the Future of AI and the Climate?

Do we have the models we need to govern AI effectively? As described in the previous section, a recent Microsoft-led article in *Nature* estimates the current carbon costs of AI based on several questionable assumptions. Having set things up in this way—*nothing to panic about now, but we should prepare for the future*—the *Nature* article then proposes funding a **consortium**, to do extensive **modelling** around the sustainability of AI, combining risks and opportunities, costs and benefits. (Interestingly, there is currently a bill before the US Senate which if it passes will mandate something broadly along these lines). The authors of the *Nature* article ask:

> How will future AI technologies develop? How will their expansion affect the global economy? And how will this affect decarbonization? Researchers simply don't know; it's too early to tell. (Luers et al. 2024)

The idea is commendable in several respects. **A more systemic, holistic approach to AI governance is needed**. Scenario-based thinking may shift the sense that there is only one future possible, and our only choices are how fast we accept it, and whether we individually benefit or lose out. However, there are problems with the proposal's details. Some problems stem from undue confidence in unsure estimates of AI's climate impacts, as described above. There is also not enough discrimination between different types of AI. Other problems stem from neglecting the importance of cumulative emissions, as opposed to target dates. It is not *entirely* too early to tell how AI is affecting decarbonisation, since at the corporate level this has been disclosed in the 2024 sustainability reports of Amazon, Google, and Microsoft. Because of AI, as Microsoft's Brad Smith memorably put it, the moon is five times further away than it once was:

> Now to meet its goals, the software giant will have to make serious progress very quickly in gaining access to green steel and concrete and less carbon-intensive chips, said Brad Smith, president of Microsoft, in an exclusive interview with Bloomberg Green. "In 2020, we unveiled what we called our carbon moonshot. That was before the explosion in artificial intelligence," he said. "So in many ways the moon is five times as far away as it was in 2020, if you just think of our own forecast for the expansion of AI and its electrical needs." (Bloomberg 2024)

When the authors say it is "too early to tell," of course, they mean something more holistic: not just direct negative impacts, but the total balance of direct and indirect positive and negative impacts. This phrase suggests we should not interfere with existing industry trajectories too much, not rush to any hasty policy decisions or strategic course-changes based on partial data. We should wait, watch, gather evidence, iterate our policies and corporate strategies, and eventually get it right. Favouring this wait-and-watch approach lines up with the tendency within climate conversations — a perilous tendency, to which we are all prone — **to think of climate change and climate transition as**

something perpetually in the future. Tech is especially used to thinking in this iterative way (we could say it is an important part of tech's sociotechnical imaginary). But this approach is not fit for all aspects of climate transition. We need to emphasise strong policy and practice on AI and the climate today, based on the best information available today, and building in risks and uncertainties that we can identify today.

In this connection, the proposal that AI should be integrated into the IPCC's Shared Socioeconomic Pathways (SSPs) and Integrated Assessment Modelling (IAMs) is oddly expressed—the SSPs have always included technological change, while being cautious to exclude speculative breakthroughs that may or may not happen. Some may say that they have been too cautious: but the crucial point is that **there is *already* a community hard at work on the next generation of scenarios for AR7** (including researchers with AI expertise). The Microsoft-led *Nature* article does not reflect much awareness of such work:

> AI should be integrated into these pathways, along with the global shocks and technological breakthroughs that might accompany it. This would require major work, including incorporating expertise from the AI community, rethinking each of the pathway narratives and exploring whether new ones need to be added. Could AI take the world to a more radically green future, or a more dystopian one? What factors define those outcomes? How plausible are they? Scenarios can help to narrow down answers.
> (Luers et al. 2024)

'The AI community' and 'the climate modelling community' are abstractions which may often be useful. In this case, however, it would be helpful to have clarity on what type of AI expertise Luers et al. propose is underrepresented within the existing climate modelling community. Certainly, one message that comes through most clearly from the climate modelling community is **the need for meaningful inclusion of the Global South and Newly Industrialised Countries**, as well as co-production with a range of societal stakeholders (not just academia and industry). Some go further and point to the need not just for inclusion, but leadership.

The proposal in *Nature*, by contrast, appears more like the kind of initiative where well-resourced research institutions and industry players concentrated in the Global North lead the agenda, with some co-production used to enrich outcomes and mitigate risks, but without meaningful leadership from the Global South, Newly Industrialised Countries, or from communities most affected. Microsoft's recent *Accelerating AI with Sustainability playbook* does at least prominently note the need for much better inclusivity.

Let's take a step back. Let's see if we can clarify the proposal, freed from the constraints of professional tact. **Perhaps the implied argument is something like this:**

1) Microsoft sustainability leadership notes the technological breakthroughs in AI since 2020. They believe that these have been so profound, that it is now actually in the climate's best interest to renegotiate Microsoft's 2020 climate pledge. **They believe that**

it may now be prudent to delay emissions abatement compared with the plan (as is actually already occurring in practice). They think this may be the correct approach, because the extra carbon emissions will eventually be outweighed by the carbon emissions savings that they enable.

2) However, Microsoft's sustainability leadership also recognise the **strong conflict of interest entailed in making such a judgment**. They also recognise that the delayed decarbonisation is already happening in practice, and probably would be happening whether or not it is compensated for by accelerated decarbonisation elsewhere, and Microsoft is therefore unlikely to be trusted as authoritative on the subject.
3) They therefore invite the climate modelling community to co-produce the necessary evidence. Since the extra carbon emissions would be on Microsoft's balance sheet, whereas the carbon savings would mostly not be, this kind of co-production is also needed to ensure Microsoft's social licence to operate.

This framing may help to clarify what is at stake. If this is what the article argues, is it credible? If it is not what the article argues, then what are the key points of difference?

Assuming it is broadly correct, one key limitation is the systematic **erasure of the IPCC's timescales**. Cumulative emissions are what matter: every day that goes by, the climate is being heated. More than halving global emissions by 2030 is as crucial to limiting global warming to Paris Agreement targets as the eventual date of achieving net zero. In 2020, Microsoft indicated its understanding of this by pledging to halve its emissions by 2030, and to scale up carbon removals to be net negative by 2030. Yet the Microsoft-led *Nature* article cites de Vries's suggestion that AI energy consumption of 85.4-134 TWh is to be expected by 2027, more than tenfold 7–11 TWh. Do the authors believe this is credible? And do they expect their proposed actions to make a significant difference to AI-related carbon in 2024, 2025, 2026, 2027, 2028, 2029, and 2030? "Researchers simply don't know; it's too early to tell" is not an adequate answer here.

The cloud giants are all major purchasers of renewable energy. Yet in 2024, Microsoft is already on the brink of exhausting the carbon budget it committed to in 2020, and explicitly identifying AI as the reason. Reaching net zero next year would be a miracle, and reaching it by 2030 is looking increasingly unlikely. The other cloud giants are similarly failing to decarbonise. Against this background, the Microsoft-led *Nature* article is a proposal to pull together funding, to set up a consortium, to do challenging research, to develop new models, to inform new policy and practice, with no particular timeframe specified. The article suggests simultaneously that it is too soon to tell what the impact of AI on net zero will be, and that AI will not lead to a major near term increase in greenhouse gas emissions. The main article that it cites to support this source, de Vries (2023), predicts rapid growth by 2027. We may well ask whether all this is a serious proposal, a delay tactic, or a symptom of wishful thinking.

Sustainable AI Innovation

There are also **hopeful signs.** Scenarios and modelling are important tools; acknowledging the expertise of the IPCC and climate modelling community is crucial; it is a breath of fresh air that progressive elements within Microsoft are considering how their activity relates to the systemic picture.

Awareness of cloud carbon is growing fast. Movements like green computing, minimal computing, digital sufficiency, digital sobriety, and organisations both big (like the Green Software Foundation, whose members include Intel and Microsoft) and small (the Digital Humanities Climate Coalition), offer routes to learning more about digital decarbonisation even for less technical audiences.

Many businesses already have quite sophisticated approaches to cloud optimisation (FinOps), capable of considering more than just a single bottom line. **FinOps** is not as sustainability focused as it should be, but does appear to have good potential to pivot. Some businesses are now starting to adopt **GreenOps** (see "FinOps and GreenOps" above).

Even AWS now offers clients a carbon emissions dashboard, which is a start, and it has promised to improve the data it provides. Partly in anticipation of future reporting requirements, companies like Greenpixie offer cloud usage data enrichment to Managed Service Providers and to enterprises with large cloud spends. There are also many open source projects, such as Cloud Carbon Footprint, which can help with estimating cloud carbon.

Policy, standards, development environments, and diagnostics are evolving. Some of this work is being conducted under the banners of **Responsible AI** and **AI Ethics.** There are Algorithmic Impact Assessment tools and AI Impact Assessment tools: to date these have mostly focused on things like bias and explainability, but we are likely to see them include environmental impacts much more in the future. There is a proposal for an Energy Star for AI eco-label.

There are also existing assessment tools like Environmental Impact Assessments, Life Cycle Assessments, and Social (Cultural) Impact Assessments which have applicability to AI. Real progress plus some greenwashing is associated with concepts such as circular economy, nature positive economy, and regenerative economy.

Meanwhile, AI researchers are constantly looking for ways to make models more efficient. Smaller ML models, non-ML approaches, model switching, and techniques such as scheduling, pruning, architecture search, quantization, and knowledge distillation, promise to improve the sustainability of AI. Quantization, for instance, involves using less precision in computations, thereby reducing resource use. Knowledge distillation is where a larger model trains a smaller one to achieve similar performance. There is some interesting work around routing queries to the most appropriate model within an ensemble of models, building in quality and sustainability considerations (Hoffmann and Majuntke 2024).

Castro (2024) points out that some larger models are being trained with less energy than their petite predecessors:

> [...] while larger models generally require more energy usage than smaller ones do, the exact figures vary significantly across different AI models. For example, researchers estimate that training GPT-3—the 175 billion parameter AI model used in the popular ChatGPT application—created 552 tCO2 emissions, but comparable AI models including OPT (a 175 billion parameter AI model created by Meta) and Gopher (a 280 billion parameter AI model created by Google) have significantly smaller carbon footprints.

Efforts are underway to build more Small Language Models, and neuromorphic computing architectures (see e.g. Cronk 2024). There is growing interest in **active inference** approaches to AI, which are distinct from Machine Learning, and have the potential to create much leaner and more interpretable models.

Lucciano et al. (2024) point out that the ambition of generality (as opposed to traditional hand-coded software or smaller models fine-tuned for specific tasks) comes at considerable environmental costs. Their study finds also that:

> 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

A 2023 literature review gives a sense of interest in greener AI:

> there has been a significant growth in Green AI publications—76% of the papers have been published since 2020. The most popular topics revolve around monitoring, hyperparameter tuning, deployment, and model benchmarking. We also highlight other emerging topics that might lead to interesting solutions—namely, *Data Centric Green AI*, *Precision/Energy Trade-off* analysis. The current body of research has already showcased promising results with energy savings from 13% up to 115%. Still, most of the existing work focuses on the training stage of the AI model. Moreover, we observe that there is little involvement of the industry (23%) and that most studies revolve around laboratory experiments. (Verdecia 2023)

Nvidia and other AI hardware manufacturers are responding to concerns with greater focus on energy efficiency.

Yet more innovations relate to **the data centres themselves**, such as using low carbon cement, innovative cooling techniques, or pumping the warmth from the servers into the local heating infrastructure (already well established in Scandinavian economies particularly; see Data Centre Heat Reuse).

> The most "eco-friendly" datacenters on the market today can be designed to have a synthetic white rubber roof, white paint to enhance albedo, and carpet with low counts of volatile organic compounds. Countertops and server racks can be made from recycled products, and mechanical and electrical systems can be set to optimal efficiency. Natural light can be used along with energy-efficient windows, skylights, and sky-tubes. A nonprofit group the "Green Grid" even publishes white papers on how to propagate the best energy-efficiency practices in datacenter design and construction.
> (Sovacool et al. 2022)

Other potential innovation relates to improving **energy storage,** to allow renewable energy to be used more flexibly. Just for example, Liang et al. (2024) propose developing greener data centres by configuring photovoltaic power generation and compressed air energy storage systems. There is also interest in expanding the use of **Direct Current** instead of Alternating Current. "Data centres are now looking at using microgrids for power. That means drawing on-site energy directly from sources such as fuel cells and solar panels. As it turns out, those sources often conveniently produce direct current" (Judge 2024). Other innovations relate to back-up power. Energy efficiencies can be unlocked by bypassing the Uninterruptible Power Supply (UPS) under normal operating conditions, or installing modular UPS systems (Sovacool et al. 2022).

So there are many promising innovations. However, policy has yet to catch up with **risks related to AI's fast growth.** Private investment in generative AI has skyrocketed, from $0.84bn in 2019 to $25.23bn in 2023 (*AI Index Report* 2024). Data centre infrastructure has been expanding rapidly (ESG News 2024). Technical efficiencies may be wiped away as big players in the AI 'race' or 'war' rush to embed AI into everyday digital tools, to make it a default dimension of how we work, play, communicate, and create. In terms of usage patterns, AI apps like ChatGPT and Gemini are already more than a search application: users will converse with them, in a way they might not with the Bing or Google search bars. A late 2022 guesstimate had ChatGPT responding to around 10 million prompts per day. There has also been a trend toward bigger and bigger AI models. ChatGPT is running on GPT-3.5, GPT-4, and its variants. We are not told how much energy it took to train GPT-4. In money terms, 'more than $100 million' is what we are hearing. Respectable rumours are rife that it might be quite a lot more. AI companies no longer like to share the details, but AI research organisation Epoch AI has provided some estimates. Their cost estimate for training GPT-4 is $78m, and for Gemini Ultra $191m. There is simultaneously an interest in achieving good results with smaller models too, of course. However, since the surge of interest in GenAI, we have seen the cloud giants fail to make timely progress on their climate pledges.

In summer 2024, there may be some signs that this AI goldrush is slowing, although we cannot forecast what the future may hold; AI certainly still seems to be a geopolitical priority for the US, China and other states. There has been plenty of pushback against AI overreach, hype and harms, including lawsuits in the creative industries. Rich Gibbons of Synyega comments: “Perhaps the only real way for organisations such as Microsoft and Google to reduce their emissions will be for the majority of customers to reject these new GenAI services until they are absolutely critical” (quoted in Donnelly 2024).

If there is some kind of AI crash or course correction, it is unclear how this would translate to changes in infrastructural expansion. Other emergent technologies such as VR are also computationally intensive, and the cloud giants “have an incentive to ensure whatever follows generative AI will similarly ramp up the amount of computation our societies will collectively require, while ensuring we’re dependent on them to provide it” (Marx 2024).

More broadly, we are frequently reminded that AI can be used to fight climate change, for instance in optimising renewable energy use, and in various AI-powered approaches to improving the performance of cloud computing itself. There are good reasons to be cautious here: we should always distinguish what specific types of AI and/or data-driven methods are being used to combat climate change, and also whether they are contributing to climate mitigation or to climate adaptation. Reports and thought leadership pieces often misleadingly pair the problems posed by AI for mitigation, with the benefits of AI for climate adaptation. For examples of analysis focused on AI for mitigation specifically, see the sidebar, as well as Microsoft/PwC 2019; Degot et al. 2021; Kaack et al. 2022.

Why insist on these distinctions? After all, within AI research, a model or technique may be developed to do one thing, and turn out to be applicable to do something else entirely. However, that shouldn't mean carte blanche to all AI R&D just in case it may end up benefiting the climate. For one thing, there may be potential for beneficial spillovers, but there is also potential for harmful spillovers. “ML has also been applied in ways that may make climate goals harder to achieve” (Kaack et al. 2022). This is not only because of the climate impacts of AI systems themselves, but the potential to enable or to encourage carbon intensive activities. In the next section, we shine further light on AI for sustainability.

Eco-labels

Energy Star is a venerable programme of the US Environmental Protection Agency (EPA), scoring the energy efficiency of electronic goods. EPEAT (Electronic Product Environmental Assessment Tool) is a global ratings system that covers aspects like energy consumption and recyclability. TCO Certified currently covers product and sustainability information, socially responsible manufacturing, environmentally responsible manufacturing, user health and safety, product performance, product lifetime extension, reduction of hazardous substances, material recovery. It is a third-party certification that aspires to be independent of IT buyers and vendors.

There are a huge variety of smaller and more niche eco-labels (or similar). For example, in the EU the WEEE label just lets you know you need to dispose of whatever it is with a registered waste handler. IEEE 1680 Standards provide criteria for the design of environmentally friendly electronic products. There's Cradle to Cradle Certified, Blue Angel (Germany), Carbon Trust Standard, Nordic Swan eco-label. There is LEED for the data centre buildings themselves.

The International Standards Organisation distinguishes between three broad types of labels. Type I (ISO 14024) labels are part of a voluntary program that requires an independent third party to evaluate a product's life cycle impacts. Type II (ISO 14021) represents self-declared environmental claims by manufacturers or sellers, without the need for external auditing. Type III (ISO/TR 14025) is about some environmental product declaration that provides quantifiable information on a product's life cycle impacts. This label presents useful data, but it kind of leaves the evaluation up to the customer or other stakeholder. What other features might we consider when mapping IT eco-labels, developing or assessing a specific eco-label, or even building a tool for selecting appropriate eco-labels? A few ideas include:

- Scope, in the sense of **types of hardware or software** it applies to.
- Whether or not the eco-label operates on a **tiered system**, indicating different levels of compliance. Furthermore, how *many* tiers, and how big a jump it is from each to the next.
- Whether there's a requirement for **independent verification**.
- If there is such a requirement, then the **associated cost** of getting certified. Some climate-aligned goods and services may be the innovations and/or passion projects of smaller companies with limited budgets, or open source communities.
- If there is a cost associated with getting certified, are there **discounts** or other support available for **social impact**?
- If there isn't independent verification, then is this a label oriented to **self-reporting**

only, or does it lend itself to use by **independent parties**? For example, is it a checklist that a procurement team or some other stakeholder could apply?
- An assessment of the actual **level of independence** of the certifier. How capable is the certifier of saying “no”?
- **Updates and latency**. How often is the label revised? How often does the certified product need to be assessed?
- Is there a system that ensures **accountability**, and how often have labels been **withdrawn?**
- What is the relationship between **who maintains the eco-label and who certifies it**.
- How **voluntary or mandatory** is the eco-label?
- Obviously eco-label **scope** in the sense of what impacts, costs, risks etc. it is seeking to avoid or minimise — such as energy efficiency, carbon intensity, water usage, minimization of hazardous substances, various supply chain considerations, resource conservation, and broader social impact — can vary widely.
- The geographic scope and **recognition** of the eco-label, whether it is recognized locally, regionally, or internationally, what types of stakeholders rely on it and for what purposes.
- The process of obtaining the eco-label, including the average **length of time** from application to certification, the **success ratio**, etc.
- **How was it created?** What kind of process was followed to ensure robust stakeholder engagement and co-production, and other appropriate forms of expertise?

AI for Climate and Sustainability

AI plausibly has a significant role to play in achieving sustainability goals, including climate mitigation and climate adaptation (Rolnick et al. 2022; Cheng et al. 2024). However, high levels of AI hype can make it difficult to sort signal from noise, and identify the genuinely promising use cases where AI can help to meet sustainability challenges.

Claims about the potentials of AI are often troublingly vague. AI has shown us it can do astonishing things in certain domains, so there is a risk that we overestimate what it can deliver across all domains. As one literature review puts it, there are questions about “whether AI contributes positively to sustainability or inadvertently accelerates resource depletion and reinforces biases” (Tripothi et al. 2024).

Rhetoric around AI is also often polarised, presenting AI as something that must be embraced or rejected in one fell swoop, instead of allowing useful distinctions—like more permissive regulation and generous incentives for sustainability-focused AI, and/or for less resource-intensive AI.

Vague promises and vaporware appear to be contributing to the legitimacy of an AI intensive future, one which poses clear dangers to climate and environment. Reviewing national strategies, Perucica and Andjelkovic (2022) suggest:

> while the majority of countries recognise the significance of AI to the environment, only a few go beyond the well-known statement that AI has the potential to help solve ecological challenges.

In discussions of AI and the climate, you will often see the point being made that AI poses risks to the climate, but that it can also contribute to addressing climate change. For example, a punchy, freewheeling article in *Fortune*, lampooning "AI's Bizarro World" still includes this proviso:

> And the potential for advanced AI systems to help tackle climate change issues—to predict weather, identify pollution, or improve agriculture, for example—is real. In addition, the massive costs of developing and running sophisticated AI models will likely continue to put pressure on companies to make them more energy-efficient. (Goldman 2024)

Similarly, Stanford University's *AI Index Report 2024*, after collating useful data on the carbon and water use impacts of AI foundation models, adds the following:

> Despite the widely recognized environmental costs of training AI systems, AI can contribute positively to environmental sustainability. Figure 2.13.5 showcases a variety of recent cases where AI supports environmental efforts. [...] These applications include enhancing thermal energy system management, improving pest control strategies, and boosting urban air quality. (*AI Index Report 2024*)

This is a highly regarded report in the AI community. It would be alarming to find that it is using unreliable sources or methodologies. Figure 2.13.5 in the report tabulates some use cases: management of thermal energy storage systems (Olabi et al. 2023), improving waste management (Fang et al. 2023),[33] more efficiently cooling buildings (Luo et al. 2022), improving pest management (Rustia et al. 2022), enhancing urban air quality (Shams et al. 2021). There is a striking, unremarked disconnect between the two sides of the equation here: one focusing mostly on the risks that AI poses to **climate change mitigation** (limiting global warming to well below 2.0 degrees), the other on the benefits it might deliver for **climate change adaptation** (surviving and thriving in this hotter and more volatile world).

The *Stanford AI Index 2024 Report* also prominently signposts a literature review, 'Artificial intelligence-based solutions for climate change: A review' (Chen et al. 2023): an article which has apparently been partly AI generated, and is **filled with unsubstantiated claims about the climate benefits of AI**. See 'A Literature Review Reviewed.'

---

[33] This one doesn't actually appear in the Works Cited, but given this description, this appears to be the right paper.

AI for Climate Mitigation?

There is also R&D that focuses clearly on **mitigation,** that is, reducing net carbon emissions (e.g. Kaack et al. 2022). A 2021 report by Boston Consulting Group, commissioned by Google, **estimated that AI could reduce GHG emissions by 5-10%** across the economy by 2030 (Degot et al. 2021). The report mentions monitoring emissions, predicting emissions, and reducing emissions.

This is welcome news, and the two specific case studies (one a steel producer, one an oil and gas company) are especially welcome. But the details are again thin, and many questions still need to be addressed.

First, the BCG report mentions having calculated the figure of 5-10%. **How did they calculate this figure?** In the absence of any data or methodology, only a reference to "our experience with clients," this is not a credible source to quote without a lot of heavy caveats.

Second, the two BCG case studies describe emissions reductions of 3% (the steel producer) and 1-1.5% (the oil and gas company). **Why are these outcomes lower** than the 5-10% that BCG forecast for the economy as a whole? It is mentioned that some optimisation initiatives had already taken place, but more detail could be very useful here. The BCG report mentions the 5-10% figure in relation to the steel industry and the economy as a whole — is this a coincidence, or are both derived from the same analysis?

Then there is the question of **how much of the improvement is truly attributable to AI, and to what types of AI.** The case study about the steel producer describes an extensive network of data-collecting sensors as part of the project. It therefore needs to be confirmed whether the 3% emissions reduction is fully attributable to the additional benefit of AI, and not partly to the collection and analysis of the data and implementation of new controls as such. As Kaack et al. (2022) point out, "When assessing the GHG emissions impacts of ML, it is important to compare ML approaches to alternatives. Such alternatives are not constrained to other ML models; they can also be other types of analytics approaches that fulfil the same purpose, or can be human decision-making." Commendably, the BCG report specifies the new AI capabilities it is referring to, such as generating synthetic data to fill gaps. So it should be possible to **benchmark** the improvement against a hypothetical baseline optimisation project — a 6+ months project with comparable financial resourcing and staff, in this case likely involving participation of over 200 staff, but without the use of new AI technology. Is this what has been done to produce the figure of 3%?

Is there a reason why BCG does not actually **name the companies** involved in the case studies? Perhaps it is simply easier not to involve legal and marketing teams and whatever extra paperwork and approvals this might require. If so, this is fair enough. However, it would be good to be able to cross-reference with other publicly available information. It does seem likely that the steel producer mentioned is **JSW Steel**, as its market cap was about $8 billion in 2021, and it was working with BCG on Project SEED. From JSW's integrated reports and new sustainability report alone, there are clearly the makings of a fascinating, granular case study, one which reflects openly on challenges and uncertainties.

JSW Steel's 2022/2023 integrated report discusses the BCG project. It provides relatively extensive coverage of both digital innovation, and of sustainability. AI is mentioned in relation to optimisation initiatives. Modelling-driven caster pull-out optimisation is mentioned, for example. At the Vijayanagar plant, Machine Learning has been deployed to help reduce defects in properties during annealing (heat treatment to improve ductility and reduce brittleness). However, such optimisations are seldom explicitly tied to sustainability in the report, and instead cost, time, and/or quality remain at the forefront.

Separately, the report contains a very interesting **inventory of specific decarbonisation interventions, together with estimated carbon savings**, e.g. 'Increased hot charging percentage in HSM by ~6% resulting in reduction of gaseous fuel rate by ~11%,' 'Reduced solid fuel rate in Corex by ~5%,' 'Replacement of old boilers resulting in an increase in steam generation rate by ~380%,' 'Optimising Centralised Gas Mixing Station's network to maximise inhouse power generation,' and many others. **Can some of these interventions be attributed to the AI-driven aspects of Project SEED?** If so, this would give more credibility to the claim that any carbon reductions are AI-driven. It is worth emphasising, however, that many of these interventions involved **investment or other expenditures beyond the cost of any AI systems**. When considering the claim that AI has enabled an emissions reduction of 3% (or could enable a reduction of 5-10%), we must also explore the extent to which AI-driven adjustments were "pure" optimisations (Pareto efficient), and to what extent they involved trade-offs, e.g. increased risks, training, workloads, skill requirements, cashflow pressures, or costs.

More recently, JSW Steel's 2024 Sustainability Report does not mention AI. Overall, for scopes 1 and 2, **JSW Steel's absolute carbon emissions have increased** from 2020 to 2024, while carbon intensity has shown some improvement, but not steady improvement (from 2.49 to 2.5 to 2.36 to 2.44 tCO_2/tcs).

BCG's 5-10% figure was quoted recently again (August 2024) in the *New York Times* (Lohr 2024). The framing was a familiar one. There are concerns about the carbon footprint of AI, but lots of sustainability benefits too, if we play our cards right.

Of course it is very reasonable to include the pros alongside the cons. But it is worth reflecting on how this works rhetorically. If the reader is not equipped to weigh the one against the other, or given a clear signal about the orders of magnitude involved, they tend to feel similar. We need to ask: **How comparable are these pros and cons?** How well quantified are they, and to what extent are they expressed in concepts and metrics which are interoperable? Just as a very simple starting point: **Is the same kind of AI being discussed in each**, or are the environmental costs of carbon-intensive foundation AI models being weighed against the environmental benefits of much more lightweight AI models or data-driven methods?

On the other hand, it may be countered that the 'pros' column is disadvantaged, since there is inherently more uncertainty in calculating emerging AI climate benefits (any given initiative may or may not work out) than there is in calculating AI climate costs (since we can at least estimate to some degree the hardware and software necessary to find out). One expert we spoke to suggested that

> If you believe AI can speed up drug discovery, it can speed up decarbonisation. It is better at science and tech type problems than for humanities stuff. There are a lot of problems that need to be tackled, and AI is already helping. I usually find industry news, backed up by word-of-mouth with experts, to be a better indicator than peer reviewed studies. A counter-argument I would take more seriously is that AI will exacerbate inequality, making societies less equitable and fragile, and hence more destructive to the environment, despite progress on the technical side.[34]

As major AI companies fail to deliver their decarbonisation pledges, we are increasingly hearing the justification that AI is responsible (or soon will be) for substantial decarbonisation outside of these companies' value chains. The argument goes that, if the growth in AI's carbon footprint is outweighed by the carbon reductions enabled by that growth, and these reductions are truly additional — i.e. would not have happened anyway, with no AI or more parsimonious and green AI — then the growth is justified. Slowing the expansion of data centres would, on this analysis, actually result in more emissions.

A variation of this argument is that data centres' demand for electricity is spurring global investment in renewable energy generation. This claim also needs to be weighed up carefully. It raises many questions. To what extent might big tech companies have been able to invest in these renewable projects while also delivering on their decarbonisation pledges? Are failed pledges really a precondition of this investment and if so, how? In the absence of such investment, could these renewable energy projects have been financed in other ways? What policy incentives might unlock such investment? Are there other pathways for renewable energy growth that are more democratic and/or equitable in their wider social impacts? Is big tech really paying a premium that other investors are unwilling or unable to pay, and if so, could there be blended finance or other policy options to make up that gap? To what extent is the

---

[34] Anonymous, personal correspondence, August 2024.

additional renewable energy capacity locked into supplying data centres, and to what extent might it be switched to other purposes?

As Bill Gates puts it: “The question is, will AI accelerate a more than 6 per cent reduction?” (quoted in Mooney and Hodgson, 2024). Gates is working on the assumption that data centres will drive a rise in global electricity usage of between 2-6 per cent. This is not an easy assessment to make, given the intrinsic uncertainties of emerging technology and future innovation, and the alarmingly unreliable quality of current academic and industry discourse on AI for the climate.

Which Sustainability?

Sustainability technically includes social flourishing too, although this is occasionally forgotten or sidelined in conversations about the sustainability of AI. There are many persuasive analyses of the limitations of sustainability as a conceptual framework, for instance from post-development perspectives.

Another approach is to focus on *many* possible environmentally sustainable futures. AI tends not to support them equally, nor make space for democratic debate on which to pursue. So even when we discover that an AI system does in fact promote a more environmentally sustainable future, that shouldn’t be the end of our questions.

One example might be the use of AI to support the transition to Electric Vehicles. In the context of a standard green growth paradigm, the International Energy Agency (IEA) projects annual nickel supplies must increase from approximately 3.4 million tonnes currently to 5 million tonnes by 2030. Similarly, copper supplies need to rise from 25 million tonnes to 35 million tonnes. This demand relates largely to Electric Vehicles (EVs) and battery storage, although there are other factors, including ICT. AI innovation is now being used to search for mineral deposits. The company KoBold, uses AI to synthesise a variety of data sources and create detailed maps of the Earth’s crust to identify potential metal deposits.

Of course, the extent to which AI has contributed to successful discoveries remains debatable, and there are potential criticisms regarding not only neocolonialism and extractivism in the mining itself. What we want to draw attention to here is the kind of environmentally sustainable future that is being presented as desirable — as opposed to a future, for example, with much more public transport, and more walkable urban environments. Sometimes an elegant technical solution can foreclose messy but necessary debates about what is really valuable. This would be a debate involving (although we won't get into it here) AI imaginaries, self-driving cars, and companies including Uber and Tesla. Or to put a more positive spin on it: conversations about AI for sustainability can be jumping-off points for exciting and important conversations about the many futures that are possible, and what we really value most.

Occasionally there are some ominous statements from AI and/or sustainability professionals, in which GenAI is hailed for its potential to revolutionise **sustainability reporting**. This may (who knows) eventually be the case, but for now seems to indicate several misunderstandings: misunderstandings of the legal and social purposes of sustainability reporting, of the many practical challenges producers and users of sustainability information face, and of the capabilities of GenAI.

Current GenAI has **some uses within sustainability reporting**. GenAI is effective at quickly producing grammatically accurate text responding to a wide variety of prompts, which if treated as a first draft and carefully reviewed and revised may speed up report-writing for many users. GenAI also has demonstrable value for quick, efficient brainstorming around a topic. This is a game-changer for many people's experience of their daily working lives.

However, this does **create the risk that GenAI outputs will not be adequately reviewed and revised**, or that insufficient time and resources will be devoted to coming up with relevant angles and themes.

Beyond this, we find no evidence that current GenAI, even in its anti-hallucinatory RAG and hybrid LLM/GOFAI forms, is fit-for-purpose beyond certain narrow aspects of sustainability reporting. GenAI is frequently touted for its capacity to summarise large amounts of information, but there is **insufficient research into the quality of these summaries within real-world regulatory settings, or their effect on sustainability workers' experiences and practices**. To date GenAI clearly does not excel at providing summaries of large quantities of information in cases where the most salient details are not positioned and expressed in ways which advertise their significance, but rather require attentive and patient detective-work to reveal. GenAI has not demonstrated itself fit-for-purpose for accurately communicating information in alignment with reporting standards, ensuring comparability, making fine distinctions, improving transparency and useability, reducing greenwashing, aligning risk management and double materiality approaches to sustainability, and clearly identifying gaps and uncertainties rather than disguising, evading, or confabulating. All this is even before we begin to consider questions of **bias**.

GenAI giants

A Literature Review Reviewed

How good is the standard of academic debate around AI for sustainability? In this section, we highlight one alarming specimen.

Chen et al. (2023) is an article referred to in the Stanford *AI Index Report 2024*. It aims to address an important gap, by compiling a variety of sustainability-oriented AI research, with admirably ambitious breadth. Right now, in September 2024, it has 63 citations: not bad for a recent article. What if the article turned out to have flaws that are not only serious, but obvious? We would need to check the citations to see if any of them are pointing out these flaws. If few or none are, this would be a bad sign about the current state of discourse around AI for sustainability. And that's exactly the case.

Chen et al. (2023) is an academic article published in an academic journal. But **even reading its abstract**, any alert reader—but especially one who has a humanities background and a sensitivity to style—will notice something odd. This academic article has stylistic similarities with upbeat tech journalism or "thought leadership" type publications. This tone should put us quickly on guard, and warn us **not to cite this article without investigating more closely**.

> Climate change is a major threat already causing system damage to urban and natural systems, and inducing global economic losses of over $500 billion. These issues may be partly solved by artificial intelligence because artificial intelligence integrates internet resources to make prompt suggestions based on accurate climate change predictions.

Upon closer investigation, errors and ambiguities are everywhere. The figure of "$500 billion" does not appear elsewhere in the article. Abstracts should not include information or figures that aren't substantiated or elaborated upon in the main body of the text. Potential sources might include 2016 UNEP estimate of annual climate adaptation costs, or a Federal Reserve estimate in 2019 of losses over the previous five years; but neither of these sources would justify this figure being used in quite this way. Moreover, *does* AI "[integrate] internet resources to make prompt suggestions based on accurate climate change predictions"? This is not an accurate description of AI in general, nor is it an accurate description of most AI. The word "accurate" is particularly troubling: climate change modelling a field which carefully taxonomises and quantifies uncertainties, not one which boasts of "accuracy" in some unqualified sense (see our *Communicating Climate Risk: A Toolkit* for more). The abstract then continues more reasonably, indicating the range of areas the article will survey:

> applications of artificial intelligence in mitigating the adverse effects of climate change, with a focus on energy efficiency, carbon sequestration and storage, weather and renewable energy forecasting, grid management, building design, transportation, precision agriculture, industrial processes, reducing deforestation, and resilient cities. We found that enhancing energy efficiency can significantly contribute to reducing the impact of climate change.

There is another slightly odd moment after this, as the abstract seems to assume that the global natural gas industry should be responsible for state-of-the-art weather forecasts, as opposed to meteorological forecasting services. There are then some impressive sounding percentages that raise plenty of questions about baselines and methodologies:

> Smart manufacturing can reduce energy consumption, waste, and carbon emissions by 30–50% and, in particular, can reduce energy consumption in buildings by 30–50%. About 70% of the global natural gas industry utilizes artificial intelligence technologies to enhance the accuracy and reliability of weather forecasts. Combining smart grids with artificial intelligence can optimize the efficiency of power systems, thereby reducing electricity bills by 10–20%. Intelligent transportation systems can reduce carbon dioxide emissions by approximately 60%.

Turning to the article itself, a striking error in the first paragraph suggests that, whether or not there is AI expertise underpinning this article, **fundamental understandings of climate change are lacking**. “The widespread use of fossil fuels in manufacturing processes is primarily responsible for the extensive carbon dioxide emissions (Yue and Gao 2018).” This identifies only one significant category alongside others such as electricity and heat, transport, manufacturing and construction, agriculture and land use change, and various others.

More pedantically, there are plenty of other tells that climate science and climate policy are relatively unfamiliar territories: “Artificial intelligence can aid in mitigating climate change in multiple ways, such as improving the prediction of extreme weather events” (Chen et al. 2023) — within climate science and climate policy, *mitigation* has a quite specific meaning, and the prediction of extreme weather events does not fall under climate change mitigation, but climate change *adaptation*. Using a word in something like its more everyday sense is not in itself a major problem. But **the accumulation of many small clues like this should lead us to scrutinise the article more closely**. That is when the major problems are revealed.

Most seriously of all, **the article misrepresents the findings and even the topics of many of the sources it cites**. For example, the cited source Yue and Gao (2018) does not actually support the erroneous claim quoted above. The closest it comes is the sentence, “Fossil fuel energy consumption remains the primary source of GHG emissions,” but fossil fuel energy consumption is not the same as the use of fossil fuels in manufacturing processes. The article by Yue and Gao is about the contributions of human systems vs. natural systems to greenhouse gas emissions, but it does not break down human systems emissions with any granularity. Likewise, looking at the use of AI for energy transition, the article again makes bold claims:

> In the energy sector, the implementation of artificial intelligence can heighten the efficiency of energy utilization by predicting energy demand, optimizing energy production and

> consumption, and realizing intelligent control, thus curtailing energy costs, lessening environmental pollution, and fostering sustainable development (Khalilpourazari et al. 2021; Lee and Yoo 2021).

The cited articles do not support these claims. Khalilpourazari et al. (2021) is not about improving the sustainability of energy production and distribution. It is an unrelated article, about machining (“removing material from a specimen using special cutting tools”), and specifically about “optimizing the turning process that uses a non-rotary single-point cutting tool to create axisymmetric parts by cutting undesirable material.” Lee and Yoo (2021) is also not about improving the sustainability of energy production and distribution. It is about training AI models on user's devices (phones, laptops etc.) instead of on the cloud.

On carbon sequestration, the article claims: "Artificial intelligence-based technologies can be harnessed to discern appropriate geological formations for carbon storage and prognosticate the behavior of carbon dioxide after it is introduced into storage sites (Abdalla et al. 2021)." The article it cites does not support this claim. Abdalla et al. (2021) is about energy storage, not carbon sequestration. It does not mention the behaviour of stored carbon dioxide. "Furthermore, artificial intelligence can optimize the injection procedure and monitor storage sites to ensure carbon dioxide is securely trapped underground (Li et al. 2021)." The cited article, Li et. al (2021), does not support this claim. It does not mention the optimization of carbon injection. It is not about carbon sequestration and does not mention carbon sequestration.

These are not the only errors, but they are more than enough to demonstrate that the article is not a credible source. **How have such errors occurred?** In our experience, similar errors are often encountered in student essays. Sometimes students will write their essay first, then look for sources to back up their claims — and when they can’t find them, will cite vaguely related papers, taking the reasonable gamble that overworked lecturers will not actually read every paper that they cite. In our experience, these types of errors have increased substantially in the past two years, as students have begun to use LLMs to help to write their essays. For example, an LLM might suggest inappropriate citations, which the student does not check. Or an LLM might hallucinate citations to papers which do not exist, which the student then replaces with real citations, but ones which do not support the claims they are making. Students may also rely on an inaccurate LLM-generated summary of a paper, rather than reading the paper. An LLM-driven research tool such as scite.ai/ may provide incorrect summaries. Occasionally inaccurate translation AI may also play a role. In our view, it is highly likely that **an inappropriate use of AI is the source of most or all of these errors.**

Alongside these major errors, the cumulative effect of **many subtle misrepresentations** is also dangerous. For example: “It [precision agriculture] is making modern agriculture more profitable and sustainable by applying artificial intelligence (Ampatzidis et al. 2020; Wei et al. 2020).” Ampatzidis et al. (2020) is indeed relevant, and supports some aspects of the claim. But Chen et al. (2023) **misrepresents the maturity of this technology**. The article features a single study, which surveyed an orchard in Florida via drones carrying low-cost RGB cameras, and analysed the results with two Convolutional Neural Networks (CNNs).

> Precise and efficient crop management in orchards depends on methods to detect and assess individual trees. A cloud- and AI-based technique (Agroview application) was developed to automatically process, analyze, and visualize UAV collected data for individual tree monitoring and assessment. This interactive application comprised a machine vision algorithm (AI-based) that uses deep learning to effectively detect individual plants on aerial maps. (Ampatzidis et al. 2020)

The results were encouraging: in these circumstances the algorithm was very good (almost as good as humans) at correctly identifying trees. It was not bad (but not as good at humans) at correctly estimating the heights and canopy areas of those trees. The algorithm produced results much more quickly and cheaply than humans would. (The experiment also aimed to estimate the health of each tree, although these outputs were not evaluated within this study). But Ampatzidis et al. (2020) does not support the claim that the technique has been successfully commercialised and implemented to "[make] modern agriculture more profitable and sustainable."

There is a similar issue with the use of Wei et al. (2020). Satellite images were taken during different stages of the carrot crop development. These images capture data in different spectral bands (types of light), such as near-infrared (NIR) and visible light (blue, green, red). Carrot yield data was collected from specific, georeferenced locations in the field. This means that the exact positions where the data was collected are known and can be matched with satellite data. As is often done in Machine Learning, the entire dataset, which includes both the ground-truth yield data and the satellite spectral data, was split into two parts: a training set to build the model and a test set to evaluate its performance. This model was tested and found to have a fairly small error rate and a high level of accuracy ($R^2 = 0.82$). A tool created from this model could forecast carrot yields over larger areas, and support farmers making decisions about where and how to plant. However, Wei et al. (2020) explain that more work is needed:

> Future works should aim to evaluate the minimum area and number of ground-truth samples necessary to faithfully represent larger areas when applying the RF regression algorithm to predict crop yield based on temporal spectral data from satellite imagery, and not only for carrot crops. In addition to that, it is also necessary to evaluate the possibility of estimating carrot yield from satellite imagery with different spatial and temporal resolutions. [...] Hopefully, this approach of applying ML techniques to datasets containing a certain number of ground-truth samples in a given area and the temporal spectral data from the crop canopy cycle will allow the creation of accurate yield maps to help support decision makers in enhancing their crop production with respect to the PA goals. (Wei et al. 2020)

At a minimum, claims could be revised to indicate the early stage of the underpinning research; something like: “Precision agriculture using AI promises to make modern agriculture more profitable and sustainable (Ampatzidis et al. 2020; Wei et al. 2020).” Better yet would be to capture the specific nature of the case studies referred to here, by providing Technology Readiness Levels (TRLs) and/or other relevant context. It is challenging to do a literature review of emerging technology, and it is appropriate to move beyond peer-reviewed literature into grey literature and other sources (press releases from AgroView, for example): the required skill is knowing how to present all sources in a self-reflective, measured, and sufficiently critical way.

After discovering the extent of these errors, we decided to contact the authors, and eventually also Springer to ask for this article to be withdrawn. We have received notification that the article is being reviewed, though after two months it still remains on the journal website.

But there is an even deeper issue. Let's imagine instead that the article went through several more revisions, and errors like the ones just identified were fixed. Even if such an article represented accurately the great variety of interesting AI for sustainability studies and demonstrators that currently exist, would we really be out of the woods yet? How broad and deep are the kinds of problems identified in this review article across the underlying literature itself? Then there would still be the issue of an underlying unreflective techno-solutionism. That is, there are not only questions about the technical efficiency of AI for sustainability, but its embedded values, and the futures it assumes as inevitable (see sidebar for a little more).

These brief forays into “AI for sustainably” discourse have revealed **a scandalous lack of rigour**. We recommend caution and careful scrutiny. The overall framing of “AI for sustainably” can also conceal important distinctions. While acknowledging of course that knowledge and tools may sometimes spill over from one area of AI research to another, we recommend maintaining these distinctions carefully. It’s essential to maintain clarity when discussing the relationship between lightweight discriminative ML models and heavyweight ML foundation models, as well as other approaches such as symbolic AI and active inference. Success with lightweight models should not inadvertently justify investment in unrelated heavyweight models. Additionally, AI tools developed for climate adaptation should not be assumed to be easily adaptable for climate mitigation purposes. When proposing AI for tasks that could potentially be accomplished by other means (including non-AI approaches), it is critical to conduct a thorough comparison, considering the benefits, drawbacks, and carbon footprints of each approach.

At a somewhat higher level of regulation, policy design, and anticipatory governance, it is also important to consider the different challenges and uncertainties of calculating the costs of AI, versus the potential benefits of AI.

More ‘AI for Sustainability’ Academic Spam

Just because there is poor-quality scholarship hyping potential for AI to benefit the climate does not mean that AI is *not* benefiting the climate, or won’t do so more in the future—it just makes it harder to find the actually promising case studies and useful evidence. Here’s another example. Various kinds of AI have been used in solar energy systems for a number of years (Garud et al. 2020).

However, the top Google Scholar result for the query “AI manage supply and demand photovoltaic systems” since 2020 is Mahjabeen and Mahjabeen (2023), ‘Revolutionizing Solar Energy: The Impact of Artificial Intelligence on Photovoltaic Systems.’ The article has 74 citations to date, so people are reading, trusting, and quoting this article. The journal does not appear in Scimago’s journal rankings, which might be a warning sign; then again, it does not appear in Beall’s List of potential predatory journals either. The breadth of the journal’s themes (*International Journal of Multidisciplinary Sciences and Arts*) could also be a warning sign, given that the article in question is not especially interdisciplinary or transdisciplinary. The journal’s information page describes a standard double-blind peer review process.

Let’s have a look at the first three sources which the paper cites. The paper claims, “The production of solar energy can be maximised using AI, which improves performance, efficiency, and total system productivity.” The cited source immediately rings an alarm bell because of its date—2001 is a long time ago in both the development of AI and the development of renewable energy. In fact the source, Liu (2001), does not support the claim at all: it is about the design of CO2 hydrogenation catalyst by an artificial neural network. The paper also describes “AI-based solar panel tracking systems” which can “dynamically change the panel orientations and angles throughout the day to maximise sunlight absorption.” But the cited source, Luna-Rubio et al. (2012), is a review of the state of methodologies used to size hybrid energy systems, and does not mention solar panel orientation. The paper also describes how “AI can forecast maintenance needs and identify potential defects before they seriously damage the system by creating patterns and correlation,” extending the lifespan of the solar panel. But the cited source, Ma et al. (2016), is about the use of bio-inspired algorithms to identify parameters to be used in modelling the behaviour of photovoltaic systems. It might in principle have some indirect relationship to solar panel maintenance, but it does not mention directly forecasting maintenance needs or identifying potential defects.

The errors continue. They strongly suggest LLM hallucinations. In the literature on AI for the climate, we have encountered some circumstantial evidence of academic research that is well below publishable standard representing likely academic misconduct going largely unnoticed and unremedied. The literature on AI for the climate needs urgently to be reviewed to determine the nature and scale of this problem. Peer review processes need to be examined, and aggregator sites (such as Google Scholar) need better mechanisms for verifying journal quality and reporting suspicious content.

Relevant UK policy on sustainable AI

Artificial Intelligence (AI) is transforming various sectors. PwC estimates that AI could contribute $15.7tr to the global economy in 2030, and have increased global productivity by up to 14%. At the same time, however, Goldman Sachs points to the perils of euphoric expectations, and *The Economist* claims provocatively that AI has had almost no economic impact to date. The last UK government made AI a priority, with AI and Data being identified as a 'Grand Challenge' in the Industrial Strategy White Paper. However, the increasing environmental impact of data centres and digital technologies, characterised by rising energy consumption, e-waste, and use of water and other resources, has prompted concerns. There are now numerous frameworks, guidelines, and regulations emerging aimed at governing the environmental impacts of AI. This space is however far from mature, and in the coming months and years, it is crucial that civil society, research and innovation communities, communities vulnerable to AI-related and climate-related risks, and other stakeholders, are actively involved in these conversations. As Nataliya Tkachenko suggests, “the current AI adoption frameworks focus exclusively on operational safety without specific mandates on environmental impact or ‘greenwashing’. In contrast, climate regulations emphasise sustainability and environmental accountability but do not specifically address the implications of AI technologies” (Tkachenko, 2024).

The UK Conservative government took a somewhat laissez-faire approach to AI regulation compared to the EU. In 2024 the new Labour government has signalled an intention to create new legislation, although at time of writing (July 2024) the details are not yet clear. Labour’s manifesto included a short reference to its AI plans. Ahead of the general election, the party said it intended to introduce “binding regulation on the handful of companies developing the most powerful AI models.” It is likely that there will be consultation, and this may be a significant opportunity for civil society and academia to inform robust and effective AI environmental governance. There are some early indications that the UK approach may continue to be less comprehensive than the EU’s. However, ambitions to rebuild post-Brexit relations with Europe may lead to some convergence with the EU’s approach (see EU AI Act, and the EU’s ongoing consultation for a general purpose AI Code of Practice).

Case study: Does AI generated art and writing have a lower carbon footprint than human art and writing?

Tomlinson et al. (2024), writing in *Nature*, compare the carbon emissions of GenAI systems (ChatGPT, Midjourney etc.) with equivalent human processes.

Elsewhere in this report, we show why academic articles like Chen et al. (2023) and Mahjabeen and Mahjabeen (2023) slipped through peer review when they really should not have, and now need to be suspended, reviewed, and withdrawn. We want to be clear that Tomlinson et al. (2024) is a different kettle of fish — while we may disagree with its methodology and conclusions, we believe it is a useful contribution to an interesting and important topic, which also reflects admirably on its own limitations. What the article shows very clearly, however, is the complexity of evaluating the environmental impacts of an activity "with AI" vs. "without AI." This is because **AI seldom if ever neatly replaces some human activity** (doing the same thing for the same reasons, only in a new way). AI transforms what we do, how we do it, and why we do it.

Tomlinson et al. want to **compare the carbon cost of a human artist creating an artwork with the carbon cost of AI creating an artwork**. They also look at writing, but we'll just focus on the visual artwork here. They come to the striking conclusion (and in our view, a highly misleading conclusion) that "AI illustration systems emit between 310 and 2900 times less CO2e per image than their human counterparts" (Tomlinson et al. (2024)). A reasonable implication is that if we want to create art in environmentally sustainable ways, we should be using AI, not commissioning human artists.

Can and should AIs be artists? The whole topic also invites deeper reflection on the intrinsic value of artistic creation — not just in terms of the final product, but in the human experience of creativity, individually and collectively. There are also ethical questions about the use of human artists' work in the training data. These are important questions, but for our purposes here we can set them aside and just focus on the carbon footprint.

The higher figure (2900 times less CO2e) comes from the assumption that the carbon cost of a US-based human artist creating an artwork is the equivalent to about 0.037% of the annual carbon footprint of an average US resident.[35] The idea is that it takes on average 3.2 hours to create an artwork, which is about 0.037% of a year. However, this means that all the activities that US residents undertake — food consumption, road travel, air travel, heating and lighting, and so on — have been lumped together as one big aggregate, and then a tiny slice taken out and attributed to making art. Hopefully the problem is already plain. By this logic, any activity that takes 3.2 hours will have exactly the same estimated carbon footprint—whether you are driving, gardening, playing a sport, going for a walk, sleeping, rioting, writing emails, reading a book, or doing something else. Furthermore, the carbon intensity of artwork creation will depend on the full range of activities undertaken by US residents, and the carbon intensity of each of these. Over time this would entail

---

[35] The authors also offer a similar calculation for India. We focus on one here for simplicity, and choose the US because it is the higher figure.

nonsensical results: **the carbon footprint of art would appear to grow if people began to take more planes, or to shrink if they began to eat less meat.** Similarly, increasing the use of image-generating AI will not actually remove human artists from the population (at least, we hope not!). A related inconsistency is that **the estimate of an AI artwork's carbon footprint explicitly excludes a human user**. The time it takes to formulate a prompt, await and review the results, and potentially iterate or "reroll" the prompt,[36] is assumed to be zero. This is before we even consider activities like editing the results to fit the brief. Overall, this methodology is not fit-for-purpose.

This points to three connected issues. First, neither a human artist using traditional methods, nor a human using an AI, will necessarily generate a single image in response to one image generation task. In fact, **it is very likely that a GenAI user might generate several images before finding one that is suitable** (this is something that can be researched empirically). Second, image-making GenAI does not only impact the creation of new images. **GenAI also impacts the re-use of old images, e.g. from stock photo libraries or public domain repositories**. Third, the ease of creating images with GenAI may lead to **users creating images for contexts where previously they would have used no image.**

In short, we should not assume we will find clear-cut mappings between "a human making an image" and "an AI making an image." Introducing AI into social practices will seldom neatly replace certain components of those practices with optimised functional equivalents. It often can transform the entire field of practice: who is participating, what they need and want, what they do, and how and why they do it. There are other issues we might discuss, but we have covered the key ones which establish **the complexity of making comparisons between AI and non-AI versions of "the same" activity**.

To make progress here, we make three suggestions. First, the paper does offer an alternative calculation for the carbon footprint of human-created artworks, based on the energy of using a device for 3.2 hours. We see this as more promising. Second, some emphasis on **carbon elasticity of income** may help to open up these questions (see Pottier 2022). Attributional approaches to carbon accounting should be supplemented by consequential approaches—what is the total carbon impact of decreasing artists' income, and allocating these funds to AI and other alternative expenditures? One approach might include an attempt to estimate not the carbon cost of one image generated, but the carbon cost of one pound or dollar spent on commissioning art vs. spent on generative AI. Thirdly and most importantly, this kind of research needs to be very richly **interdisciplinary** and ideally **co-produced** with the communities it concerns. In this case, the actual social practice of making, commissioning, curating and using artworks, should be part of the research, as should the voice of artists and other stakeholders.

---

[36] Since there is randomness involved in image generation, the same prompt will produce different images on most AI systems, unless explicit measures are taken to prevent it. Using the same prompt again and again in hope of a better result is sometimes called "rerolling" (as in rerolling dice, to get the result you want).

'Power / Profit' by Clarote

AI Regulatory Landscape Resources

There are a number of resources to help track the rapidly evolving landscape of law, policy, standards and governance tools that are directly or indirectly related to AI and the environment.

- AI Standards Hub: As part of the UK's National AI Strategy, the Hub's mission aims to advance responsible AI, focusing on standards as governance tools and innovation mechanisms. The Hub brings together industry, government, regulators, consumers, civil society, and academia to shape AI standardisation through debates, strengthen AI governance practices, increase multi-stakeholder involvement, and facilitate the assessment and use of published standards.

  The AI Standards Hub, led by The Alan Turing Institute, the British Standards Institution (BSI), and the National Physical Laboratory (NPL), is a collaboration between the UK Government and the Department for Science, Innovation and Technology (DSIT). The Alan Turing Institute, the UK's national institute for data science and AI, is a community of over 500 researchers, software engineers, and data scientists. BSI, the National Standards Body, builds trust in digital transformation by convening stakeholders to agree on priorities for action and sharing consensus-based best practices. NPL, the UK's National Metrology Institute, develops and maintains national primary measurement standards, providing confidence in AI technology and data. The DSIT leads the UK Government's strategic engagement on digital technical standards, working with the multi-stakeholder community to shape global digital technical standards for democratic values, cyber security, and UK strategic interests through science and technology.

The Hub's work comprises four pillars: observatory, community and collaboration, knowledge and training, and research and analysis. The observatory includes "the Hub's AI Standards Database, a searchable catalogue of standards being developed or published by relevant SDOs. Additional libraries keep track of other documents and publications with implications for standards, such as government strategies, standardisation roadmaps, or laws and regulatory requirements with relevance for AI technologies."[37]

At time of writing, the Hub lists fifteen standards tagged "Sustainability," from standards bodies such as ISO, IEC, IEEE, ITU, etc. The majority of these do not focus primarily on sustainability but include it within some wider remit.

- ICT Regulatory Tracker**:** The ICT Regulatory Tracker by the International Telecommunication Union (ITU) is a tool that tracks the evolution of ICT regulation, assisting decision-makers and regulators in benchmarking and identifying trends. It records the existence and features of regulatory frameworks, promoting innovation and reform for a vibrant ICT sector. It is composed of metrics based on 50 indicators and covers between 190 and 193 countries and economies over the period 2007 – 2022. Each indicator is scored based on qualitative information, with a score between 0 and 2 indicating the best possible scenario based on internationally recognised regulatory best practices adopted by the global community of regulators at the annual ITU Global Symposium for Regulators. The ICT Regulatory Tracker dataset is based on self-reported data from ITU surveys, international datasets, government research, and outreach to national authorities. It is updated every two years and may undergo further research using official government sources to enhance its completeness and accuracy. This tracker analyses global ICT regulation evolution, identifying progress areas and gaps and measuring them to facilitate digital transformation in countries through policy, regulation, and collaborative governance.

- UNESCO Global AI and Ethics Observatory: The Observatory aims to aid policymakers, regulators, academics, the private sector, and civil society in addressing AI challenges by providing information on countries' ethical AI adoption readiness. It also hosts the AI Ethics and Governance Lab, which showcases research, toolkits, and practices on AI ethics, governance, responsible innovation, standards, institutional capacities, and neurotechnologies.

- OECD Tools for Trustworthy AI: The catalogue is a global platform for AI practitioners to share and compare tools, collaborate on best practices, and accelerate OECD AI Principles implementation. The five value-based principles which are:

    - Inclusive growth, sustainable development and well-being
    - Human-centred values and fairness

---

[37] <https://aistandardshub.org/the-ai-standards-hub/>

    - Transparency and explainability
    - Robustness, security and safety
    - Accountability

  Guide countries in policy shaping and AI risk framework creation, promoting global interoperability.

- The MIT AI Risk Repository is a living database of over 700 AI risks, based primarily on literature review.

- Fairly AI's Map of Regulations: Fairly AI is a private company based in Canada, which provides AI compliance and Quality Assurance advisory services, as well as some free tools and resources (including the Map of Regulations). Fairly AI evolved from an interdisciplinary research project involving philosophy, cognitive science, and computer science, and was formally incorporated in 2020. The company's mission is to democratise safe, secure, and compliant AI.

- There are many commercial services e.g. *The Compliance People* who provide subscribers with customisable dashboards, which will track updates to legislation and regulation on selected topics, and/or according to the subscriber's company profile.

- Stanford University's annual *AI Index Report* includes some coverage of responsible AI and AI governance.

A Pro-innovation Approach to AI Regulation

The pro-innovation approach to AI regulation is a policy paper presented to UK Parliament by the Secretary of State for Science, Innovation and Technology in 2023, which encourages a regulatory environment that promotes the development and deployment of AI technologies while ensuring their safety, ethics, and societal benefits. This paper aims to achieve 3 objectives:

- Drive growth and prosperity,
- Increase public trust, and
- Strengthen the UK's position as a global leader in AI based on principles.

Existing regulators are expected to implement a framework based on the five values-focused principles, aiming to balance clarity, trust, and experimentation.

- Safety, security & robustness
- Appropriate transparency and explainability
- Fairness
- Accountability and governance
- Contestability and redress

Environmental sustainability may appear to be a glaring omission here, although it could also be argued that environmental sustainability is a cross-cutting issue implied in each of the five value-focused principles.

Comments: UK's approach in context

The UK's proposed AI regulation approach may need to effectively address the challenges and opportunities presented by AI technology compared to other countries' strategies. It needs to consider critical mechanisms that promote AI space, such as innovation and efficiency adopted in other countries.

- **The EU** currently has the most **comprehensive regulatory approach**. The EU AI Act categorises AI systems based on risk, imposes strict requirements for high-risk applications, and emphasises ethical considerations, human rights, and safety. It prohibits certain harmful use-cases. It requires transparency, accountability, and robust enforcement mechanisms for AI regulations. Annex XI requires providers of general-purpose AI models to supply "known or estimated energy consumption of the model." The EU has also introduced the European Green Deal, a comprehensive framework designed to align the EU's actions and policies with the goal of reducing net greenhouse gas emissions. As part of the Green Deal, the EU Climate Law establishes legally binding targets for achieving carbon neutrality by 2050. AI systems will be required to comply with these regulations. There are also regulatory initiatives at the national levels (e.g. in France).

- **The USA** currently uses a **piecemeal regulatory approach**, focusing on industry-specific regulations and promoting innovation and economic growth to maintain a competitive edge in AI technology. **Voluntary compliance** plays a relatively large role. One interesting development is the AI Environmental Impacts Act of 2024, a Bill under review, which would be the first piece of legislation specifically addressing the environmental implications of AI. Introduced by Senators Markey and Heinrich, along with Representatives Eshoo and Beyer, it proposes a comprehensive study on AI's environmental impacts and aims to establish a voluntary reporting system to monitor these effects. Prior to this, the White House issued the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence in October 2023. This Executive Order does not directly address environmental concerns, although it does briefly highlight the potential for AI to contribute positively to climate goals. In addition, the Federal Trade Commission (FTC) has issued guidelines on the fairness and transparency of AI systems (Luccioni 2024).

- **China's** AI strategy centres on state involvement and control, with a national agenda to become a global leader by 2030. **To date China has adopted an 'agile' approach** of releasing targeted regulations relatively rapidly and in some cases consulting and iterating them. However, a more comprehensive AI law is now in development (Yang 2024).

UK Regulator Updates

For now, the UK has adopted an outcome-based framework for regulating AI, underpinned by five principles: safety, security and robustness, appropriate transparency and explainability, fairness, accountability and governance, and contestability and redress. There are indications that legislation will be brought in eventually.

The 2023 white paper 'A pro-innovation approach to AI regulation' laid out the UK Government's approach. In February 2024, following a consultation, the UK Government confirmed its commitment to a "proportionate, context-based approach" to regulation, aiming for an approach that is both “pro-innovation” and “pro-safety.” Themes identified in the consultation response included human rights, operational resilience, data quality, international alignment, systemic risks and wider societal impacts, sustainability, and education and literacy. Key regulators were also asked for updates on how they were taking forward the white paper proposals:

- Bank of England
- Competition and Markets Authority (CMA)
- Equality and Human Rights Commission (EHRC)
- Financial Conduct Authority (FCA)
- Health and Safety Executive (HSE)
- Information Commissioner’s Office (ICO)
- Legal Services Board (LSB)
- Medicines and Healthcare products Regulatory Agency (MHRA)
- Office for Nuclear Regulation (ONR)
- Office for Standards in Education, Children’s Services and Skills (Ofsted)
- Office of Communications (Ofcom)
- Office of Gas and Electricity Markets (Ofgem)
- Office of Qualifications and Examinations Regulation (Ofqual)

These updates are useful indicators of current thinking, though mostly quite light on detail. Climate and environment are not major themes. Ofgem comments: “AI could play a big part in decarbonising the energy sector, for example it can be used to better predict weather that can help improve solar generation forecasts. Using this technology in this way means there should be less reliance on fossil fuels. This means that the use of AI technology can support the UK government’s medium and long-term net zero targets.”[38]

At time of writing, the details of the new Labour government’s approach have not yet emerged.

Government Cloud First Policy

In 2013, the UK Government introduced the Cloud First policy for all public sector organisations, mandating central government departments to consider the cloud before any other IT implementation option, and recommended all UK public sector organisations to do the same.

> When procuring new or existing services, public sector organisations should default to Public Cloud first, using other solutions only where this is not possible. This approach is mandatory for the central government, and strongly recommended to the wider public sector. [...] Organisations who do not deploy in Public Cloud should ensure they can evidence the decision, business case and value for money behind their choice.

---

[38] Artificial Intelligence (AI) within the energy sector | Ofgem.

Data Ethics and AI Guidance Landscape

In 2019 the UK's Alan Turing Institute published a framework for ethics and safety in the development and use of AI systems in the public sector. This forms a useful background to the 2023 white paper. The framework includes:

1. SUM Values: Support, Underwrite, and Motivate. The guiding values are Respect, Connect, Care, and Protect.
2. FAST Track Principles: Fairness, Accountability, Sustainability, and Transparency - actionable principles to apply ethical standards in practice during AI development and use. Fairness and sustainability establish normative criteria for design and outcomes. Accountability and transparency provide mechanisms for responsibility and justification.
3. Process-Based Governance (PBG) Framework: Technical and non-technical tools tailored to specific sectors to help implement ethical values and principles and make appropriate trade-offs.

This framework, adopted by the UK government, provides some guidance for incorporating ethics into AI from early conception through implementation, use, and maintenance. It aims to enable responsible innovation of AI in the public sector.

Also, when considering procurement of AI, assess the benefits and risks of AI deployment and ensure the system's human and socio-economic impact aligns with the social value guide principles which encompasses activities

- Promoting community unity,
- Environmental efforts to reduce waste, and
- Economic opportunities for disadvantaged groups, such as training and employment.

AI can be a double-edged sword as it can be a powerful tool for climate action, offering solutions such as improving energy grid efficiency, modelling climate change predictions, and monitoring climate treaties. However, the infrastructure required for AI is energy- and resource-intensive, with large language models like OpenAI's GPT-3 requiring significant amounts of electricity and water for cooling.

However, one crucial environmental issue often overlooked in AI debates and frameworks is the possibility of increased energy usage from using AI in daily tasks. If this risk is not acknowledged, the advancement of AI might exacerbate the climate disaster by significantly affecting overall energy use at a time when we desperately need to cut back on energy consumption. Therefore an establishment of policy measures that align AI and technology regulation with environmental sustainability should be the focus.

ISO/IEC JTC 1/SC 42

ISO/IEC JTC 1/SC 42 is a Joint Technical Committee of the International Standards Organisation and the International Electrotechnical Commission. It is currently developing a number of AI standards, including those relevant to AI and the environment. AI solutions will vary widely in their applications, architectures, designs, data required to train the applications, and other resource requirements. This will pose a significant challenge in developing standardised approaches to environmental sustainability. Standards need to provide consistency while accommodating diversity.

AI technologies are evolving, and will require flexible and adaptable standards that can also evolve alongside emerging technologies and environmental concerns. Therefore, continuous updates and revisions, with possible ongoing monitoring, might be required to keep the standard relevant and effective over time.

Of the 28 standards published to date, none focuses exclusively on environmental sustainability. A new standard in development suggests this is now becoming a priority. **ISO/IEC CD TR 20226** (Information Technology and AI and Environmental Sustainability aspects of AI systems) is still in the draft stage. However, it represents a promising initiative to address the environmental sustainability aspects of artificial intelligence (AI) systems. As of the date of this report, the status indicates it is under development within the ISO/IEC JTC 1/SC 42.[39]

ISO/IEC JTC 1/SC 39

ISO/IEC JTC 1/SC 39 is another Joint Technical Committee of the International Standards Organisation and International Electrotechnical Commission. It focuses on the standardisation of “assessment methods, design practices, operation and management aspects to support resource efficiency, resilience and environmental sustainability for and by information technology, data centres and other facilities and infrastructure necessary for service provisioning.” It has published 29 standards to date, mostly relating to the construction and operation of data centres. Several standards are currently in development, including standards relating to key metrics such as PUE, WUE, ITEEsv (IT Equipment Energy Efficiency for servers), ITEUsv (IT Equipment Utilisation for servers), among others.

UKRI Net Zero Digital Research Infrastructure Scoping Project

The operation of digital infrastructure is not carbon neutral, and the carbon emissions associated with its expansion and equipment used in research institutions are significantly growing and adding to the carbon emissions. The estimated yearly emissions of the UK digital research infrastructure are 35 kilotons of CO2e for laptops and server rooms and 40 kilotons of CO2e for large-scale computing facilities in UKRI-funded institutions. While the UK research community is not the key driver of carbon emissions, but the activities of DRI contribute and serve as a concern for the research institute.

---

[39] ISO/IEC CD TR 20226 - Information technology — Artificial intelligence — Environmental sustainability aspects of AI systems

To achieve net zero targets stated in the Climate Act 2008, and given the need and value of DRI, UKRI awarded £1.86 million to the Net Zero DRI Scoping project research, which represents a forward-looking initiative to leverage digital technologies to advance research and innovation to achieve net-zero emissions targets. This is not a regulatory framework but part of the UK's efforts to address climate change and transition to a low-carbon economy.

The NERC released the final technical report of the UKRI Net Zero DRI Scoping Project in August 2023. The report offers recommendations for actions and guidance to UKRI and its community. It also includes a roadmap that supports the institutional and national goal of achieving net zero by 2040 or earlier. The project resulted in three delivery pathways and six strategic themes.

The six strategic themes include:

1. The mission focus emphasises on continuous assessment and action to achieve sustainability
2. Recognition of shared responsibility empowers staff to take proportionate action.
3. Action-based research focuses on progress and learning from experience
4. Collaboration with peers and suppliers develops a low-carbon supply chain.
5. The initiative also aims to foster knowledge sharing by providing leadership, support, and advice for business cases and large procurements, serving as a central hub for information and institutional knowledge.
6. Green Software Engineering aims to transform code-writing approaches and support data centres by creating expertise and developing tools, metrics, and standards.

The following three delivery pathways align with the UKRI delivery areas identified in the 2020 UKRI Environmental Strategy:

1. **Policy and Governance:** This pathway is crucial in driving the necessary changes and ensuring that operations align with our environmental goals. It also underscores the role of policy and governance in collective ambition to achieve net zero.
2. **Collaboration**: The UKRI's delivery framework aims to achieve net zero ambition through partnerships with funders, facility leads, and service providers
3. **Competitive funding:** This approach will leverage the UK academic community's unparalleled creativity and diversity to help make significant strides towards achieving the goal.

Climate Change Act 2008

The Climate Change Act 2008[40] was a pioneering piece of UK legislation to address climate change. It received support across party lines, which has enhanced the stability and credibility of the regulatory framework. It is legally binding and sets ambitious targets to reduce greenhouse gas emissions in the UK. The long-term goal was initially to reduce emissions by at least 80% by 2050 compared to 1990 levels. In June 2019, the UK government amended this target to achieve 'net-zero' greenhouse gas emissions by 2050.

Implementation has been somewhat fraught, with the government's strategy for achieving these climate objectives deemed unlawful by the High Court for the second time in May 2024. The ruling supported the majority of the claims made by environmental advocacy groups such as Friends of the Earth, Client Earth, and the Good Law Project. These organisations challenged the adequacy of the government's Carbon Budget Delivery Plan, issued in March 2023, criticising its overreliance on speculative future technologies and questioning the plan's comprehensive risk assessment.

Carbon Budgets: The Climate Change Act also established the Committee on Climate Change, now the Climate Change Committee (CCC), an independent body to advise on setting and meeting decarbonisation goals. The CCC advises the government on carbon reduction targets, the contributions that different sectors could make, and the extent to which carbon budgets could be met through the use of permitted 'flexibilities,' e.g. offsetting.

Climate Change Risk Assessment and National Adaptation Programme: The Act requires the UK government to report on climate change risks and opportunities[41] every five years and assess the impact on current and future risks of the climate change implemented to date and how people and the planet are adapting[42] to those implementation strategies.

Comments: While the Climate Change Act 2008 represents a significant step toward addressing climate change in the UK, the government needs to make more progress to actually achieving its targets in line with the advice of the CCC. From the perspective of ICT, there is also room for greater international collaboration, more emphasis on adaptation, and more joined-up approach with the rapidly evolving field of digital sustainability:

- There are some tensions between the national scale and areas where international cooperation is required for the targets to be achieved, such as shipping and aviation, and Information Communications Technology. Reducing emissions from these sectors may require greater international collaboration for the UK to reach its targets.

- It appears that many of the potential benefits of AI may relate more to adaptation than to mitigation. Although the Act focuses on mitigating

---

[40] <https://www.legislation.gov.uk/ukpga/2008/27/contents>

[41] UK Climate Change Risk Assessment (CCRA)

[42] National Adaptation Programme (NAP)

measures to reduce emissions, it is essential to evaluate the resilience of the relevant stakeholders to climate policies, as well as the broader set of physical and transition climate risks. Developing more robust frameworks for cost-benefit analysis (or similar) of digital technologies with influence on both mitigation and adaptation will also be key.

- The Carbon Budget Delivery Plan does not appear to place great emphasis on ICT, with some exceptions (e.g. investment in AI to develop transformational technologies to support net zero transition; digitalisation of energy supply and increased use of smart devices; heat recapture from data centres and other industrial plants). This may be evidence that a more joined-up approach is needed.

Streamlined Energy and Carbon Reporting (SECR) for UK businesses

The UK government's Streamlined Energy and Carbon Reporting (SECR)[43] policy, implemented on 1 April 2019, requires businesses to disclose their energy and carbon emissions for financial years starting after 1 April 2019. The new regulations require approximately 11,900 UK companies to comply with this regulation. This builds on existing requirements such as mandatory greenhouse gas (GHG) reporting for quoted companies, the Energy Saving Opportunity Scheme (ESOS), Climate Change Agreements (CCA), and the EU Emissions Trading Scheme (ETS). Although there might be some overlap between compliance with the ESOS and SECR reporting, it does not replace ESOS obligations. The reporting of SECR aligns with the organisation's financial reporting year, ensuring integration with existing financial reporting processes.SECR also extends the reporting requirements for quoted companies and mandates new annual disclosures for large unquoted and limited liability partnerships (LLPs).

The UK's Streamlined Energy and Carbon Reporting (SECR) framework simplifies and consolidates businesses' energy and carbon reporting requirements. It encourages transparency and accountability regarding energy usage and carbon emissions, aligning with broader sustainability goals. It also helps businesses identify cost-saving opportunities through energy efficiency measures.

It requires businesses to report on Scope 1 and 2 emissions, which involve direct emissions from owned or controlled sources (Scope 1) and indirect emissions from purchased electricity, heat, or steam (Scope 2) but, it doesn't mandate reporting on Scope 3 emissions, which include indirect emissions from the value chain, such as purchased goods and services, business travel, and employee commuting.

Disclosures guidance

- Companies must report their global scope 1 and 2 GHG emissions, underlying global energy use, and a chosen emissions intensity ratio in

---

[43] SECR explained: Streamlined Energy & Carbon Reporting framework for UK business

their Directors reports for the current and previous reporting periods.

- Large, unquoted companies and large LLPs must report UK energy use from electricity, gas, and transport fuel and associated GHG emissions, including at least one intensity metric.
- The methodology must be disclosed, and the calculation of energy consumption and greenhouse gas emissions must align with government-approved methodologies and conversion factors. Companies are encouraged to go beyond the minimum requirements and include any other material source of energy use or GHG emissions outside these boundaries. The use of forward-looking science-based targets on emissions and adopting the reporting recommendations of the Task Force on Climate-related Financial Disclosures (TCFD) is also encouraged.
- Disclosures should cover the same annual period as the financial year, or an explanation should be provided as to why this is not the case. A 'comply or explain' clause excludes carbon and energy information where it is impractical to obtain it or in exceptional circumstances where disclosure would be seriously prejudicial to the organisation's interest.
- External verification or assurance of energy and carbon emission levels is recommended as a best practice to ensure data accuracy, completeness, and consistency for internal and external stakeholders.

The SECR framework applies to large UK-incorporated companies and LLPs and unquoted companies and LLPs that meet two or more of the following criteria:

- **250 employees or more**
- **£36 million annual turnover or more, and**
- **£18 million balance sheet total or more.**

Private sector entities that fall outside the scope of the SECR are also encouraged to report similarly voluntarily.

Exemptions clause

Exemptions are available for quoted or large unquoted companies and LLPs that confirm their energy use is low—40MWh or less over the reporting period. However, they must still include a statement confirming their low energy usage. Group-level and subsidiary-level reporting should include energy and carbon information for the parent group and subsidiaries. Still, they can exclude information for subsidiaries not obliged to report under SECR and organisations covered by the Climate Change Agreement (CCA) scheme.

Comments on SECR

SECR aims to enhance transparency regarding energy consumption and carbon emissions among UK businesses, facilitating better environmental performance and accountability. However, the report's compliance requirements might be too ambitious for small companies, which are encouraged to meet the report voluntarily as it might require additional resources to report the provision of the guidance. Therefore, Introducing incentives or rewards for businesses that exceed compliance requirements or achieve significant emissions reductions could enhance motivation and drive more substantial change.

Currently reporting Scope 3 emissions is voluntary, but strongly recommended. It is recommended that companies use a widely recognised independent standard, such as the GHG Protocol, ISO 14064-1:2018, Climate Disclosure Standards Board (CDSB), or The Global Reporting Initiative Sustainability Reporting Guidelines.

IFRS S1-Sustainability-related Standards and IFRS S2 - Climate-related Disclosure Standards

The International Sustainability Standards Board (ISSB) creates financial reporting standards centred on sustainability to meet investor needs. Financial reporting has historically tended to be taken more seriously by companies than sustainability reporting. However, in recent years there have been efforts to integrate more sustainability considerations into the financial reporting cycle. In June 2023, the ISSB released IFRS S1 and IFRS S2. The former sets general guidelines for disclosing sustainability-related financial data across a value chain. The latter is about the disclosure of climate-related risks and opportunities, incorporating GHG Protocol standards. IFRS S2 expands on the TCFD's foundational recommendations. Therefore, organisations that already adhere to TCFD recommendations or the UK regulations will likely find the transition to UK SDS relatively seamless, particularly regarding components that relate to IFRS S2.

These IFRS Sustainability Standards are not mandatory; jurisdictions can choose to adopt them for regulatory purposes. Countries like Turkey, Nigeria, and Brazil are already implementing IFRS S2 within their regulatory frameworks, and nations including New Zealand, the Philippines, Singapore, and Taiwan plan to do so. In the UK, in its March 2023 Green Finance Strategy, the government confirmed the Secretary of State for the Department for Business and Trade (DBT) would evaluate whether to endorse the ISSB sustainability standards. As outlined in an August 2023 guidance, the DBT aims to introduce the first two UK Sustainability Disclosure Standards (UK SDS) by July 2024. These standards will largely follow the International Sustainability Standards Board (ISSB) guidelines but will adapt as needed for specific UK issues. The UK SDS will be aligned with both the IFRS S2 and the UK's Climate-related Financial Disclosure Regulations, which are both based on the Task Force on Climate-related Financial Disclosures (TCFD) recommendations.

The Ten Point Plan for a Green Industrial Revolution Point

The Ten Point Plan for a Green Industrial Revolution, launched by the UK government in November 2020, was an initiative by the UK government to drive the country towards a sustainable, green economy that supports economic recovery post-COVID-19. The plan seeks to address climate change, promote economic recovery, and create jobs by investing in various green technologies and infrastructure. It also sets out the UK government's approach to "build back better" and accelerate the path to net zero. The UK plans to mobilise £12 billion in public funding, with expectation of significant co-investment from the private sector. The Ten Point Plan aims to create and support up to 250,000 green jobs.

ICT is not a significant theme in the Ten Point Plan. Artificial Intelligence is mentioned briefly, in the context of investment in research and innovation (“artificial intelligence for energy”). There is also a mention of “smart, digital-enabled technologies to drive competition and harness innovation for the benefit of consumers.”

Key objectives set out in the plan:

- End the sale of new petrol and diesel cars and vans by 2030 and support the electric vehicle (EV) infrastructure roll-out
- To produce enough offshore wind to power every home, quadrupling offshore wind power by 2030.
- Develop 5 GW of low-carbon hydrogen production capacity by 2030.
- Progress nuclear as a clean energy source, focusing on large-scale nuclear and developing Small Modular Reactors (SMRs).
- Support research and development of zero-emission aircraft and ships.
- Make homes, schools, and hospitals greener, warmer, and more energy-efficient, with a target of installing 600,000 heat pumps annually by 2028
- Capture 10 Mt of CO2 annually by 2030 through CCUS technology
- Support green finance and innovation to leverage private investment and drive economic growth
- Increase protection for natural environments and promote biodiversity

Green Finance Strategy

In 2019 and 2023 the UK government laid out strategies to make the UK the world’s first Net Zero Aligned Financial Centre. Jackson (2024) provides one critical angle on the overall approach:

> In 2017, the UK’s Industrial Strategy was thought to have marked an unconventional moment in British politics, as the state began to explicitly ‘pick the winners’ necessary to both grow and decarbonise the economy. [...] a more conventional view of this moment can be found in the pages of the Green Finance Strategies where, contrary to the assumption that a green transformation requires developing domestic capacity in emergent technologies, the UK has instead chosen to be the financier of them. It draws on a novel dataset to find another instance of ‘Treasury Control’, in which large scale investments in low carbon sectors was eschewed to instead 'green' the financial expertise located in the City of London, thereby making green finance British industrial policy. Instead of any industrial transformation of the UK political economy, the pursuit of green finance belies the fact that any such

> transition, and by extension Net Zero, is shaped by the City-Bank-Treasury nexus' desire to preserve the prevailing economic model with as few adjustments as possible.

AI is not a significant theme in the Green Finance Strategy documents, despite considerable interest in the use of AI for modelling and reporting within finance, and the emerging role of AI in many of the industries and contexts touched upon in the Green Finance Strategy documents.

Net Zero Strategy: Build Back Greener

The UK Net Zero Strategy, unveiled in October 2021, is a comprehensive plan to achieve zero greenhouse gas emissions by 2050. It provides the groundwork for a green economic recovery, building on the UK's 10-point strategy for a green industrial revolution. The strategy aims to keep the UK on track with impending carbon budgets and its 2030 Nationally Determined Contribution (NDC). It outlines precise policies and recommendations to cut emissions for various sectors, including energy, transportation, industry, agriculture, and residential.

The UK Net Zero Strategy's key policies aim to achieve 5 GW of hydrogen production capacity and 5 metric tons of carbon dioxide (MtCO2)/year of engineered greenhouse gas removals by 2030, decarbonize its power system, transition to low-carbon heating appliances by 2035, eliminate road emissions, start zero-emissions international travel, focus on sustainable farming, mobilise private investment for green projects, and establish the UK Infrastructure Bank to support climate-related projects.

The UK Net Zero Strategy is a plan and a pathway to a better future. Its strengths lie in its comprehensive approach, emphasis on economic growth and job creation, commitment to innovation, and clearer timelines and specific milestones. By implementing this strategy, the UK will create a sustainable and resilient future for the nation and lead the global fight against climate change.

Carbon Removals, Carbon Credits, and Beyond Value Chain Mitigation

Net zero doesn't mean we stop emitting carbon. Carbon is constantly entering the atmosphere, and constantly *leaving* the atmosphere and getting locked up in various forms (trees and soil, just for example). That would still be the case in any credible net zero scenario (and would be the case if humans didn't even exist). The IPCC, the world scientific authority on climate change, is clear that both sides of the equation are crucial for meeting climate targets. We need to release less carbon, *and* we also need to remove more carbon.

How do we remove carbon from the atmosphere? There are **nature-based solutions**, such as planting forests and restoring wetlands. Then there are more technological approaches to carbon removal. These are called **Negative Emissions Technologies** (NETs), although they have other names as well. Direct Air Capture (DAC) is one example of a NET — think of giant hair dryers blowing in reverse, drawing carbon out of the sky. Many NETs are proven to work on a very small scale, but we have really struggled to make them work on a scale where they make a significant impact. Large carbon removal projects have been plagued with problems, and fallen far short of their targets. So many of the big controversies about climate change are, at heart, how reliant we should be on carbon removals, as opposed to reducing the amount of carbon released in the first place.

A similar net zero logic applies at the company level. A company can get to net zero by reducing its emissions, and/or by balancing out those emissions (called "offsetting"). Here things get tricky. The usual way a company offsets its emissions is by buying carbon credits. Some companies are legally obliged to buy carbon credits (cap-and-trade schemes). Many other companies buy credits on the Voluntary Carbon Markets.

What does purchasing a carbon credit actually mean? Buying a carbon credit supposedly worth 10 tonnes of carbon doesn't necessarily mean you have really removed 10 tonnes of carbon from the atmosphere. Some carbon credits can be, shall we say, very dodgy. Some credits are based on "avoided" emissions: in essence, paying somebody not to release carbon, and counting the avoided carbon as though it had actually been removed. There can be other issues with additionality, double-counting, and durability. Some forests burn down. Others don't exist in the first place. We talk about 'high quality' and 'low quality' carbon credits, and principles and standards such as the Oxford Principles, VCMI, and ICVCM aim to codify what makes a good carbon credit. But even the highest quality carbon credit still comes with a degree of uncertainty.

The relationship of carbon credits to social justice is complex. An argument in favour of carbon credits is that they allow finance to flow to the Global South, where carbon removal tends to be cheaper. This is true to some extent, although the levels of investment are nowhere near as high as they need to be — which could be taken to suggest that the system isn't working well enough.

One promising approach to carbon credits is associated with **Beyond Value Chain Mitigation**. On this model, companies can still buy carbon credits to invest in high-quality carbon removals projects. They may invest in a range of climate projects based on all kinds of criteria — perhaps they want to prioritise projects that seem exciting and innovative, or that benefit communities in the regions where they operate. They can report proudly on these investments, for the benefit of customers, investors, regulators, employees, and other stakeholders. But crucially, companies won't translate these investments into a simple number (carbon tonnage) to be subtracted from their carbon footprint. That kind of arithmetic has always been deeply misleading, and years of trying to make it work have borne little fruit. Instead, **carbon credits purchased and a company's own carbon footprint are reported separately**, side-by-side. This would imply an important shift in the way companies report on their carbon footprints, moving away from presenting a consolidated figure — great for building misleading league tables, and incentivising greenwashing — to something much more qualitative and narrative, backed up with lots of quantitative evidence.

Amazon, Google, and Microsoft

In these sections, we won't attempt a comprehensive or consistent evaluation of the three cloud giants, nor try to rank them. This would be out of scope for this report. We do offer some closing observations, context, and signposting.

Amazon dominates the cloud market, followed by Microsoft, with Google also representing a substantial proportion.

A 2024 report by Horwood in *Computing* does attempt to rank the three cloud giants on sustainability (Horwood 2024a, 2024b). It suggests that all three are improving somewhat in terms of transparency:

> Google has made considerable progress in data quality and transparency. It breaks datacentres out from the rest of its operation, breaks down data by individual datacentre, and is even trying to get to grips with the environmental impact of third-party datacentres. [...] Microsoft too breaks down data by region, and Amazon continues to innovate and enable renewable energy projects on an extraordinary scale. (Horwood, 2024a)

Transparency is commendable. There are also undoubted efficiency gains in many areas. Nonetheless, the rapid growth in data centres and network infrastructure paints a stark picture. Policy solutions are needed, and the more progressive elements within the cloud giants need to work to make apparent the scale of the problem, and the inadequacy of existing approaches. There is pressure from some quarters to rewrite the rules to disguise failed climate pledges. Environmentalists and civil society actors also need to do all they can to hold the cloud giants accountable and to drive a step change.

In 2024, the sustainability reports of Amazon, Microsoft, and Google all attracted interest within industry and mainstream media, for poor environmental performance linked to the growth of AI. It is clear that AI, particularly GenAI, has made a negative impact. But it is also clear that these problems are not entirely new. None of the cloud giants has been on track to meet milestones for a rapid climate transition, in alignment with IPCC reports and the Paris Agreement, despite their ambitious pledges.

Adrian Cockcroft, writing for *The New Stack*, offers his analysis of the three sustainability reports, and offers some guarded optimism, pointing out for example that the "Amazon generation portfolio increased from 20GW to 28GW, making it the world's largest purchaser of renewable energy for the fourth year in a row" (Cockcroft 2024).

Amazon

AWS is the biggest of the cloud giants, with about a third of the market share. In 2019 Amazon pledged to achieve net zero by 2040. Amazon created a $2bn fund to invest in the development of sustainable technologies. By launching **The Climate Pledge** and inviting other companies to become signatories, Amazon sought to position itself as a leader in corporate climate responsibility. “We want to use our scale and our scope to lead the way,” founder Jeff Bezos commented when he announced the pledge. A year later, the company also committed to submitting its goals for verification through the Science Based Targets initiative (SBTi), the United Nations-backed entity that validates net-zero plans.

Despite its goals, **Amazon's emissions have increased by nearly 40%** since announcing its net-zero target. In response to this rise and a failure to submit policy-compliant targets, the Science Based Targets initiative dropped Amazon from its list in 2023. Soon after the Climate Pledge, two organisers of Amazon Employees for Climate Justice were fired (a move the National Labor Relations Board determined was illegal retaliation). Corporate Responsibility Monitor (2023) rates Amazon’s net zero pledge as ‘low’ for integrity and ‘very low’ transparency. This compares with moderate ratings for Google and moderate/reasonable ratings for Microsoft. In the same report, Amazon is listed as one of several companies which “provide very limited information on the renewable energy they procure, which makes it impossible to assess the integrity of their renewable energy claims.” Amazon’s net zero approach which, similar to the other cloud giants, relies heavily on “non-permanent and scarce forestry-based carbon dioxide removals,” is described as a cheap and implausible strategy. CDP, which gave Google and Microsoft “A” ratings for their questionnaire response, rated Amazon a “B” (CDP, 2023).

Amazon’s 2023 sustainability report (released in July 2024) noted a 13% decrease in **carbon intensity** between 2022 and 2023, although its **absolute carbon emissions** have increased by nearly 40% since 2019 with these modest carbon efficiencies far outweighed by its overall growth. Carbon intensity is measured by grams of carbon dioxide equivalent (CO2e) per dollar of gross merchandise sales (GMS). Thus, with continuously growing operations, a drop in carbon intensity can be a confusing and misleading metric to highlight, as it may not indicate a significant (or indeed any) decrease of an organisation’s carbon footprint overall: indeed, while between 2022 and 2023 Amazon’s overall carbon footprint reportedly decreased by 3% (from 70.74 to 68.82 to MMT CO2e), it had previously risen significantly from a value of 51.17 MMT CO2e in 2019. Unfortunately, it gets worse. Amazon has also designed a **carbon accounting** methodology which disguises the full scale of its environmental impacts.

For example, *Reveal* (2022) reveals:

> [...] while Target, Walmart and The Home Depot estimate the gas their workers burn while driving to and from work, Amazon counts emissions only from its own corporate shuttles. Asked about the discrepancy, Amazon's Davila said: "Employees can use public transportation to get to the office, and if they live nearby, they can walk or bike."
>
> (Evans, 2022)

Likewise, whereas major competitors include carbon emissions from products bought directly from manufacturers and sold to customers, Amazon only counts these if they are Amazon branded: products which account for about 1% of its sales. In general, Amazon's carbon accounting methodology seems to have been designed to downplay Amazon's climate impacts, rather than to maximise consistency and comparability.

Amazon has tried to put a positive spin on things by claiming to have met its 100% renewable energy matching target early (based on the discredited practice of purchasing **unbundled Renewable Energy Certificates**: see 'Green Data Centres'), and by emphasising **carbon intensity** rather than total carbon emissions. More details about Amazon's carbon accounting methodology can be found in Amazon's Carbon Methodology (2022). Amazon has lobbied against the shift to 24/7 hourly matching in the ongoing review of the Greenhouse Gas Protocol, proposing instead greater globalisation of renewable energy certificates, and greater emphasis on avoided emissions.

For *Computing*, Horwood (2024a) writes:

> Amazon's reporting presents a narrative of slightly reduced overall footprint despite much higher company growth. Like its closest cloud competitor, Amazon's emissions have grown; the impression of a reduction has been enabled by means of offsetting, mainly of scope 2 emissions.

What about AWS specifically? It is hard to be sure, since Amazon doesn't provide very granular data. However, the campaign group Amazon Employees for Climate Justice offer some rough estimates:

> [...] if current trends hold, NVIDIA will ship 1.5 million AI servers per year by 2027 — and that those servers will consume at least 85,400 GWh of electricity annually. Amazon won't necessarily be a prime customer of NVIDIA's, since it is heavily investing in its own AI-specialized chips for its servers, but we can use the estimated data about NVIDIA — the top global supplier currently — as a proxy, since Amazon doesn't share this data on its own AI chips. If we assume AWS's market share will remain similar through 2027, then Amazon could be deploying 465,000 new AI servers per year in the next few years. That's a potential for at least 26,500 GWh of additional energy consumed by Amazon's AI servers per year, which is nearly 70% of Amazon's current total electricity usage. (AECJ 2024).

Focusing on Scope 3 emissions in 2022, Horwood (2024b) adds:

> AWS comes out of the analysis badly for several reasons. The first is that Amazon doesn't break out AWS from the rest of its operation. Whilst Microsoft and Google don't do that fully, both provide some metrics which are specific to cloud infrastructure. [...] Amazon presents a picture of a slightly reduced overall carbon footprint. However, an appendix to the main report where carbon accounting is located gives location-based carbon emissions (before any offsets are applied) as 56,509,397 mtCO2e. After offsets are applied a market-based measure of 54,977, 815 mtCO2e is given. Amazon rounds this up to 54.98mmtCO2e (million metric tons) in the main body of the report. [...] Given that the location-based emissions FY22 were higher this year than market adjusted totals in FY21 it seems likely that the amount of carbon indirectly put into the air grew. To present this as a decrease requires an uplift in offset purchasing. [...] The lack of location-based data from FY21 enables a certain sleight of hand in Amazon's carbon accounting [...]

AWS has also been criticised by customers as the worst cloud giant for providing good granular data on the sustainability of their cloud operations.

AWS released its **Customer Carbon Footprint Tool** in 2022 which promised to support customers on their sustainability journey by providing:

> easy-to-understand visualizations to show customers their historical carbon emissions, evaluate emission trends as their use of AWS evolves, estimate the carbon emissions they have avoided by using AWS instead of an on-premises data center, and review forecasted emissions based on their current use (Amazon, 2022).

The tool works through the cloud provider's customer billing console, on the basis of comparing AWS carbon emissions across several geographic regions with a range of "surveyed enterprise data centres," in metric tons of carbon dioxide equivalent (MTCO2e). None of the tools offered by the three cloud giants is perfect, but AWS's tool is lagging behind the others. According to Targett (2022) "[t]he new tool only shows emissions data by extremely high-level geographical groupings such as EMEA (Europe, Middle-East and Africa) and AMER (North, Central and South America) not by AWS Regions themselves; a lack of precision that may frustrate some users hoping to optimise emissions reductions".

Furthermore, while Microsoft and Google have enabled their cloud customers to track Scope 1 to Scope 3 emissions since 2021, Amazon's approach still lacks the capability to provide comprehensive Scope 3 emissions data. This omission is significant, as Scope 3 emissions typically account for the majority of a cloud provider's total emissions.

Since early 2022, AWS has provided customers with Scope 1 and Scope 2 emissions via its Customer Carbon Footprint tool. Scope 3 data was promised from early 2024:

> the company confirmed Scope 3 data will be possible to track through the tool from early next year. 'We are conducting robust lifecycle assessments for our business to provide customers with high-quality Scope 3 carbon emissions data,' said an AWS spokesperson. 'We will incorporate this data into the Customer Carbon Footprint Tool in early 2024.' (Donnelly (2023))

At time of writing in mid 2024, Amazon's Customer Carbon Footprint Tool user guide refers only to scope 1 and scope 2 emissions:

> Carbon emissions data in the customer carbon footprint tool adhere to the Greenhouse Gas Protocol and ISO. Carbon footprint estimates for AWS include Scope 1 (emissions from direct operations) and Scope 2 (emissions from electricity production) data.

AWS does already offer a Scope 3 workaround for its largest clients. According to Adrian Cockcroft, former Vice President of Sustainability Architecture at AWS, the company can provide Scope 3 emissions data under non-disclosure agreements. "If you really need Scope 3 [data] from AWS and you're a big enough customer, you can escalate and they [will provide] estimates to people under NDA, but that's true of pretty much anything you want custom from AWS," Cockcroft revealed in March 2023.[44]

Cockcroft (2023a) outlines the shift from Amazon's 2021 sustainability figures which described 13 AWS regions as using "over 95% renewable" energy, to the 2022 report where 19 AWZ regions are listed as 100% renewable energy. Cockroft links these shifts (from 13 to 19 regions and 95% to 100%) predominantly to "a large increase in dedicated power purchase agreements to over 20GW, and some use of biofuels for backup generators." AWS also employs **market-based accounting** to claim zero Scope 2 emissions in some regions. Yet AWS's activities in these regions have demonstrably added carbon to the atmosphere and contributed to heating the climate. Market-based accounting, while compliant with current standards, offers a potentially skewed portrayal of actual emissions reductions. A better approach would be to include both the market-based data and **location-based data** (see Green Energy Procurement).

More recently, Cockcroft offers this update:

> The AWS Customer Carbon Footprint Tool (CCFT) was embarrassing when it was initially released in 2022, and it has made no progress in the years since then. It is going to report zero for everyone for their Scope 1 and 2 carbon footprint, according to the market methodology, has no Scope 3 data, and aggregates too much data together.
>
> I recently tried to use the CCFT data to track progress for a company, and the three usage categories EC2, S3 and Other combined with three

---

[44] Quoted in Caroline Donnoley, 'AWS Under Fire for delays in delivering Scope 3 GHG emissions data to enterprises and governments', *ComputerWeekly.com*, April 2023.

> geographies EU, Americas and Asia made it impossible to figure out what was going on. The Other category in the Americas is dominating, but that's all it tells you.
>
> Customer carbon tracking tools from GCP and Azure give you all the details you need, but with AWS you can't see which region or what service. Escalating to AWS support has produced nothing. Completely useless.
> (Cockcroft 2024).

The Greenhouse Gas Protocol (GHG Protocol), which guides corporate emissions reporting, currently allows for ambiguity in market-based versus location-based calculations for Scope 2 emissions. Revisions of the GHG Protocol, aims to clarify this and other issues, and potentially enforce more stringent reporting requirements. There is a great deal of politicking around the GHG Protocol, with Amazon and Meta spearheading an approach at odds with that of Google and Microsoft. Watch this space.

Targett (2022) further quotes AWS users who describe the tool as lacking transparency and granularity, including insufficient data on embodied emissions as well as insufficient details on methodology. According to a DEFRA sustainable IT professional quoted by Targett (2022), the carbon 'savings' element on the tool's dashboard also has tended to overemphasise AWS' sustainability credentials:

> We asked AWS to share the working, calculations, and models behind their figures and they stated they would share as much as they could. We also stated that the element on the dashboard that describes typical saving from moving from to the cloud was very 'salesy' and not based on facts i.e it was worst-case public sector data center, moving to best-case AWS.

Aiven and Thoughtworks (2024) also mention that the data AWS provides is relatively out-of-date:

> [...] Google reports comprehensively on all three Scopes of the GHG Protocol through its Cloud Carbon Footprint dashboard. However, this has a time lag of up to 21 days for access to the preceding month's data. Both AWS and Microsoft Azure focus exclusively on users' Scope 2 emissions and emphasize the indirect emissions that stem from the electricity generated to power their data centers. AWS also has a three-month delay in displaying cloud emissions data through its own Customer Carbon Footprint tool.

Alternative tools have emerged, both open source and commercial, from the likes of Cloud Carbon Footprint, Boavizta, and Greenpixie, have emerged. These tools strive to provide more accurate and inclusive emissions data, by developing independent measurement methodologies which are in line with the GHG protocol and address the gaps left by AWS. Greenpixie's methodology has been verified under ISO-14064 as compliant with the GHG protocol.

Amazon is pursuing various sustainability improvements and innovations, such as low carbon concrete for data centres, and switching to Hydrotreated Vegetable Oil for backup generators in some data centres (we don't attempt to catalogue or evaluate all these initiatives here). Cockcroft (2023a) writes:

> AWS still doesn't report scope 3 to customers, although they've said they are working on it. However they do talk in the report about using low carbon concrete and steel in many of their most recent datacenter construction projects. The Power Usage Efficiency of datacenters (how much energy is used for cooling etc.) is reported regionally by Azure, Google reports somewhat better figures by datacenter campus, and AWS doesn't report PUE, but is likely to be similar.

Amazon's Sustainability Data Initiative (ASDI) seeks to drive sustainability innovation by reducing the cost of acquiring and analysing large sustainability datasets. ASDI is currently offering a cloud grant scheme for using AWS to explore big sustainability challenges based on this data set. Amazon is of course pursuing many kinds of AI research as well. Amazon's Titan Text scored the lowest (12%) in the 2023 Foundation Model Transparency Index. Amazon is also a major investor in OpenAI rival Anthropy (makers of Claude). For a glimpse of other sustainability-related goings-on at Amazon, see this round-up of AWS re:Invent 2023 (Cockcroft 2023b).

Google

How about Google? Google Cloud Platform currently serves about 10% of the cloud market. What climate-related claims is Google making about their own business as a whole, and specifically about GCP? How transparent are they about their data centres? What tools do they offer to users of GCP to improve the sustainability of their cloud usage, and how have these tools been received? What innovations are they touting in the sustainability space, and how do these hold up to scrutiny?

The Corporate Climate Responsibility Monitor 2023 comments, "Google's headline pledge is to reach 'net-zero emissions' by 2030 while keeping its continuous goal of 'carbon neutrality' each year. We consider both claims misleading as they are not substantiated with deep emission reduction commitments" (see sidebar for more).

According to its most recent sustainability report, Google's greenhouse gas emissions have increased by 48% over the past five years. In 2023, the company's GHG pollution amounted to 14.3 million tonnes of CO2e, reflecting a 48% rise from its 2019 baseline, and a 13 percent increase since the previous year (Google, 2024).

Google indicated that this worsening performance underscores "the challenge of reducing emissions" while investing in the development of Large Language Models and other AI. Google claims that the "the future environmental impact of AI" is "complex and difficult to predict."

Kate Brandt, Chief Sustainability Officer, stated that the company is still committed to the 2030 target but, much like Microsoft, appears to be preparing the ground for failure, by emphasising the "extremely ambitious" nature of the goal, and indicating that emissions will continue to rise for now. "Ultimately this isn't just about Google," says Google, in Google's most recent sustainability report, which is about Google (Google, 2024).

Statements like these **fail to reflect the importance of *cumulative emissions***. To a casual reader, it may sound like meeting the 2030 net zero target is all that matters—rather than the total emissions between now and whatever date the company achieves and sustains net zero. Even if Google somehow does manage to hit the 2030 target through a late, steep dip, this is not the same as having made steady year-on-year progress to the 2030 target.

Writing for *Computing,* prior to Google's most recent sustainability report, Horwood (2024a) writes:

> [Google] has cut its scope 3 carbon emissions, to less than they were in 2018, despite growing as a business. Crucially, the company managed to reduce its location-based scope 3 emissions. [...] Google was less successful at reducing its scope 1 (which doubled due to the company including a source of emission that hadn't previously been included) and scope 2 emissions, which are primarily made up of emissions related to electricity consumption.

Focusing on Scope 3 emissions specifically, Horwood (2024b) was positive about Google in relation to the other cloud giants:

> All suppliers are required to sign a Code of Conduct and are assessed and audited. All are expected to report environmental data and to submit data to CDP. In 2022 Google invited 222 suppliers to respond to CDP and 90% of them did. Google also hosts a Supplier Sustainability summit where targets are set and training is provided. Whilst the mechanisms for "deep supplier decarbonisation" will of course benefit Google by reducing their indirect emissions, it is still a leading example of a powerful company using its power to raise the standards of environmental reporting and to reduce emissions.

A company of Google's scale is certain to work with many more than 222 suppliers, however — one source, which refers only to manufacture of components for data centre infrastructure, mentions over 500 suppliers.

More recently, however, Allen (2024) reports for *Computing*:

> Although datacentres represented the majority of Google's Scope 2 emissions, Scope 2 as a whole was only 24% of Google's total emissions. Scope 3 (indirect) emissions, from the up- and downstream supply chain, are a far larger source: 75% of the company's overall emissions, or 10.8 million tons of CO2e.
>
> While Scope 2 emissions rose more than Scope 3 — 37% compared to 8%

> year-on-year — the massive difference in size between the two categories meant that more emissions were added to Scope 3 as an absolute measure.
>
> Again, AI is the culprit. Google says, "We expect our Scope 3 emissions will continue to rise in the near term, in part due to increased capital expenditures and expected increases in our technical infrastructure investment to support long-term business growth and initiatives, particularly those related to AI."

Like the other cloud giants, Google is making efforts to innovate in energy efficiency and clean energy, for example adding a novel form of geothermal energy to its energy portfolio. Google's Clean Transition Tariff is a prototype financial model which Google believes will help the transition to 24/7 carbon free energy if it is adopted more widely.

Google, like Microsoft, is a signatory of the 24/7 Carbon Energy compact, and have advocated for revisions to the GHG Protocol to reflect "more geographically and temporally granular scope 2 accounting" (Google 2023). Google reported a 64% carbon free energy rate on an hourly basis for 2023, stating that this rate had been maintained from the previous year despite a total electricity load increase across all data centres by 3.5 TWh.

> This is due to both an increase in Contracted CFE (up by roughly 1.2 TWh, or 9%, from 2022) as well as improvements in overall Grid CFE. We've worked hard to continue advancing CFE in parallel with load growth across our data center portfolio. In 2023, 10 of our 44 grid regions achieved at least 90% CFE (Google 2024).

Notably, there are vast regional variations across Google's data centre operations globally, ranging from 0% in Saudi Arabia and Qatar, 4% in Singapore, 26% in the US State of Nevada, 43% in Ireland, through to 92% in Great Britain, 98% in Finland and 100% in Canada for the Hydro-Quebec powered operations (Google 2024).

Google Cloud offers customers a suite of tools to improve the environmental impact of their cloud purchases:

- Google Carbon Footprint
- Google Cloud Region Picker
- Google Active Assist

These tools are squarely within a **carbon-aware computing** paradigm, e.g. location-shifting, time-shifting, and demand-shaping (see 'Carbon-Aware Computing and Grid-Aware Computing').

As an indicator of Google's own efforts at responsible AI development, Google's PaLM 2 foundation model was in the middle of the pack (40%) in the 2023 Foundation Model Transparency Index. Like the other cloud giants, Google promotes the potential benefits of AI for sustainability (including fuel-efficient routing in Google Maps, extreme weather prediction, and minimising contrails from aviation), and highlights the potential for innovation and optimisation to improve the sustainability of AI:

> Making AI computing more efficient requires using proven methods to cut

> emissions, while also uncovering new ways to increase efficiency. To minimize the carbon footprint of AI workloads, we rely on tested practices that can reduce the energy required to train an AI model by up to 100 times and reduce associated emissions by up to 1,000 times. To support the next generation of AI advances, our Tensor Processing Units v4 is proven to be one of the fastest, most efficient and most sustainable ML infrastructure hubs in the world. Additionally, our data centers, where this AI computing takes place, are designed, built, and operated to maximize efficiency. A Google-owned and -operated data center is on average more than 1.5 times as energy efficient as a typical enterprise data center, and the average annual power usage effectiveness (PUE) for our global fleet of data centers was 1.10, compared with the industry average of 1.55. (Google)

In general, Google's public communications around AI and sustainability tend to fail these minimal transparency tests:

- Sustainability harms should be weighed up with sustainability benefits
- Sustainability benefits through improved efficiency should be weighed up with sustainability harms inflicted through growth
- Uncertainties should be identified and where possible quantified

For example, a paragraph like the following (from Google's most recent sustainability report) gives the impression that growing processing requirements of AI is a necessary and inevitable problem, rather than the outcome of human action and inaction, but that this problem will likely soon be solved through innovation.

> State-of-the-art approaches to developing AI models are varied and evolving, but one thing is clear: the desire for more precise and accurate model outputs has been leading to more complex models that rely on larger sets of training data and require more processing power. [...] These more complex models may lead to higher energy consumption, all other factors being equal. Nonetheless, it is important to note that AI model design is an evolving field, and new releases and versions of complex models consistently demonstrate improved energy efficiency while maintaining model performance. Indeed, ongoing improvements in software and algorithmic optimization have the potential to significantly enhance efficiency and decrease computational requirements. (Google 2024)

The authors then change the subject without following up to answer (or at least to pose) the important questions: is there any evidence that "improved energy efficiency while maintaining model performance" will balance out "more complex models that [...] require more processing power" quickly enough to align with a rapid transition to net zero? If not, do the social benefits of "more precise and accurate model outputs" outweigh the social harms of wildfires, droughts, famines, heat waves, hurricanes, rising sea levels, loss of animals and plants, ocean acidification, displacement of populations, climate grief, and the other social, economic, cultural, and ecological impacts of climate change? These are not rhetorical questions: the cloud giants are well-resourced entities which draw on wide and deep networks of scientific expertise. These are their sustainability reports. They could show their workings.

Corporate Climate Responsibility Monitor on Google

The *Corporate Climate Responsibility Monitor 2023* went into some depth on Google. It criticised Google's **exclusion of scope 3 emissions**, its use of **market-based accounting** for reporting on its renewable energy purchases, and its use of **low quality carbon credits to offset its emissions**.

> "Google's headline pledge is to reach 'net-zero emissions' by 2030 while keeping its continuous goal of 'carbon neutrality' each year. We consider both claims misleading as they are not substantiated with deep emission reduction commitments. The 'carbon neutrality' claim excludes major scope 3 emission sources that accounted for 58% of the company's GHG emissions in 2021 (Google, 2022b, p. 11). Emission sources covered by the target are 'neutralised' through procurement of renewable energy and offset credits that have highly contentious environmental integrity (see Table 3-2, Section 3.2.2)."

The report suggested that **Google's approach to carbon offsetting was out of date**:

> "The provided guidance of what Google perceives as 'high-quality' offsets dates back to 2011 (Google, 2011). Google could update that guidance to the newest available standards to improve the integrity of its 'neutralisation' claims."

The report also read between the lines to identify **heavy reliance on carbon credits** going forward:

> "Google's net-zero target for 2030 covers the company's entire operations and value chain emissions. In 2022, Google clarified that this target entails a 50% reduction of its market-based emissions across all three scopes by 2030 compared to 2019 baselines. This translates to an effective emission reduction of 37% of Google's emissions using a location-based accounting method (Google, 2022a, p. 5). This is an improvement compared to a year ago, when the company had not made any commitment alongside its net-zero target (Day et al., 2022). However, a 37% reduction commitment implies that Google will claim to 'neutralise' the majority (67%) of its real emission footprint with carbon offset credits by 2030 or potentially other creative accounting methods. It remains unclear what the portfolio of offset projects will look like, as Google

provides only limited information on this. Google acknowledges that a shift towards carbon dioxide removal credits is required to align with the ambition set out in the Paris Agreement but also claims that in the short-and medium-term those credits are not economically feasible at scale (Google, 2022a, p. 12). Google does not transparently disclose whether these carbon dioxide removal measures will be based on biological, geological or mineral carbon storage. Google plans to use 'avoided emission credits' until carbon dioxide removals become available at scale."

'User / Chimera' by Clarote

Microsoft

Microsoft Azure currently holds around a quarter of the market for cloud services. A 2022 *Forbes* article offers glowing praise of Microsoft's climate leadership:

> In 2012, Microsoft became carbon net zero. In 2020, it decided that this wasn't good enough and announced it would be carbon negative by 2030. If you don't think this is audacious, then listen to this. By 2050, Microsoft says it will remove the equivalent of all of the Scope 1 and Scope 2 emissions it had emitted since it was founded in 1975. Everyone, even Microsoft, recognizes that this is a "moonshot." (Bansal 2022)

Brad Smith, Vice Chair and President at Microsoft Corporation, recently offered these ominous words: "In 2020, we unveiled what we called our carbon moonshot. That was before the explosion in artificial intelligence. So in many ways the moon is five times as far away as it was in 2020, if you just think of our own forecast for the expansion of AI and its electrical needs."[45] It seems the goalposts have already shifted: what Smith is referring to here is achieving net negative, or even just net zero, by 2030.

The term **"moon shot"** is an ambiguous one. It can mean an ambitious and well-resourced venture which will produce significant results. It can also mean a brave long shot — something that is worth trying, but not blameworthy if it fails. Smith's remarks are ambiguous, but they do strongly imply priorities. The expansion of AI and electrical needs comes first, and doing this in a way which respects climate targets is a nice-to-have. It's not Microsoft's fault that the moon is running away.

Microsoft has claimed to be carbon neutral since 2012, and has pledged to achieve net zero, to be net negative by 2030 for all three scopes, and to remove its legacy emissions by 2050.[46] Microsoft also has some associations with sustainability more generally, for example via the Bill & Melinda Gates Foundation. Recently Microsoft's approach to developing and deploying AI has drawn fierce criticism from sustainability and climate transition perspectives. Ties to oil and gas have also continued to draw criticism, including from many Microsoft employees.

Despite its climate pledges, Microsoft's 2024 sustainability report revealed that GHG emissions in 2023 were 29.1% higher than its 2020 baseline. While scope 1 and 2 emissions are reported to have decreased by 6.3% from the 2020 baseline, the company's scope 3 emissions rose by 30.9%, attributed at least in part to the company's drive to expand its global share in data centres, which in turn links with the demands of its growth in AI services (Donnelly, 2024). Building more data centres increases embodied emissions both in building materials as well as hardware components. Thus, both Microsoft and Google's emissions have

---

[45] 'Akshat Rathi and Dina Bass, Microsoft's AI Push Imperils Climate Goal as Carbon Emissions Jump 30%.' Bloomberg, May 2024. <https://archive.is/V0iwb>

[46] Brad Smith, 'Microsoft will be carbon negative by 2030' (2020), <https://blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/>

increased through data centre expansions fuelled by their race to maintain dominant positions in the AI market share. In terms of cumulative emissions, Microsoft is already on the brink of exhausting the carbon budget it committed to in 2020.

Industry experts quoted in a recent *Computer Weekly* article (Donnelly, 2024) suggest that this increase in emissions linked to AI, a trend likely to continue over the next years, would have likely been anticipated by the companies - however, in a competitive market, this was a secondary concern:

> "They know [AI] is a way to lock customers in and make more money, and also that customers are going to ask for it regardless, so they have to offer [AI] or they'll go somewhere else [...] They also know it's going to increase emissions. They are simply choosing the money over the emissions." (Stephen Old, head of FinOps at independent software licensing advisory Synyega, cited in Computer Weekly [Donnelly, 2024])

The term "carbon neutral" is a misleading term, but Microsoft can at least be commended for being fairly transparent about its meaning. "Carbon neutral" means that any CO2 emissions released by Microsoft are balanced out by an equivalent amount of CO2 being removed from the atmosphere, or being avoided elsewhere, for example by paying someone not to cut down trees. However, carbon neutral does not mean that the company is no longer responsible for heating the climate. It's about balancing the books. The Science Based Targets initiative's definition of **net zero** would be a better alternative.

First, companies can claim to be carbon neutral even if they have not decarbonised their entire value chain. In 2023, Corporate Climate Responsibility Monitor suggested:

> Both Microsoft and Google currently claim to be 'carbon neutral' while only covering 2% and 12% of their full emission print with these claims, respectively. By 2030, Microsoft claims to become 'carbon negative' and Google claims to reach 'net-zero' emissions, covering their full emission footprint.

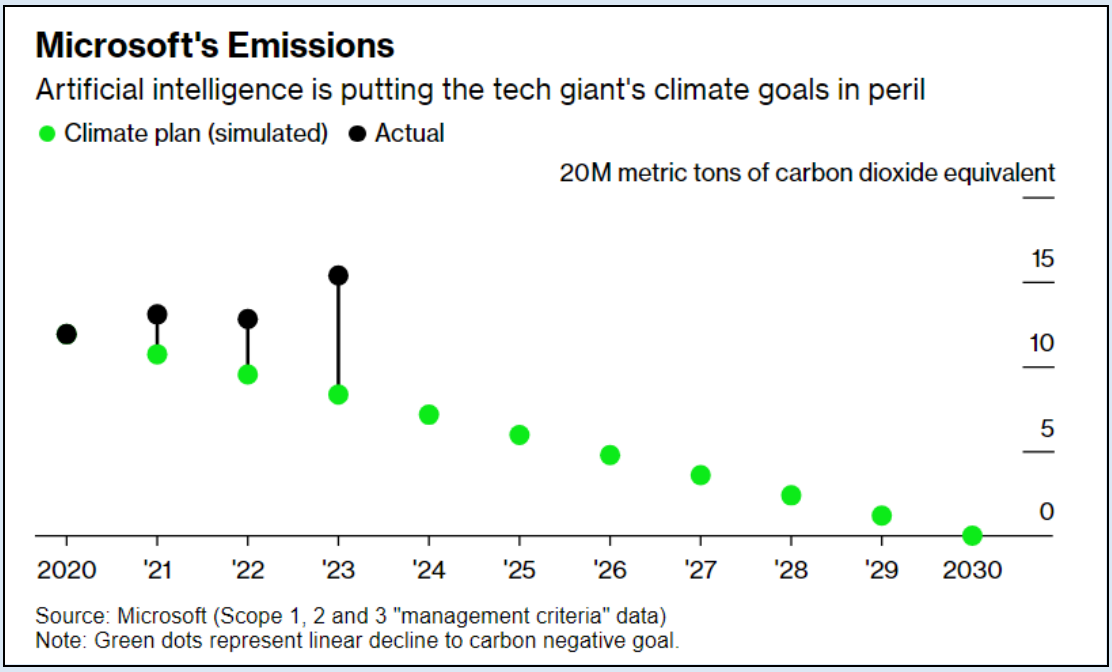
Second, although both carbon neutrality and net zero can be achieved using offsetting, net zero is somewhat more rigorous about the quality of the offsetting. Carbon neutrality permits the use of "avoided" emissions: in essence, paying somebody not to release carbon, and counting the avoided carbon as though it had actually been removed. For more details on the difference between carbon neutrality and net zero, see The Carbon Trust, SBTi, and PAS 2060, as well as the Cloud Governance Glossary in this report. The casual observer might well not realise "carbon neutral" is a technical term, let alone guess that a carbon neutral operation might well be adding carbon to the atmosphere.

So wouldn't it be better to drop the term "carbon neutral" altogether? It looks like this *may* be happening, although backsliding is always possible. The 2023 Corporate Climate Responsibility Monitor pointed out, "For the consumer it is difficult to distinguish the difference between 'carbon neutral' and 'net-zero' and make informed choices based on that information." It could send a good signal, if Microsoft were to discontinue boasting about carbon neutrality entirely. The 2024 Corporate Climate Responsibility Monitor report suggested that there has been some gradual progress in relation to such claims:

> Google and Microsoft – both of which received a *poor* rating for the integrity of their carbon neutrality claims in the 2023 Corporate Climate Responsibility Monitor – also appear to be quietly moving away from these claims, even though both companies still appear to procure carbon credits equivalent to their scope 1 and 2 emissions.

In 2020 Microsoft made a welcome **climate pledge**, although it has fallen short of delivering each year since the pledge was made, and its most recent year is particularly poor. Microsoft’s pledge included shifting to a 100 percent supply of renewable energy by 2025, reducing Scope 3 emissions by over half by 2030, and becoming carbon negative by 2030 (i.e. removing more carbon than it emits). The pledge also includes cleaning up after itself: Microsoft also pledged that by 2050, it will have removed all Scope 1 and Scope 2 carbon emitted since its founding in 1975. (As with many companies, this is tiny in comparison with its Scope 3 emissions; there is a question around who would be best placed to clean up those upstream emissions: Microsoft, the suppliers, somebody else?).

Importantly, Microsoft also pledged a linear **net zero pathway**: milestones like 2025 and 2030 can be useful to galvanise action, but it is really *cumulative* emissions that matter to the planet. It is against this pathway that we can say **Microsoft is on the verge of violating its pledge this year**, by exceeding the carbon budget of approximately 65 million metric tonnes of CO2e that it committed to in 2020. This issue can be clarified using *this chart from Bloomberg*.



Source: Microsoft (Scope 1, 2 and 3 "management criteria" data)
Note: Green dots represent linear decline to carbon negative goal.

(Rathi and Bass, 2024)

Imagine a hypothetical situation in which Microsoft continued to increase emissions until 2028 or 2029, and then rapidly decarbonised to net zero in 2030. Would it have met its climate pledge? It would not in any meaningful sense have delivered on its 2020 promise, because the area under the black dots would be significantly larger than the area under the green dots (see also 'Tech's pathway to net zero').

How does Microsoft hope to achieve its climate ambitions? In 2020 Microsoft committed to a $1 billion Climate Innovation Fund to invest in new decarbonisation technologies and projects, mentioning climate equity considerations in the use of that fund. The announcement made mention of afforestation and reforestation as well as Negative Emissions Technologies (NETs) including potentially soil carbon sequestration, bioenergy with carbon capture and storage (BECCs) and direct air capture (DAC). In building up its suite of carbon removal projects, Microsoft affirmed it would be guided by criteria of (1) scalability, (2) affordability, (3) commercial availability, (4) verifiability. Other criteria which plausibly *might* have appeared in the 2020 announcement might include stable long-term carbon storage, consistency with preserving and enhancing biodiversity, consistency with human rights and social justice, and independent assurance. Some elements of these have appeared subsequently.

Is this pledge really as ambitious as it seems? Corporate Climate Responsibility Monitor 2023 points out:

> "[...] Microsoft uses a GHG emission accounting method to make achieving this target [carbon negative by 2030] easier without having to substantially decarbonise its purchased electricity. While Microsoft tracks both its location- and market-based electricity consumption emissions, only the market-based values are included in its aggregated emissions disclosure and target coverage [...] Microsoft's location-based emissions are almost six times higher than its market-based emissions. As a result, the 2030 target only covers around 76% of Microsoft's full emission footprint. With half of these emissions set to be offset, Microsoft's 2030 target entails a commitment to an emission reduction of just 38% of its 2019 location-based emissions (50% of its market-based emissions)" (105).

Furthermore:

> "Location-based scope 3 emissions account for over 75% of Microsoft's total emissions in 2021. Most of these emissions stem from the extraction of raw materials and manufacturing of products used for datacentres and hardware products, as well as from the use of Microsoft's hard and software products (Microsoft, 2022a, p. 98). Increasing demand for data centre services may drive an increase in scope 3 emissions in the coming years" (105).

Microsoft has recently signalled some support for 24/7 hourly matching, which has potential to mitigate some of these issues. It has not been quite as active in advocating for this approach as Google. Amazon has opposed 24/7 hourly matching.

Microsoft's transparency on this issue makes this analysis relatively easy to do. The company's good intentions are not in doubt, but we should be clear that the company is pumping out huge amounts of carbon into the atmosphere each year, while offsetting a relatively small proportion through low-durability projects. In practice, the majority of the efforts to remove carbon since the pledge have involved afforestation and reforestation, through contracts with companies like Forestland Group and Natural Capital Partners. Fossil fuels are a fairly stable form of carbon storage, whereas forests are a more fragile form, especially as the climate heats and wildfire risks increase.

Microsoft's 2023 carbon removals white paper also includes the ominous phrase "Limiting warming this century to anything close to 2ºC."

> Microsoft continues to plan for a portfolio of greater than 5 million metric tonnes of carbon removal per year in 2030. We're committed to a portfolio that balances relatively proven low-durability, nature-based solutions with medium- and high-durability solutions, where low-durability options face perhaps the greatest qualitative challenges and the high-durability opportunities need the greatest scaling. Limiting warming this century to anything close to 2ºC will likely require scaling CDR to not less than 6 billion metric tonnes per year by 2050. Carbon removal will not reach the magnitude needed at mid-century without an all-of-the-above approach.

"This century" potentially represents some backsliding on the 2015 Paris Agreement, inasmuch as a prudent interpretation of "hold warming" would imply no overshoot; global action was slow following 2015 and overshoot assumptions are now fairly standard. It's really the unguarded expression "anything close to 2ºC" that is quite alarming: presumably it signifies an expectation of warming of above 2ºC, and frames this as in itself challenging to achieve. This is hardly aligned with IPCC science and Paris Agreement: while Microsoft may be ahead of its competitors on transparency, and may be investing relatively large amounts of money in carbon removals and climate technologies, to describe itself as leading on net zero generally is unambiguously greenwashing. The Paris Agreement is not to limit global warming to below 2.0 degrees, and to 1.5 degrees if possible—it is to limit it to *well* below 2.0 degrees, and to 1.5 degrees if possible. 'Well below' 2.0 degrees is sometimes informally interpreted as 1.7 degrees.

The cloud giants' own sustainability strategies offer important clues to the assumptions and values embedded in the sustainability solutions which they provide to their clients. **Microsoft Cloud for Sustainability** is an overarching term for a set of sustainability-related tools and services from Microsoft and its partners. These include Microsoft Purview Compliance Manager, Microsoft Sustainability Manager, the Emissions Impact Dashboard, and the Microsoft

Environment Credits Service. Microsoft's eight principles of sustainable software development are aimed at developers and engineers building, designing, and deploying applications.

In 2023 Olivia Byrne interviewed Andrew Quinn, a Global Sustainability Leader at Microsoft, and Andrea Coluccio, Partner Cloud Solution Architect at Microsoft. Quinn highlighted work being done to mitigate the climate impact of their cloud services. According to Quinn, Microsoft servers are "optimised to use the least amount of power" and they seek to reduce embodied emissions by recycling hardware. The recycling of data centre hardware fits in with a broader circular economy approach, although the infrastructure for doing so on a large scale is currently lacking. In 2020, Microsoft increased the rate they recycled and reused cloud hardware to 82%.

The interviewees also emphasised how Microsoft employees are also "incentivised to find things to do better" in terms of sustainability. One key initiative has been Microsoft's **internal carbon tax,** also known as a **carbon fee**, introduced in 2012 so all of their "business groups know that they have a limit to the volume of emissions their work can produce per year," and they must stick to this from a financial point of view. For example, the carbon tax is designed to incentivise software developers to write carbon efficient code, to gain the most value from each gram of carbon the product of that code emits as possible. Microsoft also gathers data remotely via telemetry on Windows products, enabling them to see where further energy efficiency improvements could be made. Microsoft offers tools and guidance for other companies to adopt a similar model. This internal carbon tax directs funds to Microsoft's internal carbon removal and reduction efforts (Willmott 2022).

Microsoft continues to hire more individuals in **dedicated sustainability roles**. Microsoft has employees in roles such as a "cloud solution architect" to advise customers on how to implement strategic solutions to reduce costs and increase the sustainability of their cloud usage.[47] Cloud solution architects consult with their clients to educate them about the sustainability tools in Microsoft's cloud offering. However, Quinn suggested that Microsoft cannot "force" consumers to implement these tools as it "comes down to them doing it."

Microsoft also has the Azure Sustainability guidance, a set of best practices which can be discussed with customers, based on best technical and up-skilling practices shared across industry to help to respond to partner and customer demands for sustainability (Taylor, 2022).

As Coluccio describes, Microsoft advises customers on "an internally developed impact dashboard which highlights the CO2 emissions generated from the workloads they are running" to design a tailored sustainability solution. As we have explored elsewhere in this report, there are controversies around the carbon savings associated with cloud migration. Critics highlight cloud migration as contributing to greater energy consumption and further GHG emissions in practice, when sustainability tools and efficiency improvements are not correctly implemented.

---

[47] Garber, D., Malik, J. and Fazio, A., 2013. *Windows Azure Hybrid Cloud*. John Wiley & Sons, Indianapolis.

Microsoft also makes efforts to wider stakeholder engagement. As Quinn described, when building data centres, Microsoft may make efforts to involve local communities to assess that they "can balance their consumption requirements [of the data centre] to make sure they are net zero." But there remains a wider problem with energy grids in different countries as each nation has a different percentage of renewable power in their national power grid networks annually, and in some areas of the world there is not enough renewable energy available to meet demand.

Microsoft still have diesel generators and batteries for back-up power to ensure that there are no power cuts so that customers can access the cloud and its applications at any time of the day.[48] Going forward, Microsoft are moving away from diesel generators to "biodiesel and hydrogen fuel cells to remove the fossil fuel dependency" (Quinn), reinforced in the literature as a more sustainable, renewable alternative to diesel.[49] It remains to be seen if these pledges are met. Cloud computing data centres currently running at maximum efficiency but still being powered by fossil fuels can only mitigate climate impact to a limited extent. Quinn pointed to Microsoft's "commitment that all the carbon emitted into the atmosphere from the day it was founded in 1975 will be removed from the atmosphere by 2050" to ensure that its historical impacts are also removed. As described above, much could hinge on exactly how carbon impacts are calculated and reported. Furthermore, many well-established data centres are located in old, energy-inefficient buildings (Monserrate, 2022), so significant challenges remain to reform these. Monserrate's research also highlighted that data centres on a smaller scale often lack the capital and resources to invest in making their data centres carbon neutral.

Microsoft has also provided bespoke cloud-based services to oil and gas giants, including Shell and Chevron, to prop up the competitiveness of oil and gas exploration and extraction (Hao 2024). Holly Alpine, former Head of Microsoft Datacenter Community Environmental Sustainability and Employee Engagement, and organiser of a 10,000 strong worker-led sustainability group called the Sustainability Connected Community, resigned "in no small part" due to these connections.

> "This work to maximize oil production with our technology is negating all of our good work, extending the age of fossil fuels, and enabling untold emissions," Alpine wrote in the email. "We are both deeply saddened to be so let down by a company we loved so much."
>
> (Stone 2024)

Since resigning, Alpine has continued to highlight Microsoft's "enabled emissions" (see e.g. Alpine 2024).

In recent years, Microsoft has also helped to drive a generative AI "race" with worrying implications for the carbon intensity of everyday computing.

---

[48] Rittinghouse, J.W., Ransome, J.F., 2017. *Cloud Computing: Implementation, Management, and Security*, 1st ed. CRC Press, Boca Raton.

[49] Cf. e.g. Kassem, Y., Çamur, H. and Bennur, K.E., 2018. Adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN) for predicting the kinematic viscosity and density of biodiesel-petroleum diesel blends. *American Journal of Computer Science and Technology*, *1*(1), pp.8-18.

Microsoft is a major investor in OpenAI, the company behind ChatGPT, DALL-E, Sora, among others. Microsoft-led thought leadership and academic research has likely underestimated the current carbon footprint of global AI (see The Current Carbon Footprint of AI: A Microsoft Case Study). A proposal has been made to shift to a more scenario-based approach to environmentally responsible AI governance: closer collaboration between the climate modelling community and the AI community is welcome, although these collaborations are also likely to be complicated, and we should be careful we're not kicking problems into the long grass.

Microsoft owns roughly 49% of OpenAI's equity, having invested about US$13 billion. Microsoft also provides computing resources to OpenAI through its Azure cloud. OpenAI's GPT-4 performed better than many comparable models on the 2023 Foundation Model Transparency Index (48%), although the creators of the index note that "[a]ll develops have significant room for improvement." Additionally, GPT-4's relatively good transparency score is based on domains such as model access, capabilities, risks, mitigations, distribution, and usage policy, and not on domains more directly applicable to environmental sustainability such as data, labor, compute, and impact. The authors of the Foundation Model Transparency Index paper comment:

> Meta and Stability AI document some aspects of compute, energy, and hardware usage, as well as the carbon footprint of model development, whereas many developers do not. Given the significant compute expenditure required to build many foundation models, the practice of documenting energy use and environmental impact is well-established along with associated tooling to measure these quantities.

Rich Gibbons of Syngena comments:

> “It is unlikely emissions will reduce [for Microsoft], as increased usage of products such as Copilot, Azure OpenAI and ChatGPT will continue to produce more emissions [...] And should use continue to grow, that may well kick-start a new round of datacentre building, too. Perhaps the only real way for organisations such as Microsoft and Google to reduce their emissions will be for the majority of customers to reject these new GenAI services until they are absolutely critical.” (Cited in *Computer Weekly* [Donnelly, 2024]).

Meanwhile Sam Altman, the CEO of OpenAI, is the main investor in the nuclear fusion energy start-up Helion. According to proponents like Altman, fusion is the likely solution to growing energy demands, and associated carbon pollution: offering the potential for a virtually limitless supply of clean energy, producing minimal greenhouse gas emissions, and significantly less long-lived radioactive waste compared to current nuclear fission power. Microsoft announced a purchasing agreement in 2023, by which Helion will supply nuclear fusion generated electricity by 2028. This is an ambitious timeline, given the early developmental stage of the technology, and many uncertainties associated with nuclear fusion.

Appendix 1: Actions and Resources

The scope of this report does not include formal recommendations. However, below we have gathered some further resources which may help readers continue to explore these issues, and come up with their own impactful actions. We also hope that many **ideas for potential actions** are nonetheless implied by the contents of this report.

Just for example, **companies** might explore whether GreenOps and/or grid-aware approaches may be suitable, to help them put their digital operations in a planetary context. They can adopt carbon-aware approaches in a constructively critical way, and contribute to pushing them forward. They can regularly engage their cloud service providers, and other IT suppliers, on sustainability issues. This includes ensuring that sustainability is central to all procurement processes. Companies can even ask their suppliers of non-IT goods and services about their own approach to sustainable cloud. Different sectors can contribute in different ways. For example, companies that create software can try to ensure that it runs on older hardware, not just the latest hardware. More broadly, companies can also actively engage with **climate scientists**, **other academics**, **environmental NGOs**, and **environmental grassroots movements**, to deepen their knowledge of the issues. They can also seek to be well ahead of climate-related legislation. This may mean adopting carbon accounting strategies for emissions across all scopes, and internal incentives and controls to accelerate decarbonisation. Companies can invest in low-carbon electricity and new energy technologies to help decarbonise the grid. They can run 'Thriftathons,' by analogy with Hackathons, to explore ways of tackling overprovisioning or to find innovative sustainability improvements. They can prioritise energy efficiency when evaluating AI models. When considering procuring or developing AI systems, they can also explore alternative analytics and approaches. Where AI is used, unnecessary AI model retraining and execution can be eliminated to conserve energy, and techniques such as routing queries to appropriately sized models can be used. Companies can work to improve the conversation around AI by refusing to spread unsubstantiated hype, and emphasising distinctions between different kinds of AI. All companies can actively engage with the cloud giants, as well as Nvidia, and other big tech companies. Together we can ask these big players for more transparency, more credible and rapid pathways to net zero, and greater emphasis on equality and justice. We can insist that any pledges for future action be mapped to IPCC timescales for decarbonisation. None of these big players is a monolith, so this may mean working with the most progressive elements within them to give those people the support, evidence, and tools they need to drive change. **Shareholders** of these companies can also exert pressure through voting, engagement, and divestment.

Likewise, **policymakers** can remove barriers to make it easier to deploy AI technologies that benefit the environment. At the same time, they can establish clear sectoral regulations to limit GHG emissions and extend the useful life of hardware. This means things like creating robust rights-of-repair and incentives to support regenerative design, not planned obsolescence. Building on established instruments such as Energy Star, TCO Certified, and SPEC Power, standards and certifications can be developed or iterated for hardware. Such standard-setting and benchmarking can be accompanied by procurement standards and fiscal policies to incentivise data centres to become more energy-efficient.

Placement programmes, curriculum redevelopment, and other interventions can incentivise data centre workers and AI professionals to develop green skills and to transition into climate-aligned roles. Public funding for research and development can assist in the innovation and improvement of ICT equipment, heat exchange and removal technologies, energy storage technology, and even software. Moratoriums may be a useful instrument: pausing the deployment of high-risk AI practices until their impacts are better understood. Moratoriums can be useful to buy time for proposals that are promising but not aligned with the IPCC's timescales. Moratoriums can also be useful tools for breaking deadlocks around perfecting a particular policy, scheme, or intervention: stakeholders can agree in advance to pause until consensus is reached.

Policymakers can take bold action at the level of infrastructure expansion, e.g. planning approvals for data centre construction or expansion, to ensure consistency with climate goals. That can seek to ensure that data centre companies pay adequate taxes and generate revenues for local communities. Progressive taxation, along with restrictions on cryptocurrency mining, and requirements for local employment and/or participatory governance. Hypothecated taxation is another tool in the policymakers' toolkit, which could raise funds for green initiatives. To enhance community well-being and acceptance, Impact Benefit Agreements (IBAs) might help ensure that data centres directly benefit their local regions. If new data centres are proposed, there needs to be an extensive stakeholder consultation delivered in accordance with recognised standards, an social impact report, an environmental impact report, and an IBA.

Policymakers can ensure that revisions to carbon accounting standards (such as the GHG Protocol) and proposed reforms of the clean energy market do not create new perverse incentives or opportunities for greenwashing. Ideally, standardised metrics and frameworks that prioritise energy efficiency in AI models could be rapidly developed and implemented, with planning for backward compatibility as metrics and standards iterate and improve. Similarly, policymakers can accelerate the development of interoperability standards to minimise lock-in. Funding can be allocated for credible AI for sustainability R&D, and for environmentally sustainable AI research. Interdisciplinary projects, including social sciences, the arts and humanities, and co-production with affected communities, remain as important as ever. Finally, funding can be allocated for credible AI for sustainability R&D and for environmentally sustainable AI research. Interdisciplinary projects, including social sciences, the arts and humanities, and co-production with affected communities, remain as important as ever.

We can all do things to learn more, have conversations, raise awareness, create bold proposals in our individual contexts, support and inspire one another, and take action. Discourses of delay, wishful thinking, and inequitable impacts can be challenged by anyone, wherever they arise. Some resources that may help include the following.

- The IPCC is where to get your climate science.
- For companies considering adopting a GreenOps approach, more information is available from Greenpixie, Environment Variables, Posetiv, NTT Data, techUK, The FinOps Foundation.
- SustainableIT.org is a membership organisation dedicated to developing and sharing best practice. "CIO, CTOs, IT Sustainability officers, their core teams, and business partners can benefit from SustainableIT's wealth of data insights, executive collaborative and networking events, and training and awareness programs." SustainableIT.org also steward a set of IT-specific ESG standards, which are a good place to begin for medium to large companies.
- The Green Software Foundation is "a non-profit with the mission to create a trusted ecosystem of people, standards, tooling, and best practices for building green software." There are all kinds of projects and standards, including the Real Time Cloud project, whose goal is "a standard mechanism for cloud providers to share more information, and more useful, by having the same data schema for all cloud providers, and to support updates to that data in real time, which could be minute level granularity for energy usage, and hourly or daily granularity for carbon intensity." Environment Variables is a great podcast from the Green Software Foundation for keeping up to date with news, policy, research, innovation, and generally what is going on in sustainable IT. 'Thinking About Using AI?' is a recent succinct report from the GSF on the environmental impacts of AI, which makes useful distinctions about what will be relatively easy or relatively hard for AI users to influence.
- The Digital Humanities Climate Coalition Toolkit is aimed primarily at academic researchers and Digital Humanities professionals, offering tips and tools for making your digital practices more sustainable. Some of the advice, such as green website design, may be useful more widely. Try starting with the "I Want To…" section.
- Lannelongue and Inouye (2023) offer advice on different approaches to estimating the carbon impact of computation.
- *Branch* is "an online magazine written by and for people who dream of a sustainable and just internet for all."
- Climate Fresk and Digital Collage are engaging interactive training workshops. Climate Fresk is an introduction to climate change, and Digital Collage focuses on climate change and digital technology. Digital Collage also incorporates themes like the relationship between digital technology, social media, and mental health.

- More tools! Cloud Carbon Footprint provides tooling to monitor cloud carbon emissions. Green Algorithms is a tool for estimating CO2 impact of a computational process. CO2.js is a JS library that helps developers estimate emissions related to use of their websites and apps. Greenframe.io estimates the carbon footprint of web apps for developers. Code Carbon is a Python package that estimates carbon footprint, again based on power consumption and regional carbon intensity. Scaphandre is a monitoring agent that makes it possible to see the power being used by a single process. Kepler uses eBPF to probe CPU performance. Kube Green is a Kubernetes operator to reduce the carbon footprint of your clusters. ML CO2 Impact Calculator is a calculator to estimate the carbon emissions of an ML process. Green-coding.ai estimates the carbon footprint of GPT queries, and is discussed in this 2024 episode of the Environment Variables podcast. Carbontracker estimates the carbon cost of training models.

- HotCarbon “aims to bring together researchers and practitioners in computer and networked systems to engage in a lively discussion around sustainability throughout the entire computing lifecycle, focusing on both the operational and embodied impact of computer systems.”

- Careful Industries, based in the UK, offers workshops and training for organisations who want to ensure a socially and environmentally responsible approach to AI.

- RAND Corporation’s Avoiding the Anti-Patterns of AI report covers common ways AI projects can go wrong (‘anti-patterns’).

- The MIT Risk Repository is an attempt to thoroughly map the AI risk landscape.

- Zero Carbon Analytics offer a useful greenwashing guide focused on carbon accounting, and other handy explainers.

- Greenpeace is a global environmental campaigning network, with a rich history, who did some formative work on the climate impacts of tech in the 2010s.

- Carbon Market Watch is a useful source of information on corporate carbon accounting and the voluntary carbon credit markets.

- WEALL and the Donut Economics Action Lab both offer information and resources on sustainability generally, spanning grassroots campaigning and corporate contexts.

- The Sustainable AI Innovation section of this report mentions other interesting emerging possibilities.

- Climate Acuity is a group of researchers at the University of Sussex interested in exploring closer collaborations between industry and academia around AI, the cloud, and the climate. Get in touch at climateacuity@sussex.ac.uk.

Appendix 2: Case Study: The SHL Digital Server

Nicolas Seymour-Smith, February 2024 (updated October 2024)

Our research group, SHL Digital (formerly the Sussex Humanities Lab), has for some time run its own server somewhat separate from central university IT Services. In 2023 we conducted a systemic needs analysis and formulated several options.

Context

SHL Digital is a multi-disciplinary digital humanities lab that relies on digital infrastructure to support research and collaboration. In many cases these infrastructure requirements cannot be met by running software on personal computers, e.g. because the computational processing power required is too high, or the hosted service requires a permanent online presence.

To meet these needs so far, SHL Digital has been relying on its own servers and staff to provide researchers with a platform to run their software. This platform can manage most computational tasks (short of machine-learning applications that require high amounts of GPU power), and can run any custom software that runs on the Linux operating system. In late 2023 and early 2024 we carried out a review, including estimates of carbon impacts.

Hardware and basic CO2 estimates

SHL Digital servers are three high-power computers which were bought together in 2018. These were fairly standard commercial servers for the time, and Dell provides their own estimates of the CO2 impact of these servers based on a 4 year life span:

- Two Dell Poweredge R440 servers, 2x7360 kg CO2, 2x1155 kg of which is carbon produced in manufacturing, and the majority of the rest is from estimated computational usage
- One Dell Poweredge R740xd server, 9180 kg CO2, 1321 kg of which is carbon produced in manufacturing, and the majority of the rest is from estimated computational usage

From Dell's carbon footprint reports: "Dell uses PAIA (Product Attribute to Impact Algorithm) to perform product carbon footprints. PAIA is a streamlined LCA tool developed by MITs Materials System Laboratory. It takes into consideration important attributes of the product which can be correlated to activities in order to calculate the product carbon footprint."

However none of their documentation explains what level of usage is assumed in making these estimates.

Improving usage estimates

Given that we can monitor the CPU usage of our servers and the number of visitors to our websites, we can adjust Dell's estimates by substituting an impact based on our real usage.

The DHCC Toolkit provides links to resources for calculating carbon impact of computation in a few different contexts. The most relevant for us are:

- Green Algorithms for estimating CO2 impact based on various relevant properties of the server, including the 'real usage factor' of the CPU, and
- CO2.js, which can help us calculate the emissions associated with the number of bytes transmitted from our websites to visitors across internet infrastructure that has its own CO2 impact. We do note studies suggesting that the energy usage of network infrastructure is inelastic in relation to data transfer.

Monitoring real computational usage

Plenty of tools and services exist for aggregating CPU usage over long periods. netdata.cloud provides one such free tool that is very simple to install on all platforms and provides a simple web based user interface that can be accessed either locally or through the netdata.cloud website.

Using this tool to monitor the CPU load of our servers for 2 weeks, we got average values of 7.65%, and 1% for our R440 servers, and 1% for our R750xd server. Entering this and other relevant details into http://calculator.green-algorithms.org/, we got 945g and 891g of CO2 per day for our R440s and 450g for our R750xd server.

It's interesting to note that while our CPU usage is low on both our R440s, it is significantly lower on our second unit, and yet the CO2 calculation is not far different. This implies that at low usage numbers at least, base power usage could be dominating the CO2 output. Further, while the R750xd has similar usage to one of our R440s, the CO2 impact came out roughly half. Given that this unit has half as much memory, perhaps a lot of that base power usage is going into memory use. We could spend some time delving into the details of the calculator to learn more.

All told, the total CO2 impact based on 'real usage' of processing power is estimated at 2.3 kg/day or 834 kg/year.

Monitoring real data transmission

It's also possible to get the total data transmission values from the netdata.cloud service, but it's a little more difficult to coax out the value as a total rather than a rate, and so instead I used a separate tool called goaccess. There's a handy tutorial for setting this up for long term monitoring here.

This provided a monthly data transmission value of 4 GB over a period of a month, or roughly 5 MB per hour. The paper “Network energy use not directly proportional to data volume: The power model approach for more reliable network energy consumption calculations” (10.1111/jiec.13512) suggests that for volumes as low as this, there is negligible impact on the CO2 consumption of the internet infrastructure that supports this data transmission.

We can now add the manufacturing impact documented in Dell’s documentation to our ’real’ estimates of computational and transmission impacts to get a potentially improved estimate:

- Manufacturing: 3632 kg
- Real CPU: 834 kg/year
- Real data transmission: negligible

Taking the same lifespan assumption as the Dell documentation (4 years): that gives us an overall impact of 7036 kg, which is a third of Dell’s 23900 kg estimate based on an unknown usage factor. However our local carbon intensity is likely different from whatever Dell is using.

Given our very low usage statistics and the disparity with the Dell estimate, we might assume that the majority of our CO2 impact comes from basic power requirements of the idle system.

Conclusion

This work was undertaken as part of a broader analysis of SHL Digital’s resources and needs, in order to plan for efficient provision of these services in the future. This analysis allowed us to weigh sustainability as a factor in those plans. From this perspective, we can ask the following questions:

- Could downsizing our server infrastructure to more closely match our real-terms usage and traffic reduce our carbon impact (by reducing manufacturing related CO2 and basic power requirements) without running us into issues of processing power? At first glance this seems likely.
- Whether downsizing could also be accompanied by migration to externally hosted and or shared services could be an interesting follow on.

Image credits

'AI Mural' by Clarote. Image credit: www.clarote.net / www.ai4media.eu / Better Images of AI

'Labour / Resources' by Clarote. Image credit: www.clarote.net / www.ai4media.eu / Better Images of AI

'Mirror D' by Comuzi. Image credit: Comuzi.xyz / Better Images of AI

'Power / Profit' by Clarote. Image credit: www.clarote.net / www.ai4media.eu / Better Images of AI

'Tree' by David Man and Tristan Ferne. Image credit: David Man and Tristan Ferne / Better Images of AI

'User / Chimera' by Clarote. Image credit: www.clarote.net / www.ai4media.eu / Better Images of AI

Blade inspection, Colorado. Image credit: Dennis Schroeder / GPA Photo Archive

Burning e-waste, Agbogbloshie. Image credit: Fairphone

Data labelling. Image credit: Nacho Kamenov & Humans in the Loop / Better Images of AI

Flooding in Sirajganj. Image credit: Moniruzzaman Sazal / Climate Visuals Countdown

GenAI giants. Image credit: Midjourney

Photovoltaic power station. Image credit: andrewwatsonuk.com. Image cropped.

Planting in Montana. Image credit: USDA Forest Service

Planting in the Kubuqi desert. Image credit: Ian Teh

Planting on Mount Brown. Image credit: Glacier National Park

Prototype artificial glacier in Ladakh. Image credit: Ankit Tanwar

Scientist studying coral reefs in Virgin Islands National Park. Image credit: NPS

Solar powered light on Lake Victoria. Image credit: Anthony Ochieng / Climate Visuals Countdown

Treating coati after wildfire. Image credit: Maria Magdalena Arrellaga / Climate Visuals Countdown

Find out more about these images

"Have you noticed that news stories and marketing material about Artificial Intelligence are typically illustrated with clichéd and misleading images? Humanoid robots, glowing brains, outstretched robot hands, blue backgrounds, and the Terminator. These stereotypes are not just overworked, they can be surprisingly unhelpful." (Better Images of AI).

"Natural England commissioned Climate Outreach to speak with conservation organisations, community groups, online influencers and nature enthusiasts to explore how we can diversify the images of people and nature, resulting in a practical, evidence-based report" (Climate Visuals).

References

References are provided as hyperlinks within the digital edition of this report.

https://doi.org/10.5281/zenodo.13850067

https://doi.org/10.5281/zenodo.13850067

Sussex Digital Humanities Lab