TEXTE
68/2025

Carbon leakage in AI-driven data center growth?

An assessment of drivers and barriers to the localization of data center operations and investments with respect to carbon pricing policies

by:

Nicolas Schmid, Martin Eichler, Moritz Reisser, Jürg Füssler
INFRAS and Digital Climate Innovation, Zurich

Vlad C. Coroamă, Oana Dumbravă
Roegen Centre for Sustainability, Zurich

Lynn H. Kaack
Hertie School Berlin

Janna Axenbeck
German Environment Agency

with input from:
George Kamiya

publisher:
German Environmnent Agency

TEXTE 68/2025

Project No. 192630

FB001747

Carbon leakage in AI-driven data center growth?

An assessment of drivers and barriers to the localization
of data center operations and investments with respect
to carbon pricing policies

by

Nicolas Schmid, Martin Eichler, Moritz Reisser, Jürg Füssler
INFRAS and Digital Climate Innovation, Zurich

Vlad C. Coroamă, Oana Dumbravă
Roegen Centre for Sustainability, Zurich

Lynn H. Kaack
Hertie School Berlin

Janna Axenbeck
German Environment Agency

with input from:

George Kamiya

Please cite as:

Schmid, N., Coroamă, V. C., Dumbravă, O., Eichler, M., Reisser, M., Kaack, L. H., Axenbeck, J., Füssler, J. (2025). *Carbon leakage in AI-driven data center growth? An assessment of drivers and barriers to the localization of data center operations and investments with respect to carbon pricing policies*. INFRAS and Roegen Centre for Sustainability on behalf of the Umweltbundesamt (UBA), Dessau-Roßlau, Germany, doi: https://doi.org/10.60810/openumwelt-7756.

On behalf of the German Environment Agency

Imprint

Publisher
Umweltbundesamt
Wörlitzer Platz 1
06844 Dessau-Roßlau
Tel: +49 340-2103-0
Fax: +49 340-2103-2285
buergerservice@uba.de
Internet: www.umweltbundesamt.de

Report performed by:
INFRAS Research and Consulting
Binzstrasse 23
8045 Zurich
Switzerland

Roegen Centre for Sustainability
Rossbergstrasse 28
8002 Zurich
Switzerland

Report completed in:
December 2024

Edited by:
Z 2.3 Digitale Transformation und Beratungsstelle Green IT
Janna Axenbeck

DOI:
https://doi.org/10.60810/openumwelt-7756

ISSN 1862-4804

Dessau-Roßlau, June 2025

The responsibility for the content of this publication lies with the authors.

Abstract

The AI-driven data center growth raises concerns about its climate impacts, including embodied emissions in digital trade and the potential for carbon leakage. The latter occurs when businesses shift operations in the short term or investments in the long term to jurisdictions with less stringent climate policies to avoid higher production costs driven by carbon pricing. In principle, a large share of AI-driven data center operations and investments are geographically flexible and can thus contribute to carbon leakage in different temporal dimensions.

Short-term, operational carbon leakage can result from shifting workloads to data centers in regions with cheaper and emission-intensive electricity, while long-term carbon leakage may emerge through strategic localization of new data center investments in response to anticipated electricity cost differences due to carbon pricing.

This study offers a comprehensive first assessment of the global carbon leakage potential associated with AI-driven data center operation and investment. The focus is on carbon leakage from costs imposed by emission trading systems (ETS) on data center electricity consumption. The study estimates AI's current and near-future electricity consumption and evaluates the technological feasibility of shifting compute loads. Additionally, it maps global compute capacity against carbon intensities and the presence of ETS.

The findings reveal that current AI compute capacity and expected growth is predominantly concentrated in the United States and China, jurisdictions with less stringent climate policies than the EU. This disparity between compute capacity concentration and regional climate policy ambition prompts consideration of whether carbon pricing through ETS induces carbon leakage.

Economic theory suggests that the extent of carbon leakage is moderate. This is primarily due to the current growth stage of the AI industry and the ability of AI data centers and companies to pass-through additional costs. Furthermore, compliance with the climate goals announced by AI companies, as well as regulatory requirements in the area of data protection, are likely to reduce carbon leakage. Despite the limited current likelihood of significant carbon leakage from ETS, potential conditions for increased carbon leakage in AI data centers can be identified, including slow renewable energy uptake, which may perpetuate or increase grid mix differences, and the potential for future overcapacity in data centers running on emission-intensive electricity, which would increase their ability to receive shifted workloads.

Additionally, data center growth may strain decarbonization efforts in other sectors, such as transport and industry, by competing for renewable electricity. Further, an increase in cross-border data flows may amplify the embodied emissions in data trade, even if such trade flows are not intentionally avoiding carbon pricing.

This report provides a first assessment of the potential and extent of carbon leakage from AI-driven data center operation and investment growth. Based on a mixed-methods approach, the report draws a research agenda for further analyses and offers policy recommendations to mitigate potential carbon leakage from ETS and embodied in digital trade. Addressing the broader challenge of energy and resource efficiency, meeting data center electricity demand growth with renewables, and tackling embodied emissions in global data flows, is critical for climate-aligned AI development.

Zusammenfassung

Das KI-bedingte Wachstum von Rechenzentren wirft Bedenken hinsichtlich Klimafolgen auf, darunter graue Emissionen im digitalen Handel und die Gefahr von Carbon Leakage. Letzteres tritt auf, wenn Unternehmen aufgrund höherer Produktionskosten durch CO_2-Bepreisung ihre Tätigkeiten oder Investitionen in Länder mit weniger strikten Klimapolitiken verlagern.

Grundsätzlich sind ein Großteil des Betriebs und der Investitionen in Rechenzentren geographisch flexibel, was zu Carbon Leakage beitragen kann. Kurzfristig kann operatives Carbon Leakage entstehen, wenn Rechenlasten in Regionen mit günstigerem, aber emissionsintensiven Strom verlagert werden. Langfristig kann Carbon Leakage durch strategische Standortwahl für Investitionen in Rechenzentren entstehen, basierend auf erwarteten Stromkostenunterschieden aufgrund der CO_2-Bepreisung.

Diese Studie bietet eine erste Einschätzung globaler Anreize für Carbon Leakage von KI-Rechenzentren. Der Fokus liegt auf erhöhten Stromkosten von Rechenzentren durch Emissionshandelssysteme (ETS). Die Studie schätzt den aktuellen und zukünftigen Stromverbrauch von KI, bewertet die technische Machbarkeit der Verlagerung von Rechenlasten und untersucht die Plausibilität der Verlagerung von Rechenlasten. Zusätzlich kartiert sie die globale Rechenkapazität im Verhältnis zur Kohlenstoffintensität von Strom und der Präsenz von ETS.

Die Ergebnisse zeigen, dass aktuelle KI-Rechenkapazitäten und das erwartete Wachstum vorwiegend in den USA und China konzentriert sind – zwei Rechtsräume mit weniger strengen Klimapolitiken als die der EU. Diese Diskrepanz zwischen der Konzentration von Rechenkapazitäten und regionaler Klimapolitik wirft die Frage auf, ob eine CO_2-Bepreisung durch das ETS Carbon Leakage verursacht.

Wirtschaftstheoretische Argumente legen jedoch nahe, dass das Ausmaß von Carbon Leakage moderat ist. Dies liegt hauptsächlich an der gegenwärtigen Wachstumsphase der KI-Branche und der Fähigkeit von KI-Rechenzentren und Unternehmen, zusätzliche Kosten weiterzugeben. Darüber hinaus dürften die Einhaltung der von KI-Unternehmen angekündigten Klimaziele sowie regulatorische Anforderungen im Bereich des Datenschutzes Carbon Leakage verringern. Trotz der aktuell begrenzten Wahrscheinlichkeit für Carbon Leakage werden potentielle Bedingungen für eine Zunahme von Carbon Leakage identifiziert, darunter eine langsame Verbreitung erneuerbarer Energien, welche die Unterschiede im Strommix aufrechterhalten oder verstärken könnte und die Möglichkeit künftiger Überkapazitäten in emissionsintensiven Rechenzentren, was ihre Fähigkeit erhöhen würde, verlagerte Rechenlasten zu empfangen.

Darüber hinaus könnte das Wachstum von Rechenzentren die Dekarbonisierung anderer Sektoren wie Verkehr und Industrie verlangsamen, indem Rechenzentren um erneuerbaren Strom konkurrieren. Weiter könnten zunehmende grenzüberschreitende Datenflüsse importierte und exportierte graue Emissionen erhöhen, selbst wenn solche Datenflüsse nicht absichtlich CO_2-Bepreisung umgehen.

Dieser Bericht bietet eine erste Bewertung des Potentials und des Ausmaßes von Carbon Leakage in Rechenzentren. Auf Basis eines Mixed-Methods-Ansatzes formuliert der Bericht eine Forschungsagenda für weitere Analysen und gibt politische Empfehlungen zur Minderung potentieller Carbon Leakage-Risiken und grauer Emissionen in grenzüberscheitenden Datenflüssen. Für eine klimafreundliche Entwicklung von KI ist es entscheidend, den wachsenden Strombedarf von Rechenzentren mit erneuerbaren Energien zu decken, sowie die grauen Emissionen in grenzüberschreitenden Datenflüssen zu adressieren.

Table of content

List of figures

List of tables

1 Introduction

The rapid growth of artificial intelligence (AI) applications and systems has spurred an unprecedented demand for computing capacity and expansion of data centers (DCs) globally. The International Energy Agency (IEA) predicts that the electricity consumption of data centers will rise sharply in all major regions of the world (IEA, 2024a). A significant share of this increase is attributed to the electricity consumption of AI-related computing: By 2026, it could be at least ten times higher than in 2022/2023. This rising electricity demand poses a challenge to achieving climate mitigation goals (Kaack *et al.*, 2022; Bieser *et al.*, 2023; ITU and World Bank, 2024). One of these challenges is the potential of carbon leakage which occurs if, for reasons of costs related to carbon pricing policies, business operations or investments intentionally transfer to jurisdiction with less ambitious climate policies.

Data center compute loads – AI training and inference loads included – are flexible and can in principle be shifted geographically. This means that data center operators and developers can react to costs imposed by climate policy such as carbon pricing, which contribute to price differences in electricity. Data center operators or developers may choose to evade regions with high electricity prices for those with lower prices. If the lower-priced electricity entails carbon emissions and is in a less stringent climate jurisdiction, carbon leakage takes place. The shifting of compute loads and possible consequent carbon leakage can be a short-term occurrence in which AI loads are sent over the Internet to another data center. In the long term, the localization of new AI data center capacity, i.e. investment decisions, may also be influenced by current and expected carbon prices and their effects on electricity, leading to long-term carbon leakage. Although the literature on climate impacts of AI is expanding fast, carbon leakage has so far received little attention. In this context, the current study presents a first assessment of the global potential for carbon leakage due to AI. The study specifically explores the possible impact of carbon pricing instruments, in particular emissions trading systems (ETS) and with a special focus on the European Union, on AI-triggered carbon leakage in data center operations and investments.

To achieve these goals, the study estimates the current and near-future global electricity consumption of AI. To do so, Section 4, first establishes the global energy footprint of AI in data centers. Assessing in a second step the technological flexibility of compute loads, it sets an upper bound for the global electricity consumption of AI loads that can be shifted. Subsequently, it estimates the current distribution of AI compute capacity around the world. Combined with the carbon intensities of national grid mixes, and building on the energy of shiftable compute loads already assessed, it then determines short-term upper bounds of emissions that can be outsourced and absorbed due to AI in the context of carbon leakage. Forth, in Section 5, the study compares the geographic distribution of compute capacity to the distribution of ETS across jurisdictions. In Section 6, the study then explores arguments based on economic theory on the extent to which the shift of compute loads is dependent on electricity price variations from carbon pricing instruments such as ETS, and also highlights other influencing factors such as corporate climate strategies, data security and localization policy, and general locational factors, such as price differences (in particular with respect to the overall energy price). Through this analysis, the report determines both how plausible it is that the short-term carbon leakage upper bound might realistically take place, but also – and arguably more important – the possible long-term carbon leakage from shifts in investments into data centers resulting from expected electricity price increases due to ETS. Finally, Section 7 summarizes the key findings, provides a research agenda and policy recommendations.

Based on a mixed-methods, this report provides a first assessment of AI-related carbon leakage risks from carbon pricing instruments. The report highlights areas for future research to deepen our understanding of this evolving challenge. It provides policy recommendations to address potential carbon leakage and the more general challenge to address electricity demand increases from data center growth and embodied emissions in cross-border data flows.

2 Problem definition and methodology

According to the definition by the European Commission, carbon leakage occurs because of differences in climate policies between jurisdictions, which make companies shift operation to regions with less stringent climate policies:

Carbon leakage definition by the European Commission

"Carbon leakage refers to the situation that may occur if, for reasons of costs related to climate policies, businesses were to transfer production to other countries with laxer emission constraints. This could lead to an increase in their total emissions. The risk of carbon leakage may be higher in certain energy-intensive industries." (European Commission, 2021)

Data centers, where the vast majority of AI computing takes place, require substantial amounts of electricity and could thus be at heightened risk of carbon leakage as carbon pricing may increase their operational costs from electricity. The aim of this study is thus to *explore the indirect carbon leakage in AI-driven data center operation and growth that might occur due to carbon pricing instruments. The focus of this study is on costs imposed by emission trading schemes*. *Costs related to other climate policies are not covered due to the limited scope of this study.*

Such carbon leakage within AI-driven data center operation and growth can be at least of two fundamentally distinct types.[1] We distinguish between:

- *Short-term or operational carbon leakage*, which entails the temporary spatial shift of specific computing loads to a geography with lower electricity prices and higher emission intensities, and
- *Long-term or investment carbon leakage*, which stems from changed localization decisions of investments in new data centers due to expected future electricity prices.

To assess the potential and likelihood of carbon leakage from data center growth, the report is based on a mixed-methods approach and a variety of data sources. We combine the collection and analysis of quantitative data from data center operations and investments with a review of existing literature and selected expert interviews. A more detailed description of the methodology can be found in the Appendix C and in the individual chapters.

Main data sources include company reports by data center operators and users, academic and grey literature on (AI-driven) data center growth, energy consumption, as well as role of ETS in electricity prices and carbon leakage. Table 8 in Appendix A lists the interviewees representing industry, research, and international organizations. The experts' opinions are reflected in the report were relevant, but no personal attribution is made (unless explicitly allowed). They all provided valuable feedback in two main rounds of review as well as along the entire project.

[1] It is also possible that final consumers may shift their non-AI demand to AI services originating from regions without stringent climate policies, as these offerings might become comparatively more attractive due to the absence of carbon pricing. However, exploring this phenomenon is beyond the scope of this study.

3 Estimated AI energy consumption and feasibility of location-independent AI computing

This section lies the foundation for AI's environmental footprint, and its relevance in the context of carbon leakage, which will be discussed in later sections.

3.1 AI model life cycle and relevance for carbon leakage

The life cycle of an AI model, and how it relates to environmental life cycle assessment, is schematically represented in **Fehler! Verweisquelle konnte nicht gefunden werden.**. As in any economic activity, energy consumption and environmental impacts happen during the production of devices, their use, and end-of-life (EoL). While hardware production and EoL can of course also have varying environmental impacts depending on the deployed industrial processes, their displacement requires the shift of physical goods and even entire production processes. Such displacement has larger inertia and coarser granularity than that of computing loads, which can be easily shifted over the Internet to a different geography, and therefore entail a greater risk of carbon leakage from data center operations and investments. In this study, we thus focus on AIs use phase, i.e. the electricity consumption of devices employed for AI algorithms. For the use phase, the AI model lifecycle has been categorized in several ways. One known categorization distinguishes three main phases (Kaack *et al.*, 2022): i) model development, ii) model training, and iii) model deployment and inference:

a) **Model development** focuses on the conceptualization and design of the model. It includes problem definition and scoping, hyperparameter tuning and architecture search, model selection (i.e., choosing an appropriate algorithm and architecture, including trial-and-error on different model options), and possibly feature engineering.

b) **Model training** includes model training, fine-tuning, evaluation, and possible retraining (which can also happen later, during deployment).

c) **Model deployment and inference** marks the usage of the model. Inference refers to using the model to make predictions on new, unseen data. In the deployment phase, the model may be integrated in a production environment, such as a web application or mobile app.

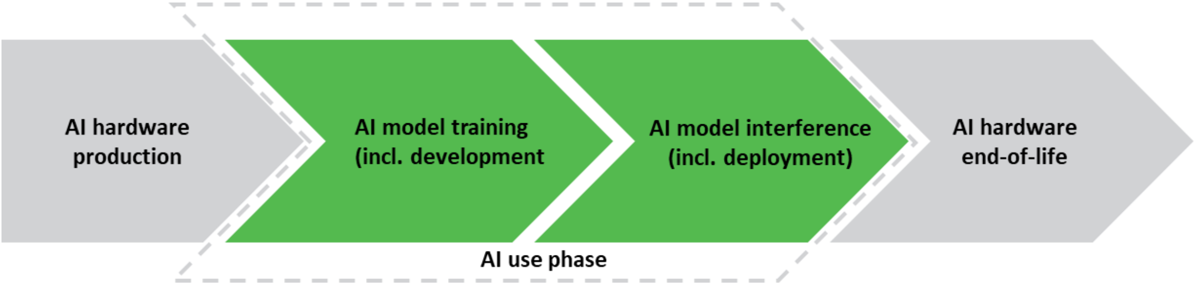
In reality, the picture is more complex and cyclic, involving several sub-steps and auxiliary activities such as performance monitoring and maintenance, and retraining or updating the model if needed. Other sources include data acquisition for training as part of model development and data preparation (cleaning, transforming, and partitioning it into training, validation, and test sets) as part of training. These distinctions are not so relevant for our simplified model.

The development phase consists of many training runs and can therefore be summarized as training, which has been done in other scholarly work that does not distinguish between development and training (Luccioni, Jernite and Strubell, 2024). Various industry sources also do not distinguish a development phase, seeing model selection as part of training, and only distinguishing a "data collection and/or preparation" phase ahead of the training.[2] Hence, this study distinguishes two main phases in terms of energy consumption and flexibility: "model training" (which includes development) and "model inference" (which includes deployment), as shown in Figure 1.

[2] See, for example, Yurushkin, M. How do Machine Learning Pipelines Work? https://broutonlab.com/blog/how-machine-learning-pipelines-work/ (Accessed: 2 December 2024) and Datatron Blog (HRSG). What is a Machine Learning Pipeline? https://datatron.com/what-is-a-machine-learning-pipeline/ (Accessed: 2 December 2024).

Figure 1: AI model life cycle and its embedding into the environmental life cycle assessment.

The environmental life cycle assessment includes production of devices, their usage for AI, and finally their end-of-life (EoL). The AI model is fully included in the use phase of the devices, and consists of training and inference.


AI hardware production
AI model training (incl. development
AI model interference (incl. deployment)
AI hardware end-of-life
AI use phase

Source: own illustration, Roegen Centre for Sustainability and INFRAS.

3.2 The energy footprint of AI

As a widely deployable technology, AI is possibly the most important general-purpose technology of our times, and perhaps even on par with technological revolutions such as the steam engine or electricity (Brynjolfsson and McAfee, 2017). As such, AI is impacting all walks of life. And while this development encompasses the potential to help addressing environmental (Kaack *et al.*, 2022; Rolnick *et al.*, 2022) or societal (De-Arteaga *et al.*, 2018; Hager *et al.*, 2019) issues, the potential environmental impact of such a ubiquitous technology has rightfully become a source of concern.

These two types of AI impact are often referred to as *direct impacts*, which are directly computing-related, and *indirect impacts*, which relate to the application of the AI approach, and can be immediate or longer-term and systemic (Kaack *et al.*, 2022). This distinction reflects the discussion on direct and indirect environmental impacts of information and communication technologies (ICT) more generally (Coroamă *et al.*, 2020; Bremer *et al.*, 2023; Axenbeck, Berner and Kneib, 2024).

This study focuses on the *direct energy and carbon impact of AI's use-phase*. Data on its electricity consumption and related carbon impact, however, are quite sparse and inconsistent. Additionally, existing assessments are fragmented across individual phases of a model's lifecycle (as presented below) and various levels of abstraction and related functional units: Some assessments quantify the impact of one single model, others of individual model inferences, other the global yearly impact of AI.

Training energy versus inference energy

Early studies on the environmental footprint of AI focused on the energy and carbon impacts associated with *training* the AI models (Lacoste *et al.*, 2019; Strubell, Ganesh and McCallum, 2019, 2020; Schwartz *et al.*, 2020). A recent review (Verdecchia, Sallou and Cruz, 2023) also shows that the training phase has been the focus of research on the energy and GHG impact of AI. And this focus stands to reason when large deep neural networks are intensively trained to then be deployed relatively seldomly, as is the case for a Go-playing ML model such as AlphaGo or an ML model deployed in medicine for the pre-screening of possible cancers.

With the advent of *generative AI*, the deployment phase has come more in the focus of research. While large language models (LLMs), such as ChatGPT and Google's Gemini, or image generators, such as Midjourney or DALL-E, require large amounts of energy to be trained, they are also widely used, so the *inference* phase additionally requires substantial amounts of energy; the exact amount of energy and the ratio between training and inference for different application areas are still a matter of debate:

- Early industry data from Meta and Google indicate that around 2021-2022, the training phase already accounted for 20-40% ML-related energy consumption in their DCs, while inference accounted for 60-80% (Patterson *et al.*, 2022; Wu *et al.*, 2022). These data stem from before ChatGPTs public release and widespread consumer use of such models.
- In 2023, it has been estimated that the inference of ChatGPT requires as much as 564 MWh of energy per day (De Vries, 2023); this would mean that a little more than two days of inference already outweigh GPT's estimated training costs of 1,287 MWh (Luccioni, Viguier and Ligozat, 2024).
- The amount of inference instances required to outweigh the costs of training is the metric used by Luccioni et al. (2024). Unsurprisingly, their analysis shows that both training and inference energy grow with the size of the model. However, at least for the BLOOMz family of models analyzed, the training energy grew faster with the number of model parameters than the inference energy did. Hence, for the 1.7 billion (B) parameters BLOOMz-1B model, the threshold of energy parity was at 290 million inferences, while for the 7B parameters BLOOMz-7B model, it was almost double at 592 million inferences. Extrapolating these results to ChatGPT and its reported 10 million daily users at the time of paper writing, this would amount to a threshold of a couple of weeks / few months – albeit under the quite conservative assumption of just one query per user (Luccioni, Jernite and Strubell, 2024).
- One of the interviewees for this study presented a (yet unpublished) very detailed systems dynamics model they have developed, which shows that in all its global scenarios, generative AI inference energy is closing the gap to training energy, but has not surpassed it yet. And even after it will have surpassed it, it will not become all-dominant, but training will still be a sizeable portion of the overall AI energy.

While this last result is seemingly contradictory to the results above, it might in fact not be the case. In a different context, that of the energy consumption and energy intensity of the Internet (measured in TWh/year and in kWh/GB, respectively), both bottom-up and top-down assessments have been shown to yield systematic, and opposing errors (Coroamă, 2021): Since they draw system boundaries too narrow, not accounting for example for redundancy or legacy equipment, bottom-up assessments consistently yield underestimates. Similarly, top-down assessments often draw system boundaries too wide, yielding overstatements (Coroamă, 2021).

Given the complex process of AI model training and that model development might entail various trial-and-error experimentation and validation steps – even as many as "thousands of training runs" (Kaack *et al.*, 2022) – as well as possibly frequent retraining, bottom-up models such as those in Luccioni, Jernite and Strubell (2024) and Luccioni, Viguier and Ligozat (2024) might indeed draw too narrow system boundaries and consistently yield understatements of the energy consumption required for training.

Given all these considerations and the ultimately unknown ratio between training and inference energy, this study deploys assumptions (rooted in the studies and considerations listed above) for the energy ratio of training, $er_T(y)$, as well as the complementary energy ratio of inference,

$$er_I(y) = 1 - er_t(y) \tag{1}$$

For the present, we assume that

$$er_T(2024) = er_I(2024) = 0.5 \tag{2}$$

while for the future it is assumed that inference will indeed start dominating, with

$$er_T(2028) = 0.3, er_I(2028) = 0.7 \qquad (3)$$

Current global and predicted future global energy footprint of AI

In the context of exploring the (necessarily global) potential for carbon leakage, relevant is not the footprint of single models or inference instances, but the global yearly energy and carbon footprint of AI. Data on this level, however, are relatively sparse. Although numerous estimates of global data center energy usage exist, there are only few estimates that single out the footprint of AI within the entirety of DCs; all of them being quite recent and uncertain.

Most of these studies conclude that until very recently (and despite all the hype surrounding AI and its energy consumption), the energy impact of AI was – and still is – extremely limited from a global perspective.[3] At the same time, considering the impressive growth rates of AI deployment, GPU and TPU chip production to support the seemingly "endless hunger" for AI computation, they all predict a substantial growth over the next few years, albeit to different extents. Existing assessments of the global energy consumption of AI (both recent developments and projects until 2030) are the following:

- A study by Schneider Electric (2023) sees AI average power consumption growing more moderately from 4.5 GW in 2023 to 14.0 – 18.6 GW in 2028; taking an average value of 16.3 GW for 2028, these correspond to 39.4 and 142.8 TWh for 2023 and 2028, respectively.
- Based on estimated Nvidia's GPU sales, and assuming a 100% utilization rate (De Vries, 2023) suggests that these GPUs could consume in 2023 "up to" 5.7 – 8.9 TWh/year of electricity, while – due to planned expansions – the consumption of newly produced devices is expected to grow to 85.4 – 134 TWh yearly by 2027. The 100% utilization rate is certainly unrealistic, but then this Nvidia assessment leaves aside further (less important) GPU producers as well as Google's AI consumption based on its self-produced TPUs. We thus consider the average values of these ranges (i.e., 7.8 and 109.7 TWh, respectively) as reasonable approximations of the *additive AI energy consumption* for 2023 and 2027, respectively. Assuming a starting value of 8 TWh/year for 2022 from (Goldman Sachs, 2024) and linear interpolation between 2023 and 2027, yields the values from Table 1.
- Also based on Nvidia sales, an IEA report (2024a) puts forward a similar 7.3 TWh for 2023, stating that by 2026, AI "is expected to have grown exponentially to consume at least ten times its demand in 2023" – i.e., 73 TWh.
- Based on a proprietary analysis, Morgan Stanley (2024) expects AI energy to grow from 13 TWh in 2023 and 46 TWh in 2024 to 224 TWh by 2027.
- Goldman Sachs (2024), one of only two estimates to start earlier than 2022, devises negligible 5-8 TWh yearly for 2020 – 2022, still quite low 12 and 30 TWh for 2023 and 2024, respectively, followed by more substantially growing consumption, reaching 209 TWh by 2030.
- Finally, Semianalysis (2024) estimated the yearly average annual power used by AI in DCs between 2020 and 2028, from very modest 318 MW (corresponding to less than 3 TWh) in 2020 to a massive 56.3 GW (corresponding to 493 TWh) in 2028.

[3] Locally, however, due to its high-power density, AI can already cause energy shortages.

The detailed numerical results (after transformation, where needed) are shown in Table 2 and then graphically represented in Figure 2. As both show, the sources agree that until 2022, AI's energy consumption was negligible: A few TWh among the estimated 240 – 340 TWh global yearly DC energy consumption in 2022 (IEA, 2023), which itself corresponded to only about 1-1.5% of global electricity consumption. By 2024, however, the consumption is estimated to have grown to already 30 – 74 TWh. Future projections vary, reaching from substantial but not so worrisome 73 TWh in 2026 (IEA, 2024a) or 209 TWh by 2030 (Goldman Sachs, 2024) to more alarming 400 – 500 TWh already by 2027-2028 (De Vries, 2023; Semianalysis, 2024).

Table 1: Numerical values (in TWh/year) for AI's yearly global energy consumption.

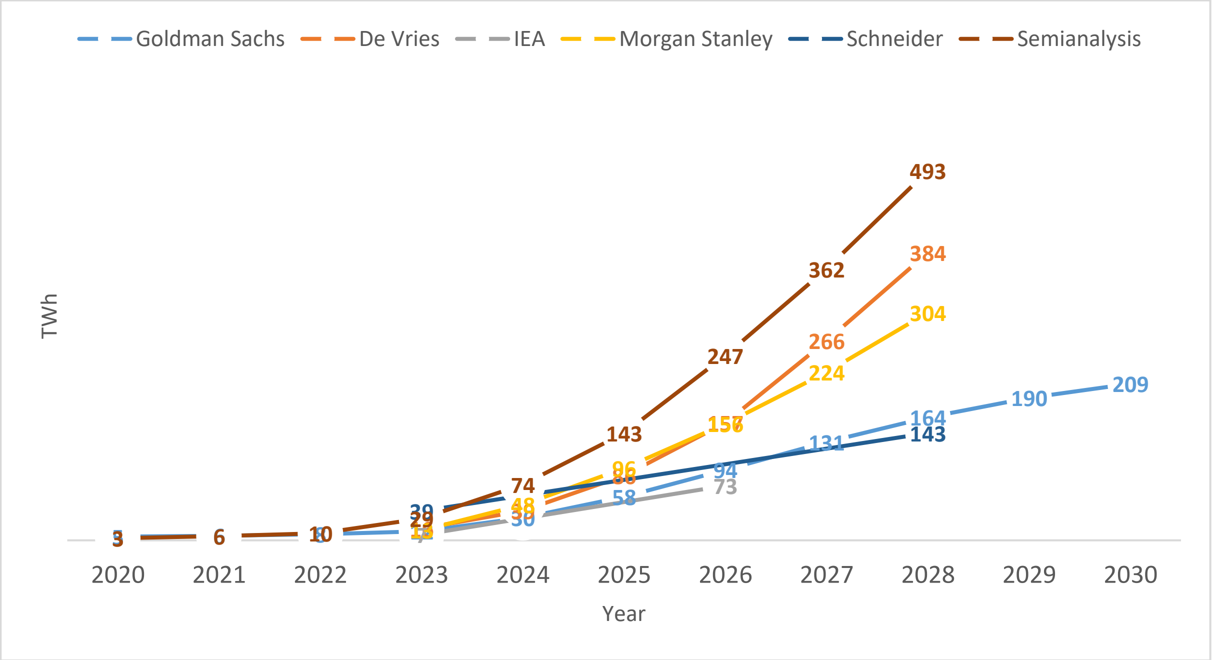
The years 2020 – 2030 are analyzed.

Study	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Schneider Electric (2023)				39					143		
De Vries (2023)				16	39	86	157	266	384		
IEA (2024a)				7			73				
Morgan Stanley (2024)				13	48	96	156	224	304		
Goldman Sachs (2024)	5	6	8	12	30	58	94	131	164	190	209
Semianalysis (2024)	3	6	10	29	74	143	247	362	493		

Source: own analysis based on the sources cited, Roegen Centre for Sustainability and INFRAS. The two own extrapolations to 2028 marked blue. Data preparation performed where necessary, as described in the text. Valuable input was provided by George Kamiya.

Figure 2: Recent developments and projections for yearly global AI energy consumption.

Studies: Schneider Electric (2023), De Vries (2023), IEA (2024a); Morgan Stanley (2024); Goldman Sachs (2024), and Semianalysis (2024).



Source: own illustration based on the data from Table 1, Roegen Centre for Sustainability and INFRAS.

The numbers for today are plausible. The energy consumption of Google and Microsoft, most probably the largest users of GPUs and TPUs for AI, grew jointly from 17.5 TWh in 2018 to 48 TWh in 2023 (Google, 2024; Microsoft, 2024a). This growth is likely to a substantial extent (but not exclusively) attributable to AI; on the other hand, numerous other (albeit smaller) users exist. It thus stands to reason to use the median value of 43 TWh as likely number for AIs total energy consumption in 2024. For future projections, we extrapolate the tendencies in De Vries, (2023) and Morgan Stanley (2024) by one year to 2028 (yielding 384 and 304 TWh, respectively) – highlighted in blue in Table 1 – and work with the median value of 304 TWh for 2028 (which is also very close to the average of 298 TWh), and which represents about 1% of the expected global electricity consumption in 2028. Overall, we thus have:

$$EC_{AI}(2024) = 43\,TWh;\; EC_{AI}(2028) = 304\,TWh \tag{4}$$

3.3 Flexibility of AI computing loads

This section addresses the technical extent to which AI computing loads are geographically shiftable, possibly leading to operational carbon leakage as a consequence. This discussion only sets a theoretical upper bound for what might be shifted in practice. In reality, there are numerous further constraints (such as availability of computing power in other geographic locations, legal requirements, etc.) that influence the appetence for load shifting. They will be discussed in Section 6.

For the flexibility analysis, ML *training* and *inference* need to be discussed separately.

Flexibility of training

In principle, training has a high geographic flexibility: It is usually not time-critical and is a batch process which can be performed offline. Additionally, due to the sheer size of the models and the time it takes to train them in one location, training on geographically distributed clusters also becomes more and more of a necessity.

BLOOM, an open-access LLM which reveals such data, for example, has 176 billion parameters and its training took approximately 3.5 months on a supercomputer, corresponding to about 1 million compute hours.[4] Other models are even larger: Although the exact size is not disclosed by the popular GPT-4 and Gemini Advanced LLMs, the former has substantially more parameters than the known 175 billion parameters of its predecessor GPT-3,[5] and both GPT-4 and Gemini have been estimated to more than 1 trillion parameters each, perhaps as many as 1.8 trillion.[6]

Consequently, distributed training becomes a necessity for LLMs. While distributed training (or "learning", as it is often called in the computing literature, as seen from the model's perspective) is not new (Sergeev and Balso, 2018; Ben-Nun and Hoefler, 2019); it is, however, increasingly deployed for also across geographically distributed computing clusters (Duan *et al.*, 2024). Distributed training can take advantage of various mechanisms such as

- *data parallelism*, in which the training data is split across different GPU clusters, each cluster trains a copy of the model on its data, while the model parameters are periodically synchronized across clusters,[7]

[4] See Bekman, S. (14 July 2022). The Technology Behind BLOOM Training. https://huggingface.co/blog/bloom-megatron-deepspeed (Accessed: 2 December 2024).

[5] See Kudesia A. (5 March 2024). Gemini vs. GPT-4: Which one is better? https://fireflies.ai/blog/gemini-vs-gpt-4 (Accessed: 2 December 2024).

[6] See Schreiner M. (11 July 2023). GPT-4 architecture, datasets, costs and more leaked. https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/ (Accessed: 2 December 2024).

[7] See Huggingface (HRSG). Model Parallelism. https://huggingface.co/docs/transformers/v4.13.0/en/parallelism (Accessed: 2 December 2024).

- *model parallelism*, in which the model itself is split across clusters and different parts of the model are trained on different clusters,[8], and
- *pipeline parallelism*, where the model is split into sequential stages, each stage being processed by a different cluster.[9]

Various parallel developments of synchronization and communication techniques (such as asynchronous updates or gradient compression and Nvidia's Collective Communications Library, respectively), additionally smoothen the distributed training process (Duan et al., 2024). Given the (coordinated) segmentation of the training process that underlies it, distributed training consequently also facilitates the shift of training loads across geographically distributed clusters.

There are, however, also opposing tendencies, which make geographically distributed training more challenging: Network latency, for example, can be an issue; transferring the large datasets needed for training can become so time-consuming, energy-intensive and expensive, that it might be required to train the model where the data is located. The data might also not be transferrable due to legal or privacy-related reasons: If training data is sensitive or subject to strict regulations (e.g., medical data), training might need to occur close to the data source.

Finally, too much training granularity might paradoxically hurt the potential for geographically distributed training. *Federated learning* is a type of machine learning which deploys an extreme type of data parallelism: The data typically does not leave its origin which might be the computers or even devices such as smartphones of numerous users. A first model is sent to each such device, which uses its own data to improve and/or personalize it. The devices send only the improvements they made to the model, but not their personal data used to achieve them, to a central server, which combines them to obtain a better model (Zhang *et al.*, 2021).

Federated learning can make ML model training not only more efficient and personalized, but also comply to higher privacy requirements, as also highlighted by the European Data Protection Supervisor,[10] and is thus an expanding model. And while it has been traditionally reserved for smaller-scale models, it is now increasingly proposed in the training of LLMs as well, both by the academic community (Ye *et al.*, 2024) and industry.[11] As mentioned above, although federated learning advances distribution and granularity, it also ties this distribution to very precise locations (i.e., those where the training data is), reducing the flexibility of geographic distribution – and thus flexibility – to a minimum.

Flexibility of inference

As opposed to training, ML inference must often display low or very low latency (Gao *et al.*, 2018). For autonomous vehicles, for example, which continuously evaluate sensor data, the ability to make split-second decisions is crucial to road safety;[12] even small delays potentially leading to the loss of human lives. As the model may also not be subject to any communication disturbances, it needs to be on the vehicle itself and not in a data center, placing this example outside the scope of this study.

[8] See AWS (HRSG). Introduction to Model Parallelism. https://docs.aws.amazon.com/sagemaker/latest/dg/model-parallel-intro.html (Accessed: 2 December 2024).

[9] See PyTorch (HRSG). Pipeline Parallelism. https://pytorch.org/docs/stable/distributed.pipelining.html (Accessed: 2 December 2024).

[10] See Lareo X. Federated Learning. https://www.edps.europa.eu/press-publications/publications/techsonar/federated-learning_en (Accessed: 2 December 2024).

[11] See Roth H., Xu Z., and Renduchintala, A. (10 July 2023). Adapting LLMs to Downstream Tasks Using Federated Learning on Distributed Datasets. https://developer.nvidia.com/blog/adapting-llms-to-downstream-tasks-using-federated-learning-on-distributed-datasets/ (Accessed: 2 December 2024).

[12] See ultralytics (HRSG). Inferenz in Echtzeit. https://www.ultralytics.com/glossary/real-time-inference (Accessed: 2 December 2024).

Another example is healthcare, where AI-supported monitoring of patients can also be extremely time-sensitive; this example was brought up during one of the interviews for this study. But applications need not be life-threatening to require low latency; it suffices that they should be (near) real-time. For all its drawbacks, generative AI can also be a productivity booster for skilled users.[13] Users are thus likely to expect a (near) real-time experience and not tolerate large delays, which would probably induce a competitive disadvantage. AI-based answering services and virtual receptionists are also quickly expanding; this is another of the many fields requiring fast responses and little delay.

Except for a few extremely time-sensitive applications such as self-driving cars and some healthcare applications, network delays should not be hindering the real-timeliness of a majority of AI applications. For a natural dialogue across the Internet, for example, a round-trip time of no more than 300 milliseconds is required; in an early example of videoconference-based distributed conference, even a 27,000 km connection (more than half of Earth's circumference) passing over 20 Internet nodes on its way, could easily fit within this limit (Coroamă, Hilty and Birtel, 2012). Replicas of models are nevertheless often placed in cloud or edge data centers close to their users; not because the distance in itself were a problem, but the possible network congestion over long distances. It is the same reason why thought-after contents (such as the newest episode of a popular series or the most recent software update of a widely used software) are replicated numerous times across continents – either at DCs owned by the content provider or at externally contracted content distribution networks (CDNs).[14]

Overall flexibility of ML compute loads

The literature does not provide any quantitative hints towards the geographic flexibility of ML training or inference. Based on the considerations above, however, this study can deploy reasonable assumptions. Training generally has a higher geographic flexibility than inference. The trend, however, seems to be rather towards further restrictions on the geographic flexibility of both training and inference: training because of the infrastructure required for large data volumes, and inference due to the trends towards edge computing and real-time AI applications.

With these considerations in mind, our assumptions for the shares of training and inference that could theoretically be shifted, are:

$$flex_T(2024) = 0.6;\ flex_I(2024) = 0.4 \tag{5}$$

$$flex_T(2028) = 0.5;\ flex_I(2028) = 0.35 \tag{6}$$

3.4 Short-term shiftable AI power consumption

In Section 6, we will conclude that most data centers are likely to be built independently of carbon pricing over the next years. Hence, short-term leakage, i.e., the geographic shift of computing activities rather than entire data centers, may be more relevant for carbon leakage. In the following, we therefore focus on the potential of such short-term leakage. Bringing all the insights from this section together, the yearly operational flexibility potential of AI can be computed as the flexible share of both training and inference taken together:

$$FlexEn_{AI}(y) = \left(er_T(y) * flex_T(y) + er_I(y) * flex_I(y)\right) * EC_{AI}(y) \tag{7}$$

Using the results and assumptions from Equations 2 – 6 in Equation 7 yields

[13] See Roberts, S. Study Finds Applying Generative AI Correctly Can Improve Productivity by 40%. https://verbit.ai/enterprise/how-smart-application-of-generative-ai-can-improve-productivity/ (Accessed: 2 December 2024).

[14] See Hempenius, K. Content Delivery Networks (CDNs). https://web.dev/articles/content-delivery-networks (Accessed: 2 December 2024).

$$FlexEn_{AI}(2024) = 21.5\,TWh;\ FlexEn_{AI}(2028) = 138.3\,TWh \tag{8}$$

This estimate represents the energy upper bound of the loads that may be shifted; an energy upper bound at the source, as it were. The loads that can be shifted do not only depend on the flexibility at the source, however, but also on the absorption potential at the target. Concretely, it is limited by the available capacity of AI accelerators such as GPUs and TPUs. The available capacity is the inverse of the utilization rate:

$$avCap_{AI} = 1 - ut_{AI} \tag{9}$$

In terms of energy, this sets the additional constraint of the energy of the maximum yearly absorption capacity:

$$AbsorbEn_{AI}(y) = (1 - ut_{AI}(y)) * EC_{AI}(y) \tag{10}$$

Taking these two constraints together, the energy of AI shiftable loads can be computed as the minimum between flexibility and absorption:

$$ShiftEn_{AI}(y) = Min\big(FlexEn_{AI}(y), AbsorbEn_{AI}(y)\big) \tag{11}$$

Using Equations 7 and 10, Equation 11 can be more precisely rewritten as:

$$ShiftEn_{AI}(y) = EC_{AI}(y) * Min\Big((1 - ut_{AI}(y)), \big(er_T(y) * flex_T(y) + er_I(y) * flex_I(y)\big)\Big) \tag{12}$$

The value resulting from Equation 12 sets an upper bound for the energy that could be feasibly shifted within a year because it is both technically flexible and can be absorbed elsewhere. In practice, there will be numerous other constraints (e.g., economic, geopolitical, organizational, etc.) further limiting this value; Section 6 discusses them in detail. As for the utilization rate of AI accelerators, several of the experts interviewed for this study agreed that there is currently very little availability, all accelerators being intensely used, almost at full capacity. Everyone agreed that current utilization rates are 80% or above. This statement applies even for the Leibniz supercomputing center. Its director, Prof. Kranzlmüller, mentioned that the GPUs in the Leibniz supercomputer are differently optimized than those of commercial hyperscalers. As the supercomputer is often needed really large scientific simulations, the GPUs that are progressively freed from other tasks are not immediately occupied, but kept for the upcoming simulation. But although maximizing utilization rate is not the main criterion for the supercomputer, he still estimated a current GPU utilization rate of 80-85%. Given both the demand pressure for the exceptionally thought-after AI accelerators, and the prevailing opinion among interview partners, this study assumes a utilization rate of 80% for AI accelerators not only for today, but also for 2028. In the longer run, it is conceivable that an AI bubble might burst, freeing AI compute capacity around the world. Currently, however, it is a seller's market, and that moment seems farther away than just a couple of years. With this assumption, Equation 10 instantiates to

$$AbsorbEn_{AI}(2024) = 0.2 * 43 = 8.6\,TWh;\ AbsorbEn_{AI}(2028) = 0.2 * 304 = 60.8\,TWh \tag{13}$$

The absorption potential being lower than the flexibility, the shiftable energy coincides with the former:

$$ShiftEn_{AI}(2024) = 8.6\,TWh;\ ShiftEn_{AI}(2028) = 60.8\,TWh \tag{14}$$

Summary of chapter 3

To sum up, AI is a transformative technology with profound societal and environmental impacts, and a growing energy and carbon footprint. By 2028, AI might consume about the same amount of energy as all data center loads combined just a few years ago. To assess its flexibility potential and related maximum operational leakage, training and inference need to be analyzed individually. Training has more flexibility than inference, but both potentials might be slowly decreasing. Analyzing operational flexibility is not enough, however; available AI compute capacity at the receiving end is also required and represents the bottleneck today for shifting emissions in the short term. Today, according to our assumptions, less than 10 TWh of electricity might thus be shiftable; for 2028, the theoretical capacity-constrained upper limit of electricity that may be shifted in AI-related data center operations could reach around 60 TWh/year. The question is whether these potentials are leveraged and for which reason.

4 Geographic distribution of AI compute capacity

The aim of this section is to estimate the current global geographical distribution of AI computing energy. This information is valuable in the assessment of both operational (i.e., short term) leakage as well as long-term carbon leakage from AI-driven data center investments, as follows:

- For operational carbon leakage (short-term): Assuming similar utilization rates across geographies implies that the unused capacity available for shifting AI compute loads is distributed proportionally to the overall AI compute capacity. This allows conclusions on the maximum possible shifts to each geography and, combined with data on the individual carbon intensities of electricity, on the maximum amount of operational carbon leakage.
- For investment carbon leakage (long-term): The current distribution of AI compute loads is a proxy for where future capacities would be built according to the same proportionality, in a business-as-usual (BAU) scenario (Appendix B motivates this assumption). This sets a helpful baseline/counterfactual when analyzing the effects of climate policy measures that might change DC placement decisions.

4.1 Poor data availability and chosen assessment method

For an accurate estimate of the geographic distribution of AI compute capacity, country-level aggregates based on metered AI electricity consumption would be the most precise data source. Even for total consumption in DCs, however, such data only exists for a couple of countries, most of them within Europe, i.e., Ireland, the Netherlands, and Finland (Kamiya and Bertoldi, 2024).[15] For AI-specific consumption, no country-level data is so far available.

The second-best option would be company-wide estimates of large AI operators around the world, ideally in conjunction with information on their geographic distribution. Unfortunately, while such data exists for general-purpose DCs, none of the major DC operators devises the energy consumption of AI separately.

With these two options not feasible, the following approach was chosen: Data on the geographic distribution of general DC energy consumption can be assessed (Appendix C.1 discusses how). And it is reasonable to assume that AI compute capacity (and thus its energy consumption) is distributed similarly to general DC compute capacity (and energy consumption); Appendix C.1 motivates this as well. A two-step method was thus deployed in this study:

- to first estimate the geographic distribution of DCs generally; i.e., to initially compute the overall yearly energy consumption of all large hyperscale and colocation DCs globally, $EC_{DCs}(year, global)$, and to subsequently allocate this consumption to individual countries, $EC_{DCs}(year, C), \forall C \in \{worldwide\ countries\}$,
- and then to distribute accordingly both the global energy consumption of AI, $EC_{AI}(year, global)$, and the energy of the shiftable AI capacity, $ShiftEn(year, global)$, as computed in Equations 4 and 14, respectively. The resulting values are denoted as $EC_{AI}(year, C)$ and $ShiftEn_{AI}(year, C),\ \forall C \in \{worldwide\ countries\}$, respectively.

[15] This is about to change with the new, EU-wide data center reporting scheme, launched in 2024. In the near future, aggregated country-level data on yearly DC energy consumption will be available throughout the EU, see European Commission. Data centres in Europe – reporting scheme. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13818-Data-centres-in-Europe-reporting-scheme_en (Accessed: 2 December 2024). However, this will not address country-wide data availability outside the EU, nor will it generate AI-specific data, not even in the EU.

As motivated above and in more detail in Appendix C.1, country-level data is only available for a few countries. To compute the geographic distribution of total DC energy consumption, the energy consumption of large DC operators from around the world was used and distributed to individual countries. Appendix C.2 presents the 22 large global hyperscale and colocation operators used in this study to assess the geographic distribution of general DC energy consumption, as well as the data sources for their total energy consumption and geographic distribution of data centers.

4.2 An estimate for country-level data center energy consumption

To assess per-company power consumption, the sustainability or environmental reports of important worldwide hyperscale and colocation operators were analyzed. To extract data on their data center energy consumption, the electricity consumption was used (if indicated); otherwise, the Scope 2 energy consumption was used as a proxy, which for hyperscale or colocation operators is to an overwhelming amount electricity for their DCs.

The geographic distribution of each company's data centers was then also retrieved, typically from a different source. For lack of more information, an equal energy consumption for all of a company's DCs was assumed, and thus the total consumption distributed evenly among them. Then, the individual consumptions across operators were aggregated per country. The sources for both total energy consumption and data center distribution are discussed in Appendix C.2.

Finally, corrections were applied for the two European countries with the largest overall DC energy consumption: Germany and France. Appendix C.3 explains these corrections in detail and argues why their necessity does not hint towards a fundamental problem of the deployed methodology.

The total energy consumption of the analyzed hyperscale and colocation DCs amounts to almost 200 TWh/year for 2023 globally, more precisely

$$EC_{DCS}(2023) = 196\,TWh \tag{15}$$

According to the literature, this represents roughly 50% of all data center energy consumption, when accounting for the smaller providers and enterprise DCs as well (Malmodin *et al.*, 2024).

Table 2 lists the top 12 countries in terms of DC energy consumption according to this analysis, while Figure 3 maps all results on a world map. As the table shows, the US has by far the highest DC energy consumption in large hyperscale and colocation DCs (82.4 TWh and 42.1% of the global consumption), followed by China with 39 TWh annually (corresponding to about 20% of the global consumption). The EU member states Germany and Ireland are third and fourth, followed by Singapore and Japan, all of them with consumptions far lower than the US and China. With a total of 31.2 TWh and 16% of the global consumption (not shown in the table), even the entire EU still comes behind the US and China.

Taken together, the US, China, and the EU account for 78% of the worldwide consumption. Non-EU European countries amount for another 1.6% (Russia), 1.5% (the UK), 0.6% (Switzerland), another 0.6% (Iceland), and 0.4% (Norway), etc. Together with EU's 16%, this yields a total of just over 20% for all of Europe.

The data corroborates well with data computed with an entirely different, top-down method (i.e., based on market data and vendor revenues) by the Synergy Research Group, which yields 51% for the US, 16% for China, and 17% for all of Europe.[16]

Table 2: The top 12 countries in terms of DC energy consumption according to this analysis.

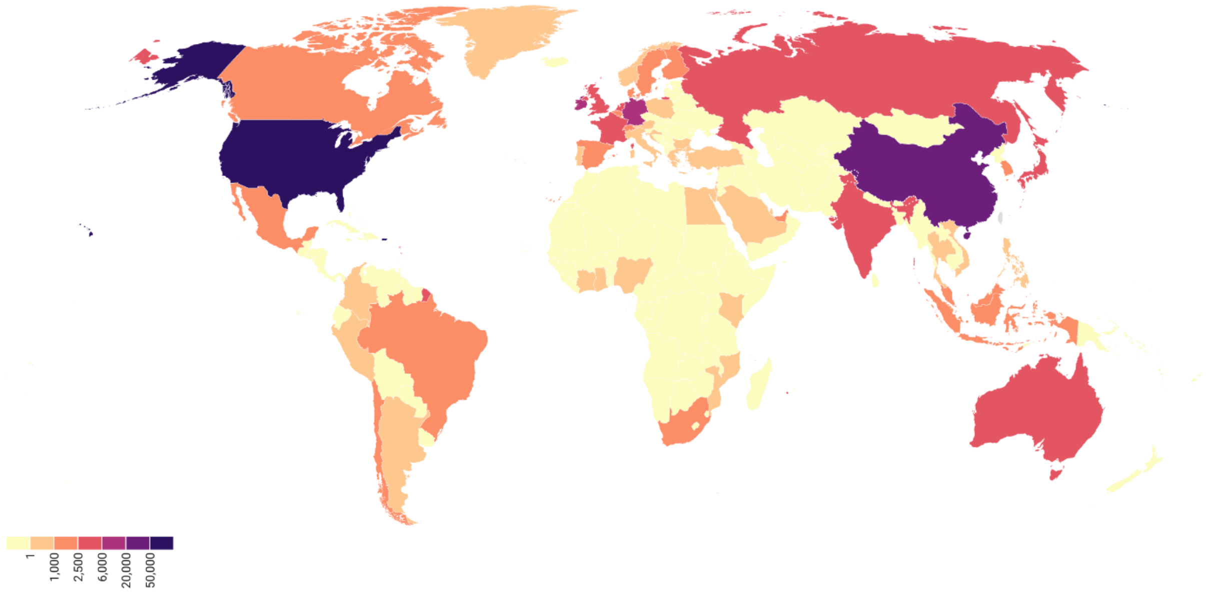
Focus is exclusively on large hyperscale and colocation DCs.

Country	Estimated 2023 DC energy (hyperscale & colocation) [GWh]	Percentage	Country	Estimated 2023 DC energy (hyperscale & colocation) [GWh]	Percentage
US	82,423	42.1%	France	4,050	2.1%
China	38,966	19.9%	Australia	3,932	2.0%
Germany	8,055	4.1%	India	3,749	1.9%
Ireland	6,394	3.3%	Netherlands	3,278	1.7%
Singapore	4,282	2.2%	Russia	3,186	1.6%
Japan	4,241	2.2%	UK	2,878	1.5%

Source: own calculations based on company reports and assumptions, Roegen Centre for Sustainability and INFRAS.

Figure 3: Mapping of 22 large hyperscale and colocation operators in the world.

The 22 considered hyperscale and colocation operators are together responsible for about 196 TWh of electricity consumption annually. With 82 TWh, the US is responsible for 42% of the total, and China for another 20% (39 TWh).



Source: own visualization of the calculations presented in Table 2, Roegen Centre for Sustainability and INFRAS (values in GWh/year).

[16] See Synergy Research Group (17 April 2024). Hyperscale Data Centers Hit the Thousand Mark; Total Capacity is Doubling Every Four Years https://www.srgresearch.com/articles/hyperscale-data-centers-hit-the-thousand-mark-total-capacity-is-doubling-every-four-years (Accessed: 2 December 2024).

4.3 Country-level AI energy consumption and GHG emissions

As stated, this study uses the current distribution of hyperscale and large colocation data centers as proxy for the distribution of AI computing capacity, both currently and in near-future. This AI distribution, in turn, is used as baseline for the further analysis of possible carbon leakage. Appendix B motivates this assumption. Still, we acknowledge that our resulting estimates are based on limited data quality. The estimated global energy consumption of AI in data centers for 2024 and 2028 was $EC_{AI}(2024) = 43\ TWh$ and $EC_{AI}(2028) = 304\ TWh$, respectively, as derived in Equation 4. Allocating it to individual countries according to the relative distribution of DC energy consumption from Table 2, yields the results shown in columns 2 and 3 of Table 3. Using the average 2024 carbon intensity of the grid mix for these countries (column 4) yields the AI-related carbon emissions per country for these years (columns 5 and 6, respectively).

In absence of future forecasts for the grid carbon intensities, 2024 values are used for 2028 as well. Given the accelerated adoption of renewable electricity around the world, this is most certainly a pessimistic assumption for most countries. Additionally, hyperscalers are large investors in low-carbon sources such as renewable and nuclear power, as discussed in Section 6. Rows in Table 4 are ordered according to carbon footprint; not the energy footprint.

Table 3: Energy consumption and GHG emissions for the top 20 emitting countries due to AI in data centers, for 2024 and 2028.

For emissions, the energy is multiplied by the carbon intensity (CI) of the grid mix; today's mix is used for 2028 as well.

Country	2024 AI energy [GWh]	2028 AI energy [GWh]	2024 grid mix [g CO_2/kWh]	2024 AI GHGs [kt CO_2eq]	2028 AI GHGs [kt CO_2eq]
US	18,083	127,842	369	6,673	47,174
China	8,549	60,437	582	4,975	35,174
Germany	1,767	12,494	381	673	4,760
India	823	5,815	713	586	4,146
Australia	863	6,098	549	474	3,348
Japan	930	6,577	485	451	3,190
Singapore	939	6,642	471	442	3,128
Ireland	1,403	9,917	291	408	2,886
Russia	699	4,942	441	308	2,180
South Africa	335	2,369	708	237	1,677
Malaysia	367	2,593	606	222	1,571
UAE	348	2,460	561	195	1,380
Netherlands	719	5,085	268	193	1,363
Indonesia	263	1,863	676	178	1,259
Mexico	346	2,447	507	175	1,241
UK	631	4,464	238	150	1,063
South Korea	265	1,872	431	114	807

Country	2024 AI energy [GWh]	2028 AI energy [GWh]	2024 grid mix [g CO_2/kWh]	2024 AI GHGs [kt CO_2eq]	2028 AI GHGs [kt CO_2eq]
Israel	179	1,267	583	105	739
Bahrain	103	729	905	93	660
Canada	548	3,872	170	93	658

Source: own calculations, Roegen Centre for Sustainability and INFRAS.

Finally, for the same top 20 countries, Table 4 sets in relation their entire AI-generated carbon emissions with the emissions of their operationally flexible loads (i.e., upper bound for emissions reductions if those loads were to be fully shifted – emissions that can be outsourced due to AI [FlexEn]) and their operational absorbing capacity (i.e., upper bound for emissions increase, if all available capacity was used [ShiftEn]). If, e.g., Germany was to introduce a very high carbon price in 2028, up to 2,166 kt CO_2eq could be shifted abroad, while if it was to become competitively cheap, an increase of 952 kt CO_2eq could happen. This corresponds to a volume of 0.32% (FlexEn) and 0.1% (ShiftEn) of Germany's total greenhouse gas emissions projected for 2028.[17] Moreover, for 2028, e.g., FlexEn is equivalent to the amount of greenhouse gases emitted by about 200,000 people in Germany in one year.[18]

Table 4: Per-country total AI-related carbon emissions, GHG emissions of the loads that are flexible, and GHG emissions of the capacity that is free to absorb incoming shifted loads, for 2024 and 2028, respectively.

Only top 20 countries in terms of AI-generated GHG emissions listed. 2024 grid mix used for 2028 as well.

Country	2024 all AI [kt CO_2eq]	2024 FlexEn GHGs [kt CO_2eq]	2024 ShiftEn GHGs [kt CO_2eq]	2028 all AI [kt CO_2eq]	2028 FlexEn GHGs [kt CO_2eq]	2028 ShiftEn GHGs [kt CO_2eq]
US	6,673	3,336	1,335	47,174	21,461	9,435
China	4,975	2,488	995	35,174	16,002	7,035
Germany	673	337	135	4,760	2,166	952
India	586	293	117	4,146	1,886	829
Australia	474	237	95	3,348	1,523	670
Japan	451	226	90	3,190	1,451	638
Singapore	442	221	88	3,128	1,423	626
Ireland	408	204	82	2,886	1,313	577
Russia	308	154	62	2,180	992	436
South Africa	237	119	47	1,677	763	335

[17] German Environment Agency (27 March 2024). Indicator: Greenhouse gas emissions. https://www.umweltbundesamt.de/en/data/environmental-indicators/indicator-greenhouse-gas-emissions (Accessed: 2 December 2024).

[18] German Environment Agency (30 January 2025). Wie hoch sind die Treibhausgasemissionen pro Person in Deutschland und wie viel wäre klimaverträglich? https://www.umweltbundesamt.de/service/uba-fragen/wie-hoch-sind-die-treibhausgasemissionen-pro-person#:~:text=Der%20deutsche%20Aussto%C3%9F%20an%20Treibhausgasen,(CO2e)%20pro%20Jahr (Accessed: 2 December 2024).

Country	2024 all AI [kt CO_2eq]	2024 FlexEn GHGs [kt CO_2eq]	2024 ShiftEn GHGs [kt CO_2eq]	2028 all AI [kt CO_2eq]	2028 FlexEn GHGs [kt CO_2eq]	2028 ShiftEn GHGs [kt CO_2eq]
Malaysia	222	111	44	1,571	715	314
UAE	195	98	39	1,380	628	276
Netherlands	193	96	39	1,363	620	273
Indonesia	178	89	36	1,259	573	252
Mexico	175	88	35	1,241	564	248
UK	150	75	30	1,063	483	213
South Korea	114	57	23	807	367	161
Israel	105	52	21	739	336	148
Bahrain	93	47	19	660	300	132
Canada	93	47	19	658	299	132

Source: own calculations, Roegen Centre for Sustainability and INFRAS.

Summary of chapter 4

Starting from the distribution of data center energy consumption on general, this section estimates the global geographic distribution of AI-related electricity consumption and the theoretical upper limit for geographically shiftable AI data center operations. The capacity-constrained upper limit describes the potential for short-term, operational carbon leakage. It considers unused compute capacity in different regions and their electricity grid carbon intensities. Key findings include that the US and China dominate data center electricity use, collectively accounting for over 60% of global consumption, with the EU trailing behind. The analysis highlights challenges in data availability but provides a foundational assessment of AI electricity use and emissions by country.

5 Comparing the data center landscape with the geographic scope of emission trading systems

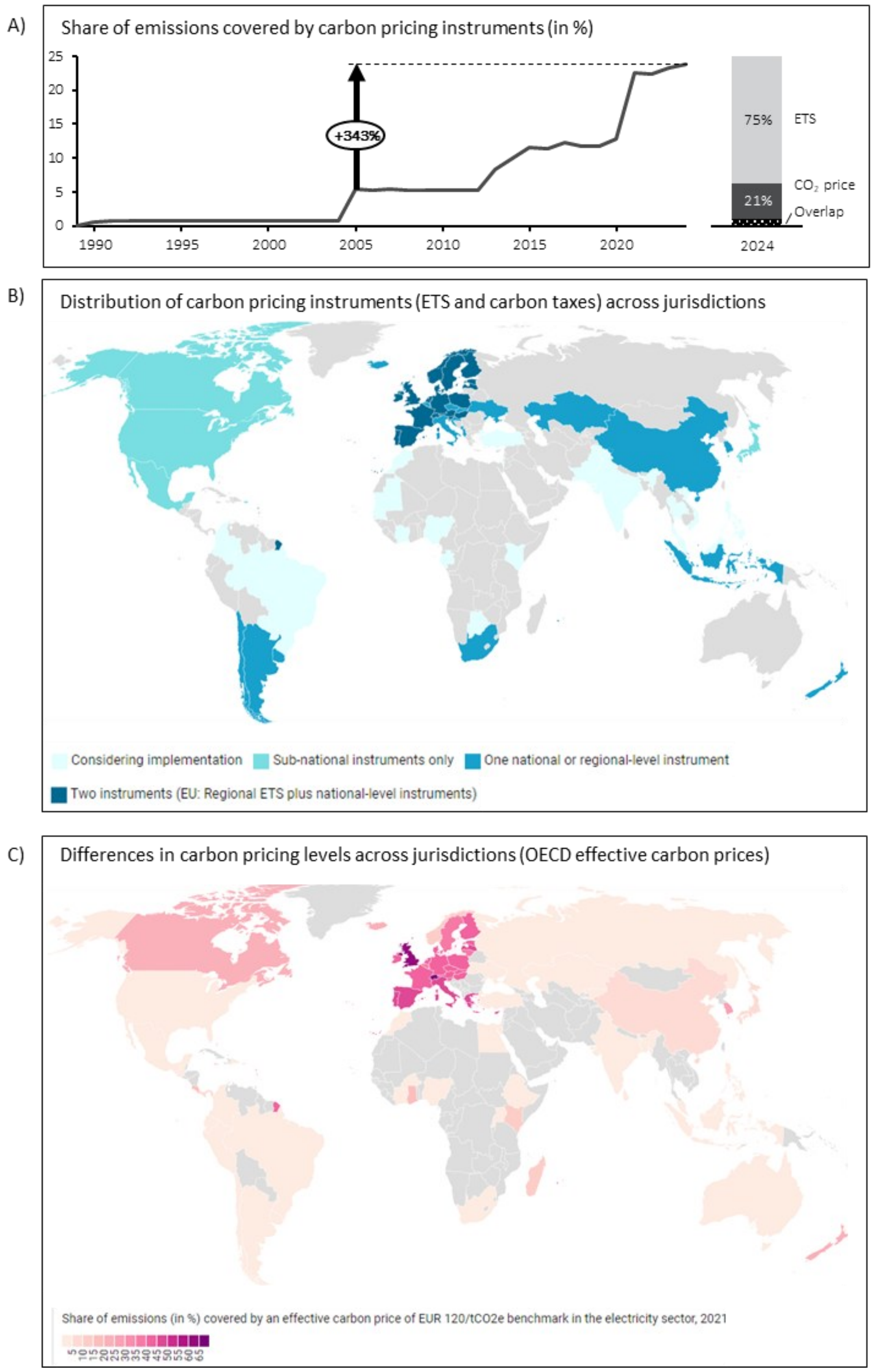
Section 4 has shown that the US and China dominate current data center operations and are likely to host most projected data center investments. Here, we compare this global "data center landscape" (see Figure 3 above) with the global scope of emissions trading systems (ETS). The emissions impact of AI-driven data center growth depends on the local electricity mix, which is influenced by whether data centers are located in countries with ambitious climate policies. Climate policies, including instruments like ETS, are designed to raise the cost of fossil-fuel-based electricity relative to renewables and nuclear by pricing carbon emissions. Ideally, these policies should discourage data centers from using carbon-intensive electricity by increasing associated operational costs. However, the effectiveness of carbon pricing policies in curbing fossil-fueled data center expansion depends on two key factors. First, the stringency of carbon policies; only a high carbon price or a strict emissions cap can significantly impact fossil electricity prices, thus influencing the operational costs of data centers. Second, disparities in carbon pricing across jurisdictions create a risk of "carbon leakage". To avoid costs, AI-driven data center operations and investments may intentionally shift from regions with substantial carbon pricing to areas with weaker or no such policies, undermining climate targets.

Comparing the global distribution of data centers shown above (Figure 3) to the distribution of ETS across countries (Figure 1B) underscores this challenge. While carbon pricing instruments cover an increasing share of global emissions (Figure 2A), still a majority of existing and anticipated data center growth occurs in jurisdictions without national-level ETS, such as the US. The US has subnational ETS (RGGI in the Northeast, Californian ETS); however, a significant share of US data center growth happens outside of these regions. Anecdotal evidence points towards partial data center expansion based on fossil gas in these regions, as recently reported for xAI, Elon Musk's AI company building a new data center in Tennessee. Furthermore, countries without any form of ETS such as India remain potential hotspots for future data center expansion. This uneven landscape of carbon pricing and data center localization highlights a tangible risk of carbon leakage and embodied emissions in digital trade, which could weaken the intended effects of stringent climate policies in specific regions such as the European Union with its ETS.

Even among those regions with ETS, differences in policy design, i.e. the level of the carbon price and projected future carbon price increases, can lead to changing investments and thus potential carbon leakage. The stringency of these instruments varies considerably. Figure 3C depicts the effective carbon price on electricity, an indicator by the OECD representing the sum of fuel excise taxes, carbon taxes, and tradeable permits that effectively put a price on carbon emissions. The figure shows the share of emissions from electricity generation that are covered by a 120€/ton CO_2e benchmark. In 2021, with a few exceptions, strict carbon pricing in the electricity sector was only present in the EU, UK, and Switzerland, with more than 30% of emissions effectively covered by a carbon price of 120€/ton CO2e benchmark.

Figure 5A and B synthesize insights on the geographic distribution of data center electricity consumption and the scope and ambition of carbon pricing. Comparing the presence or absence of carbon pricing instruments with data center electricity consumption does not yield a clear pattern (Figure 5A). However, comparing data center electricity consumption to the effective carbon price on electricity (using an OECD indicator) shows that three sets of countries can be identified: countries with high carbon prices and moderate data center localization, countries with low carbon prices and moderate (to low) data center localization, and finally the US and China with low effective carbon prices but high data center localization.

Figure 4: Distribution of carbon pricing instruments over time and across countries.



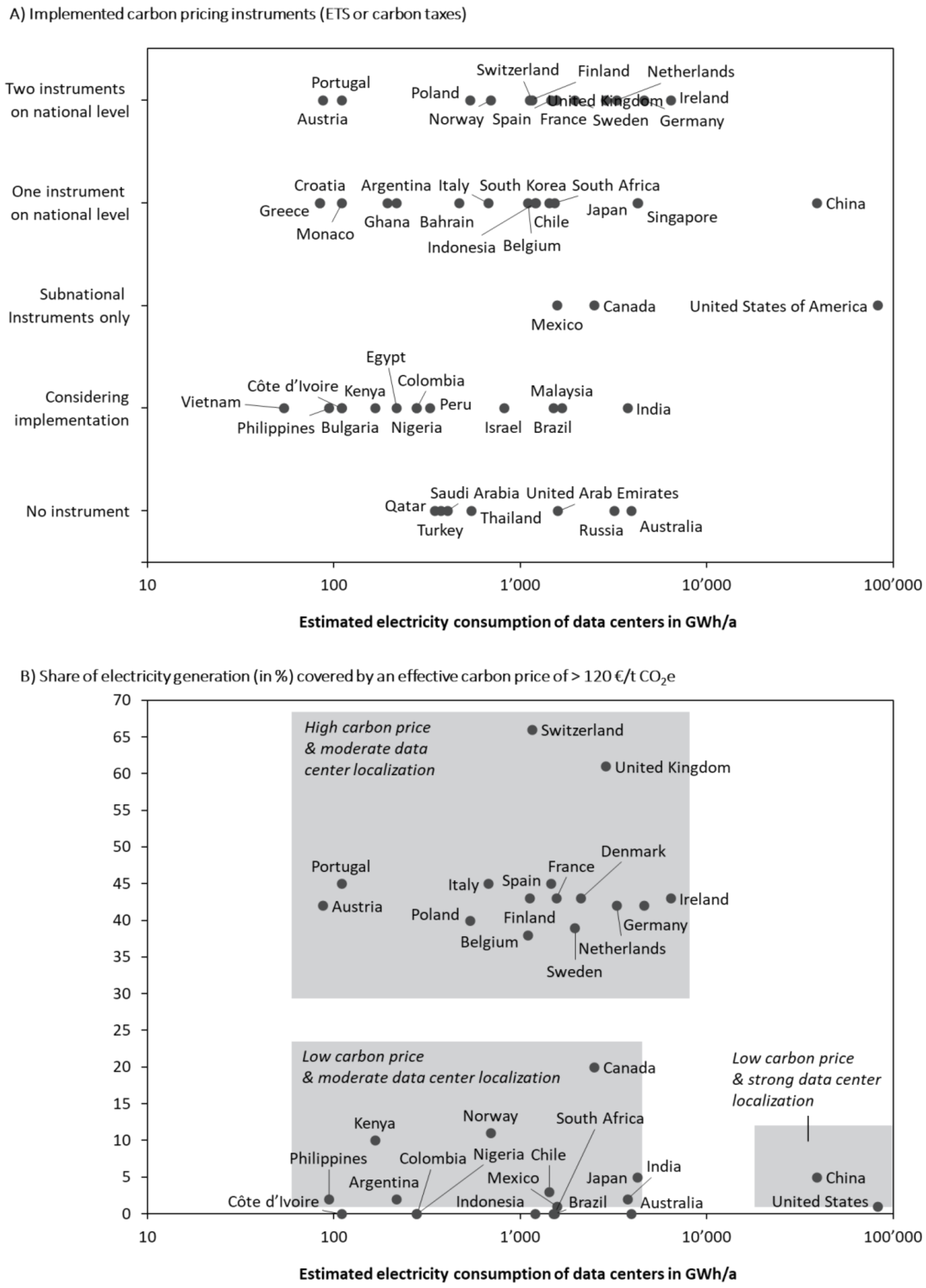
A) Share of emissions covered by carbon pricing instruments (in %)

B) Distribution of carbon pricing instruments (ETS and carbon taxes) across jurisdictions

C) Differences in carbon pricing levels across jurisdictions (OECD effective carbon prices)

Source: own illustration, INFRAS and Roegen Centre for Sustainability. Based on A) and B) World Bank Carbon Pricing Dashboard, accessed on 11.11.24. C) OECD Effective Carbon Rates Dataset, accessed on 24.10.24.

Figure 5: Comparing data center electricity consumption with the geographic scope of carbon pricing instruments (A) and stringency level of instruments (B).



A) Implemented carbon pricing instruments (ETS or carbon taxes)

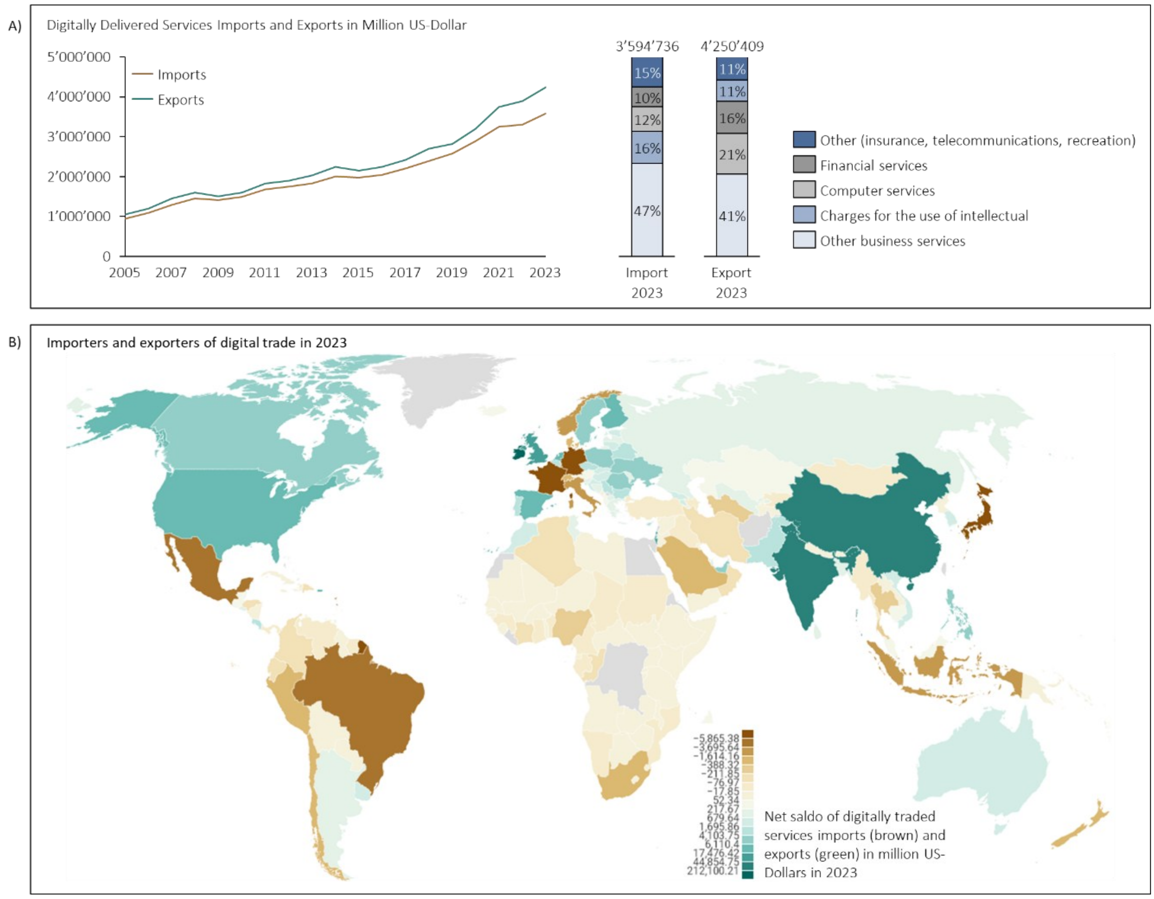
B) Share of electricity generation (in %) covered by an effective carbon price of > 120 €/t CO_2e

Source: own illustration, INFRAS and Roegen Centre for Sustainability. Sources: see Figure 4.1.

Excursion: Emissions embodied in cross-border data flows

Next to carbon leakage, the unequal geographical distribution of data centers raises the question whether the AI-driven expansion of data center operations and investments increases embodied emissions in digital trade flows (trade in services not goods) – independent of whether this is driven by climate policy or not. Here, we shortly discuss this point; the remainder of the report focuses on carbon leakage caused by costs imposed by climate policy as defined by the European Commission (see Section 2). There is no statistical classification or official statistics on cross-border data flows (OECD, 2022b). As a rough proxy for data flows, Figure 6A depicts the change over time of global imports and exports (in million US-Dollar) of digitally delivered services, as estimated by the World Trade Organization (WTO, 2023). The value of both, imports and exports of digital services, have roughly quadrupled between 2005 and 2023 (global imports and exports differ due to statistical asymmetries and discrepancies in how trade is recorded across countries). The share of the indicator “computer services” which includes data center services accounts for 21% of worldwide service exports. Using these trade flows jointly with the respective emission intensities of service sectors for the respective countries can provide a first insight into traded emissions from digital services.

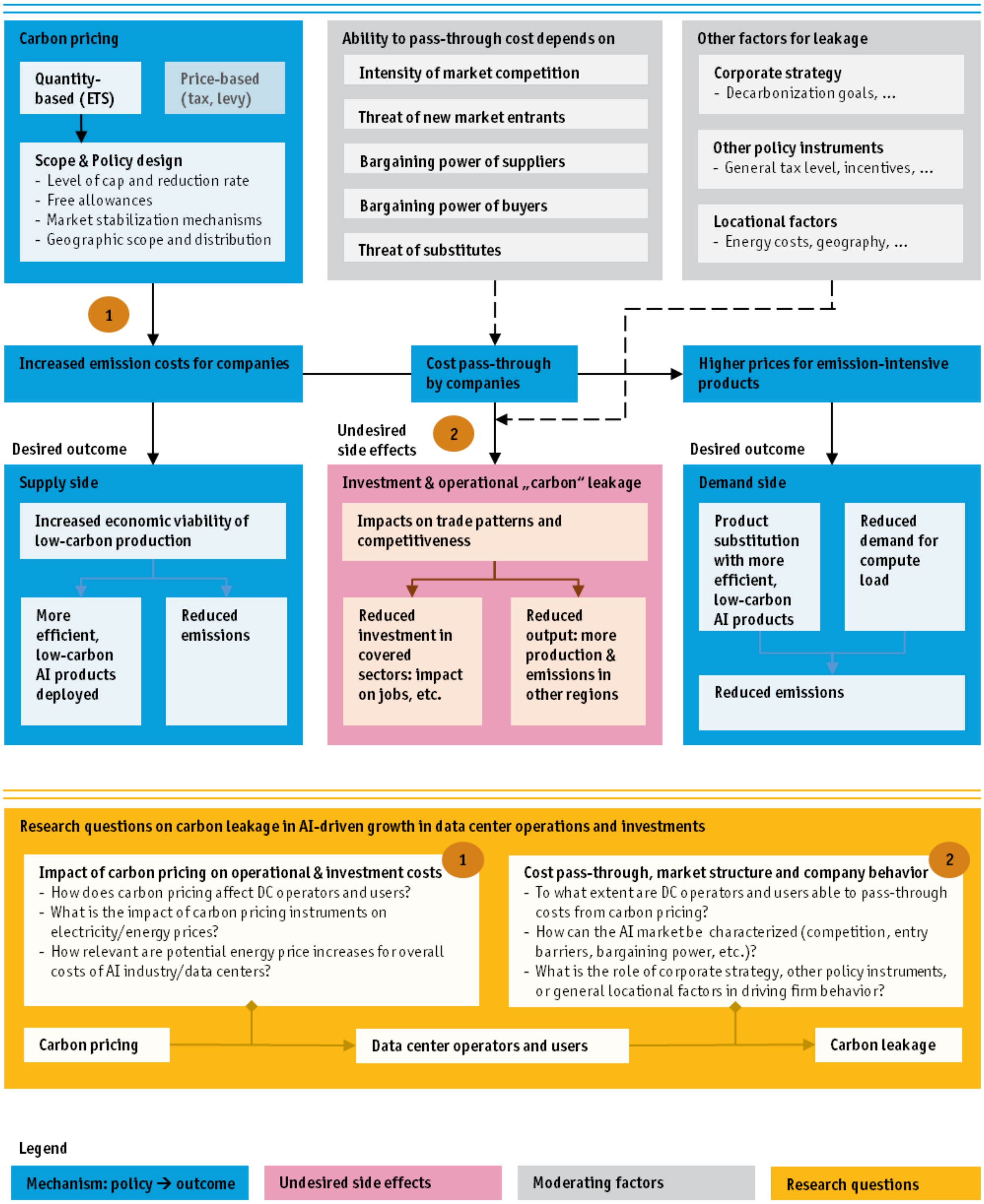
Figure 6: Digitally delivered services trade as a proxy for cross-border data flows.



Source: own illustration, INFRAS and Roegen Centre for Sustainability. Based on WTO estimates.

Figure 1Figure 6B provides a geographic overview on which countries are net importers (brown) and net exporters (green) of the indicator “computer services” measured in US Dollar for 2023. Most European countries are net importers of computer services. Major countries such as China, India, and the US are net exporters.

Figure 7: The relation between carbon pricing instruments and carbon leakage.


Carbon pricing
Quantity-based (ETS)
Price-based (tax, levy)
Scope & Policy design
- Level of cap and reduction rate
- Free allowances
- Market stabilization mechanisms
- Geographic scope and distribution
Ability to pass-through cost depends on
Intensity of market competition
Threat of new market entrants
Bargaining power of suppliers
Bargaining power of buyers
Threat of substitutes
Other factors for leakage
Corporate strategy
- Decarbonization goals, ...
Other policy instruments
- General tax level, incentives, ...
Locational factors
- Energy costs, geography, ...
1
Increased emission costs for companies
Cost pass-through by companies
Higher prices for emission-intensive products
Desired outcome
Undesired side effects
2
Desired outcome
Supply side
Increased economic viability of low-carbon production
More efficient, low-carbon AI products deployed
Reduced emissions
Investment & operational „carbon" leakage
Impacts on trade patterns and competitiveness
Reduced investment in covered sectors: impact on jobs, etc.
Reduced output: more production & emissions in other regions
Demand side
Product substitution with more efficient, low-carbon AI products
Reduced demand for compute load
Reduced emissions
Research questions on carbon leakage in AI-driven growth in data center operations and investments
1
Impact of carbon pricing on operational & investment costs
- How does carbon pricing affect DC operators and users?
- What is the impact of carbon pricing instruments on electricity/energy prices?
- How relevant are potential energy price increases for overall costs of AI industry/data centers?
2
Cost pass-through, market structure and company behavior
- To what extent are DC operators and users able to pass-through costs from carbon pricing?
- How can the AI market be characterized (competition, entry barriers, bargaining power, etc.)?
- What is the role of corporate strategy, other policy instruments, or general locational factors in driving firm behavior?
Carbon pricing
Data center operators and users
Carbon leakage
Legend
Mechanism: policy → outcome
Undesired side effects
Moderating factors
Research questions

Source: own illustration, INFRAS and Roegen Centre for Sustainability.

6 Assessing the extent of carbon leakage from AI-driven data center operations and investments

As argued in Section 4, current and expected data center capacity is predominantly located in jurisdictions without carbon pricing such as much of the US. This raises the question whether the relatively ambitious carbon pricing schemes in Europe have led to carbon leakage, i.e. the intentional shift in operations and investments of data centers beyond Europe. Figure 7 shows desired and undesired effects of carbon pricing on the supply and demand side, as well as how these effects translate into research questions on the potential for operational and investment carbon leakage from AI-driven data center growth.

6.1 Impact of carbon pricing on operational and investment costs

How can carbon pricing affect data center operators and users?

Carbon pricing instruments such as ETS can *increase the operational costs* for data center operators and users[19]. Operational cost can increase if the electricity consumed comes from power plants that are covered by carbon pricing.[20] Since under ETS power plants need to purchase emissions allowances, any increase in carbon costs may be passed on to consumers in the form of higher electricity prices. This can lead to higher costs for training and running AI models (for the AI model lifecycle see Figure 1 in Section 3). Data center operators and users are thus indirectly affected by ETS through its effects on electricity prices. This indirect effect on electricity prices may lead to i) short-term geographical shifts in the operation of data centers and ii) long term shifts in data center investment decisions. While the short term or operational carbon leakage (see definition in Section 2) would result from current electricity price differences, long term or investment-related carbon leakage would occur from *anticipated* electricity price differences over the next years and decades.[21] Finally, although it is likely a minor effect, carbon pricing instruments can potentially also *increase the investment costs*. Emission-intensive material needed by the data center operators can become more expensive because of carbon pricing of upstream industries such as steel or cement.

What is the impact of carbon pricing instruments on electricity prices?

The effect of carbon pricing on DC operators largely depends on its impact on electricity prices, as carbon costs from fossil fuel generation might be internalized. In an electricity market based on the merit order,[22] the impact of carbon prices on the electricity price depends on the technology, efficiency, and the amount of carbon emissions of the price-setting plant. Studies on whether this internalization takes place show mixed results. Next to carbon prices, long-term prices on future markets are influenced primarily by natural gas and coal prices, while short-term prices on spot markets depend on renewables and demand (Mosquera-López and Nursimulu, 2019). Under the EU ETS, Kosch et al. (2022) found a 1€/tCO_2 rise led to a 0.5-1€/MWh increase, depending on the power markets production portfolio. In Spain, a 1% carbon price change caused a 0.24% increase in electricity prices (Freitas and Silva, 2015). Most studies suggest only higher costs are passed to consumers, showing an asymmetric effect (Aatola,

[19] Data center operators include colocation operators, hyperscalers, or inhouse data centers at universities or companies. Data center users include AI developers and service providers as well as third parties using AI models through cloud services.

[20] Renewable power purchase agreements, which are predominantly used by hyperscalers, can also increase operational costs.

[21] The EU ETS II starting in 2027 will also put a price on fuel combustion in buildings and industrial activities which may affect fuel prices for back-up fossil turbines in data centers. However, given that backup capacity is rarely used, the effects on operational costs are likely to be low. From the expert interviews we gathered that in regions with reliable grids such as Germany or Switzerland, backup generators are fired up every 4-6 weeks for tests, lasting about 30-60 minutes.

[22] In short, the merit order is the ranking of electricity sources from cheapest to most expensive, ensuring the use of the most cost-effective options first and only turn to pricier ones when demand is higher.

Ollikainen and Toppinen, 2013). Short-term effects are evident in highly competitive markets like the Nordic region, where emission allowances influence electricity prices on spot markets (Jabłońska *et al.*, 2012). However, some studies show conflicting results. Jouvet and Solier (2013) found both positive and negative impacts of ETS on electricity prices across European markets (in some contexts, electricity prices decreased after an increase of the ETS price), especially in Italy. Wolff and Feuerriegel (2019) found a negative impact of emissions pricing on intraday market prices in phase III of the EU ETS. Outside the EU ETS, Palmer et al. (2018) found rising regional ETS prices in the US had little effect on prices. Woo et al. (2017) observed that a 1$/t increase in California's CO_2 price raised electricity prices by 0.15-0.59 $/MWh. Cotton and Mello (2014) found minimal impact from Australia's ETS, and Apergis (2018) identified a long-term asymmetric effect in New Zealand, with carbon price increases fully passed through.

What explains the impact of carbon pricing instruments on electricity prices?

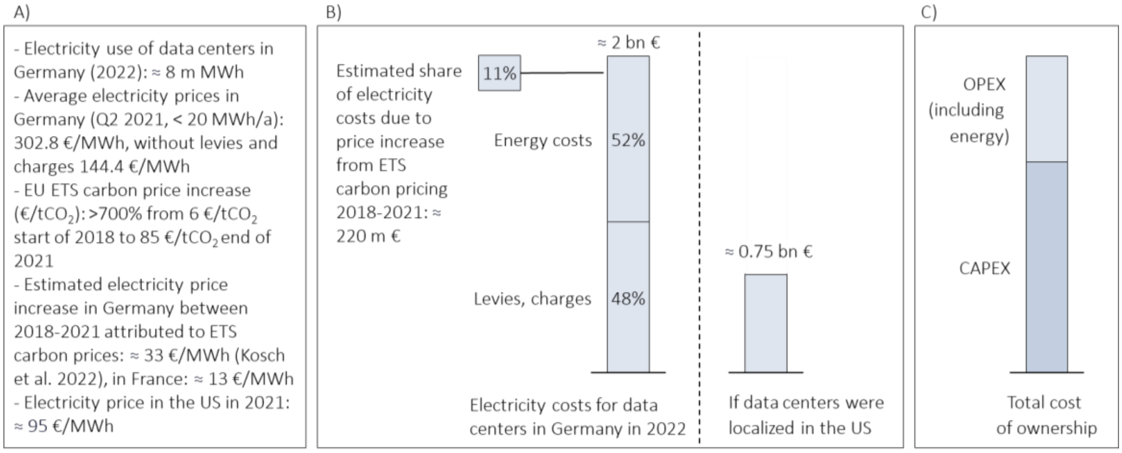
The *instrument design* is crucial to understanding impact on prices (Wolff and Feuerriegel, 2019). Historically, carbon pricing instruments have had modest ambition and thus low prices. In the EU ETS, recent changes in design elements may lead to higher electricity prices. Most importantly, in 2023, the reduction rate of the allowance cap has been increased to 2.2% per year. As the cap tightens, the scarcity of emission allowances increases, raising their price. Design changes are expected to lead to increases of carbon prices and reach between approximately 120 and 160 €/t according to most ETS models (Refinitiv, BloombergNEF, ICIS, Enerdata, PIK, CAKE) reviewed by Pahle et al. (2022). However, if low-carbon electricity sources expand and increasingly substitute electricity sources covered by carbon pricing, the overall effect on electricity prices can be much lower (see below). Pietzcker et al. (2021) find that under the scenario of a strongly tightening EU ETS cap (−63% of allowances instead of −43% in 2030), total discounted power system costs in the EU would only increase by 5% (€3680 billion vs. €3500 billion) for the period 2018–2052. Such moderate impact of the ETS on electricity prices even under a tightening cap is largely explained by the dominance of *supply factors* like the diffusion of renewable energy technologies, general prices for fossil fuels, and *demand factors* such as overall economic activity in shaping electricity prices. In the EU, the share of clean electricity generation from renewables has continuingly increased (from almost 0% in 2000 to 27% of electricity generation in 2023), decreasing demand for emission allowances from the electricity sector. In 2023, wind produced more electricity than gas for the first time (Ember, 2024). This trend is also visible on the global scale with emission intensities of the electricity mix in all world regions expected to moderately fall from 2023 until 2030 as the share of renewables in final energy consumption is expected to increase from 13% to 20% (IEA, 2024a, 2024b), although this trend is likely to be heterogenous across countries. For example, emission intensities in China are expected to fall from roughly 600 gCO2/kWh in 2022 to below 400 gCO2/kWh until 2030 with currently implemented policies (IEA, 2024b). However, China is a key source of uncertainty in projecting future developments (IEA, 2024a). Finally, carbon pricing affects only part of the final electricity price for end consumers. *Regulatory factors* such as taxes and grid fees account for a significant share of the price (these can also be related to climate policy but are not covered due to the limited scope of this study).

How relevant are potential electricity price increases from carbon pricing?

Even if carbon pricing affects electricity prices to a significant degree, the question remains whether such price differential matters for data center operators and users. Based on a set of simple assumptions, a back of the envelope calculation for the example of data centers in Germany in 2022 (Figure 8) shows that electricity price increases from carbon pricing can account for a significant increase of electricity costs (an estimated cumulative 11% from ETS price increases 2018-2021) for data centers.[23] This particular case represents an upper bound for the effects of ETS on electricity prices, due to the strong increase of ETS prices in the examined period and the selection of a country with significant share of coal and gas in the electricity generation. For the few countries such as France with substantially lower shares of fossil electricity generation the calculation would yield a remarkable lower estimate. In total, such electricity price increase may have led to an increase of operational expenditure (OPEX) for data centers of around cumulative 220 m € considering the carbon price increase 2018-2021.

Figure 8: Back of the envelope calculation on the role of ETS prices on data center OPEX.

It shows the estimated impact of ETS price increases 2018-2021 for electricity costs by German data centers in 2022. Given strong ETS price increase and fossil shares in the German electricity mix in this period it constitutes an upper bound.



A)

- Electricity use of data centers in Germany (2022): ≈ 8 m MWh
- Average electricity prices in Germany (Q2 2021, < 20 MWh/a): 302.8 €/MWh, without levies and charges 144.4 €/MWh
- EU ETS carbon price increase (€/tCO_2): >700% from 6 €/tCO_2 start of 2018 to 85 €/tCO_2 end of 2021
- Estimated electricity price increase in Germany between 2018-2021 attributed to ETS carbon prices: ≈ 33 €/MWh (Kosch et al. 2022), in France: ≈ 13 €/MWh
- Electricity price in the US in 2021: ≈ 95 €/MWh

B)

Estimated share of electricity costs due to price increase from ETS carbon pricing 2018-2021: ≈ 220 m €

Source: own illustration, INFRAS and Roegen Centre for Sustainability. Based on Statistisches Bundesamt, World Bank Carbon Pricing Dashboard, European Commission (2024), Kosch et al. (2022) and Table 2 above. Approximate distribution of total cost ownership (C) based on expert interviews for an average data center in Europe.

This rough approximation needs to be compared to the estimated 2 bn € that data centers spent overall on electricity in 2022 and the other cost components of electricity such as levies and charges (Figure 8B). It also needs to be compared to the cost differential that results from hypothetically running these data centers in the US with its much lower electricity costs (here: average industrial electricity price that disregards local price differences within the US), resulting in hypothetical electricity expenditure of roughly 0.75 bn €. Finally, electricity costs are not the only cost factor of data centers and account only for maximally a third of the total cost of ownership (based on expert interviews), see Figure 8C. More importantly, capital expenditure (CAPEX) for IT equipment (GPUs, CPUs) and infrastructure (racks, cooling systems, servers) make up a larger part of total cost of ownership.

Data on the cost structure of AI model development suggests that energy costs (assuming US energy costs; $85/MWh) account for a moderate part of the overall costs, e.g., 6% for GPT-4 or 4% for Gemini 1.0 Ultra (Cottier *et al.*, 2024). Energy costs for interference may be higher compared to these numbers given that, e.g., spending on R&D will be smaller. Moreover, assuming that electricity demand stays constant the impact of carbon pricing on electricity prices is likely to be mitigated to some extent in most countries over the next years even under a tightening of carbon pricing instruments (as targeted by the EU Commission, the share of renewables is supposed to increase from 24 % to 42% within the EU).

[23] Assuming that the strong ETS price increase between 2018 and 2021 has led to an electricity price increase of 33 €/MWh in Germany (Kosch, Blech and Abrell, 2022).

Summary of chapter 6.1

Short term, operational carbon leakage may occur if ETS increases current electricity prices. Long term carbon leakage may occur if data centers investments are shifted geographically due to anticipated electricity price differences due to carbon pricing. To assess the role of ETS for electricity prices, this chapter has qualitatively reviewed existing literature. In sum, we deduce that carbon pricing is likely to have had a moderate effect on electricity prices until 2021 (in the EU but also in other jurisdictions), however, more current insights are missing. Moreover, the effect of carbon pricing on electricity prices is contingent on the electricity mix of each country. Itis likely to be more significant in countries with higher shares of fossil fuels in electricity generation and to become weaker when the reliance on fossil fuels decreases. Still, even if carbon pricing leads to significantly higher electricity prices in an economic meaningful way, it is an open question to what extent this electricity price increase relates to carbon leakage, i.e. intentional shifts in data center investments due to the EU ETS. We discuss this question in the next section.

6.2 Carbon leakage from data center operations and investments

What do we know about carbon leakage from data center operations and investments?

The discussion above shows that there is potential for carbon leakage if climate policy costs increased substantially, and companies decide to act on it. To our knowledge, there are no studies that explicitly address carbon leakage in data center operations or investments or the wider ICT sector yet. Empirical studies from other sectors generally find *evidence* of moderate carbon leakage. For instance, an analysis by the OECD (2024) suggests that carbon leakage through international trade offsets around 13% of domestic emission reductions. Wang and Kuusi (2024) confirm that EU ETS has caused the carbon content of the EU's imports to increase to a moderate extent. Saussay and Sato (2024) find that energy price differences are related to investment location decisions. Nonetheless the impact is heterogeneous. Saussay and Sato measure, no effect in most cases, especially between industrialized countries, but stronger effects in energy-intensive industries. In a literature review, Verde (2020) finds no clear evidence of carbon leakage in earlier EU ETS phases. Grubb et al. (2022) confirm these findings and explain the absence of leakage in early phases mainly with the grandfathering of allowances which shields key sectors and maintains competitiveness of European firms (Joltreau and Sommerfeld, 2019).

Also, most studies use data from periods of low CO_2 prices (Saussay and Sato, 2024). More recently announced protective measures such as the EU carbon border adjustment mechanism (CBAM) which may affect forward-looking investment decisions are also not yet examined empirically. Furthermore, carbon leakage effects on long-term investment decisions are difficult to model and only rarely examined specifically.

How likely is carbon leakage in AI-related data center operations and investments?

Carbon leakage is of particular concern in energy-intensive sectors of an open economy with trade competition, in which companies may struggle to pass on these carbon costs along the value chain and thus decide to offshore their production (Saussay and Sato, 2024). Most affected are companies in the cement, lime and plaster, fertilizer, or steel industries (Sato *et al.*, 2015; Grubb *et al.*, 2022). These industries are affected both directly through the ETS as onsite emitters as well as indirectly through higher energy costs. As indicated in Figure 8 above, the energy costs account for a majority of data center OPEX, which, in turn, account for roughly a third of total cost of ownership for data centers. Data centers can thus be considered an energy-intensive sector (comparable to air transport and higher than cement or iron and steel production, see Figure 10 in Appendix D).

Besides the energy cost share in overall production cost of an industry, a key determinant for carbon leakage is the *ability of companies to pass on costs* from climate policy along the value chain (i.e., the lower this ability, the higher the chance for leakage). Companies' ability to do so depends on a set of factors commonly referred to as *Porter's five forces* which basically describe the structure and nature of a given market section: market competition, new entrants, supplier and buyer power, as well as substitutes. In a perfect market with fierce competition and constant marginal costs, companies would have no other choice than a 100% pass-through of additional costs. However, from a theoretical perspective, if marginal costs are not constant but increase, monopolistic markets will show greater cost pass-through rates (Ritz, 2024). Moreover, it is possible that data centers have increasing marginal costs due their large investment and idle-state costs.

We describe the market in data center operation and use, as it relates to AI in Table 5 below. The table is based on our expert judgements.[24] To the best of our knowledge, there are no studies describing the market structure of the data center and AI market so far and how it affects location decisions. Future research is required for robust assessments of the AI market structure.

Table 5: Market structure in data center operation and use.

Actor Type	Market Competition	New Entrants	Supplier Power[25]	Buyer Power	Substitutes
Colocation data center operators	Medium, with several competitors, but fairly high switching costs due to inertia and lock-in.	Low - Medium-High entry barriers due to necessary capital expenditure, specialized knowledge.	Medium, with traditionally monopolistic energy markets, but reforms and growing shares of renewables increase competition.	Low to moderate depending, e.g., on size of buyer and size of colocation operator.	Low to medium - end user devices are the closest substitute and not able to process fast amounts of data.
Hyperscalers and cloud services	Medium - Established market and oligopoly with large players including Amazon AWS, Google Cloud.	Low -Very high entry barriers due to high capital expenditure, technology know-how.	Medium, with traditionally monopolistic energy markets, but reforms and growing shares of renewables increase competition.	Low, as consumers have little negotiating power over compute costs.	Low to median - end user devices are the closest substitute and not able to process fast amounts of data.

Source: own table, INFRAS and Roegen Centre for Sustainability.

Considering investment carbon leakage, based on these characteristics the data center market faces a low level of competition and it is likely that they will pass-through costs if market characteristics do not change (cf., OECD, 2022c).

[24] Calculating the Herfindahl-Hirschman Index (HHI) is beyond the scope of this study. Another common indication for market power is the profit margin, as low competition and monopolistic markets are often associated with very high margins. However, only a limited number of sources report profit margins for DCs, which vary widely between 5% and 50%, providing no clear indication regarding the intensity of competition.

[25] This is only discussed from an energy perspective, e.g., AI hardware suppliers (such as Nvidia) are not discussed.

With respect to short-term carbon leakage, we provide the following arguments:

- In a typical pricing model, colocation data center operators can pass-through higher electricity costs from carbon pricing to customers. While customers have the option to switch colocation operators, it is unlikely to happen below a certain electricity cost threshold. For instance, insights from an expert interview suggest that the energy crisis in 2022/2023 with strongly increasing electricity prices in Europe has not led to major geographical shifts to colocation data centers outside of Europe.
- Hyperscalers offering cloud services may set different prices according to different end uses. It is not clear to what extent prices differ according to the location of data centers used, i.e. along different electricity price levels.
- From the perspective of data center customers, cost pass-through may differ also according to AI application and thus different electricity needs and customer base (due to different price elasticities). For example, ultra-high frequency applications like AI-assisted financial trading may be more or less prone to cost pass-through than occasional training runs at supercomputers for research purposes.

These are current market conditions and competition intensity may increase over time which can affect cost past-through rates and cost structures.

There are, to our knowledge, no studies on the cost pass-through in the data center and AI sector. Data from other sectors can be informative: A study on behalf of the European Commission has examined cost pass-through rates in European manufacturing sectors (Directorate-General for Climate Action, 2015). Cost pass-through seems to be present in many sectors and countries, but heterogenous. Indicative cost pass-through rates – if present – range between 35% to more than 100% depending on the sector and country, see also (UBA, 2020).

6.3 Other factors driving data center operations and investments

The decision to (re)locate investment and shift operations in data centers can only be partially driven by carbon prices and their effects on current or – relevant for investment decisions – anticipated operational expenditure. Data center localization also depends on other factors namely corporate strategy, non-climate related policy, and general locational factors. These may moderate the potential effect of carbon pricing on short-term operational shifts in compute loads and long-term investment decisions.

Corporate strategy by data center operators and users

Even if able to pass on costs imposed by climate policy, data center operators and users *may not wish to do so* due to corporate strategy. This can have several aspects:

- **Stage of the industry life cycle:** The AI industry is currently in the growth stage of the industry life cycle (McKinsey, 2024)., Rather than maximizing efficiency and profitability, in this phase companies typically focus on market shares and scaling. In platform-based business models such as used in many AI applications, the first mover advantage can be significant as benefits may arise from capturing a large market share and using resulting network effects. Thus, there is a strong emphasis on quickly developing new AI technologies and applications and differentiate AI products and services. Efforts are directed towards scaling operations and infrastructure to handle increased demand. This affects AI developers, and their operational decision-making, as well as data center siting and investment decisions by large corporations who provide both computing infrastructure and AI solutions.

- **Climate commitments:** At the same time, major data center operators and users are focusing on reducing the climate footprint in their value chain. Non-binding initiatives like the European Code of Conduct on Data Centre Energy Efficiency and private sector commitments, such as the Climate Neutral Data Centre Pact and RE100, showcase voluntary efforts towards energy efficiency and renewable energy use. These initiatives suggest that many data center operators are already pursuing environmental goals, reducing the likelihood of significant regulatory-driven carbon leakage. Table 10 in Appendix E provides an overview on voluntary cross-company initiatives. Table 6 shows the extent to which these players source their electricity from renewable power purchase agreements (PPAs) and sums up key aspects of their climate commitments. One key question will be how these players will meet their climate targets in time (and with conflicting timelines of renewables or nuclear built out vs. data center built out). Green washing claims (mainly due to the purchase of renewable energy certificates and different methods of calculating Scope 2 & 3 emissions) also exists, however, these are heterogeneously distributed across companies.[26]

Table 6: Corporate climate commitments by four of the largest data center users.

Company	Cumulative renewable PPA offtake in GW (and share of global volume of PPA)	Summary of climate commitments
Amazon	27.8 GW (16.3%)	- Co-founded The Climate Pledge (net-zero by 2040). - Invests in technologies to accelerate its path to net-zero.
Meta	12.1 GW (7.1%)	- Aims to reduce Scope 1 and 2 emissions by 42% by 2031. - Achieved net-zero for global operations in 2020.
Microsoft	10.8 GW (6.3%)	- Pledged to be carbon negative by 2030. - Created a $1B Climate Innovation Fund for sustainable tech.
Google	9.9 GW (5.8%)	- Targets 50% Scope 1 and 2 emissions reduction by 2030. - Plans to operate on 24/7 carbon-free energy globally by 2030.

Source: own table, INFRAS and Roegen Centre for Sustainability. Based on Bloomberg NEF PPA Data, corporate sustainability reports.

Further EU policy instruments and regulation

Public policy beyond carbon pricing instruments can impact localization decisions of data center operators and users (Ebert *et al.*, 2024)[27] or the integration of emissions into ETS systems. **Fehler! Verweisquelle konnte nicht gefunden werden.**7 summarizes the impact of major p olicies within the EU related to the location on data centers (Table 10 in Appendix E provides more detail).

Potentially relevant regulations include the EU Carbon Border Adjustment Mechanism (CBAM), which imposes costs on imported carbon-intensive goods and aims to level the playing field between operations inside the EU and those located in regions with less stringent climate

[26] O'Brien I. (15 September 2025). Data center emissions probably 662% higher than big tech claims. Can it keep up the ruse? https://www.theguardian.com/technology/2024/sep/15/data-center-gas-emissions-tech (Accessed: 2 December 2024).

[27] Two datasets map relevant digital/AI policies: Marcus, J. S., Sekut, K., Zenner, K. (06 June 2024). A dataset on EU legislation for the digital world. https://www.bruegel.org/dataset/dataset-eu-legislation-digital-world ; OECD.AI (2021), powered by EC/OECD (2021), database of national AI policies. https://oecd.ai/en/dashboards/overview/policy (Accessed: 2 December 2024).

policies. However, the supposed "leakage prevention"-effect on data centers is minimal/zero since it does not include services in general. Another potentially relevant instrument are electricity price compensation mechanisms under the ETS (implemented by member states such as Germany) which aim to prevent carbon leakage by offsetting high electricity costs for manufacturing. However, AI-related data center compute is not covered under these measures.

Further, the EU Energy Efficiency Directive (EED) has indirect effects on data centers, encouraging countries to meet energy-saving targets. The German "Energieeffizienzgesetz" (EnEfG) from 2023, however, might have some influence, since (unlike the EU regulation which focuses on reporting and only encourages efficiency measures for DCs) it imposes strict energy efficiency and renewable energy targets for data centers. While many large IT companies already meet these efficiency standards, meeting (on balance) the 100% renewable electricity sourcing target from 2027 onwards as well as waste heat recovery may be a challenge to some, especially the latter.

Table 7: Important policies affecting data center operation and investment.

Policy	Short description	Influence on carbon leakage from data centers
EU Carbon Border Adjustment Mechanism (CBAM)	A mechanism to price carbon emissions of imports like steel, cement, and electricity to prevent carbon leakage and promote cleaner production globally.	Limited impact on data centers as they consume EU-produced electricity, regulated under the EU ETS. CBAM focuses on imported products with high emissions, not services. IT, data center, and AI services are excluded.
EU Energy Efficiency Directive (EED)	Sets energy savings targets for countries, including measures for data centers like energy audits, efficiency reporting, and management systems.	Obligations like reporting and audits marginally increase operational costs. However, leading IT firms already meet/exceed efficiency standards, reducing carbon leakage risk. IT associations are concerned about the potential relocation of businesses within the European market when it comes to the use of waste heat.[28]
EU AI Act	Regulates AI systems based on risk levels, requiring compliance for high-risk applications through special approvals and oversight.	Regulates development and deployment of high-risk AI within the EU. Provides transparency on energy consumption used for AI training abroad.
General Data Protection Regulation (GDPR)	Protects personal data privacy, including restrictions on cross-border data transfers outside the EU.	Partially prevents offshoring of data processing involving EU citizens, limiting carbon leakage risks. International privacy laws similarly restrict data use. However, data can be shifted across borders if countries have comparable data policy protection policies.

Source: own table, INFRAS and Roegen Centre for Sustainability.

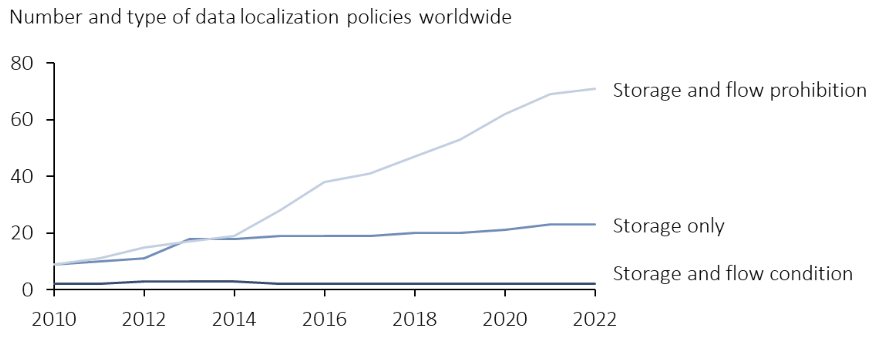
Moreover, data sovereignty and localization policies such as the General Data Protection Regulation (GDPR) that aim at restricting the use of data to geographical boundaries put limits on how fungible data can be used across jurisdictions (Selby, 2017; Cory and Dascoli, 2021; CIPL, 2023). A good example for this is restrictions on the use of medical data outside of national borders. AI is also covered by international trade agreements like the WTO's GATS and regional agreements under regulation for trade in services (OECD, 2022a). As highlighted by the OECD's Services Trade Restrictiveness Index (STRI), domestic regulations present significant barriers in

[28] Bitkom (HRSG). Presseinformation: Bitkom zur Verabschiedung des Energieeffizienzgesetzes. https://www.bitkom.org/Presse/Presseinformation/Bitkom-Verabschiedung-Energieeffizienzgesetz (Accessed: 2 December 2024).

areas like communication infrastructure and data connectivity. Over recent years, the regulatory environment for digitally enabled services trade has become increasingly restrictive. Figure 9Figure 8 below shows this trend based on data from the OECD on data localization policies[29].

The EU AI Act puts constraints and regulatory costs on AI development and deployment of high-risk and general-purpose AI applications used in the EU. A report for the European Commission estimates the total global compliance cost for the AI industry in 2025 to range from € 1.6 billion to € 3.3 billion, assuming 10% of AI units are classified as "high risk" (Renda *et al.*, 2021). Given that all AI systems put into service in the EU are regulated, independent of who is developing them and where the compute is located, we do not anticipate geographic shifts from this regulation. In addition, the AI Act will provide transparency for the energy used to train AI models, but may not for the inference stage or embodied carbon emissions (this will depend on the specific guidelines of the AI Act that are currently defined). This may help to shed more precise light on the environmental impacts of AI.

Figure 9: Diffusion of data localization policies over time.



Source: own illustration, INFRAS and Roegen Centre for Sustainability. Based on OECD data on "Cross-border data flows" available under https://www.oecd.org/en/topics/cross-border-data-flows.html, accessed on 11 November 2024.

As of early 2023, nearly 100 data localization measures were implemented across 40 countries, over half introduced within the last decade. These measures have become more restrictive, with more than two-thirds requiring data storage within national borders and restrictions on data flows. These considerations are partially integrated in the estimations from Sections 3 & 4.

Other locational factors

Additional relevant location factors for data centers operators and users beyond specific policies or corporate strategies can affect siting decisions and may further diminish the effect of carbon prices on localization of data centers. They include[30]:

- **Focus on securing energy supply:** As data center growth is expected to significantly increase energy demand, a key struggle for AI companies is to secure reliable electricity supply in time. Due to interconnection queues, general long-time horizons of energy projects (compared to data center construction time), and network bottlenecks, securing enough electricity is likely to be more important than electricity price differences. According to Morgan Stanley, data center developers may even be willing to pay a large premium to come online fast (Morgan Stanley, 2024). In addition, data centers have a demanding load profile

[29] It remains open to what extent these policies are implemented, and how stringent they are.
[30] Kamiya, G. & Kvarnström, O. Data centres and energy – from global headlines to local headaches? https://www.iea.org/commentaries/data-centres-and-energy-from-global-headlines-to-local-headaches (Accessed: 2 December 2024).

as they need 24/7 uninterrupted power (momentary lapses in power can cause major economic damages) with multiple layers of backup power generation. However, a stable energy supply is more difficult to achieve with a large share of renewables.[31]

- **Connectivity and infrastructure**: A location with robust, reliable access to high-speed data networks is crucial for data center operations. Data centers require fast, secure internet connectivity to ensure minimal downtime. High-quality IT infrastructure with high bandwidth and low-latency connections is essential for smooth data processing and communication between facilities and end-users (i.e., clients). These aspects are typically prioritized in locations near major telecom hubs and energy grids (e.g., major cities).

- **Availability of water**: Access to water can be essential for non-closed cooling systems in data centers, particularly in energy-intensive AI data centers and in areas where air cooling is not sufficient. Data centers often require large amounts of water to prevent overheating, especially in warmer regions.

- **Regional industry networks and collaboration**: Proximity to other technology companies or universities can be key for driving innovation or finding appropriately skilled workforce. Skilled professionals are necessary to operate and manage data centers, as they are needed for tasks such as IT support and network management. Data centers benefit from being in areas with a concentration of digital businesses and research institutions.

- **Climate conditions**: A cooler climate can significantly reduce energy costs associated with cooling, which accounts for a large proportion of a data center's total energy consumption. Natural methods are more effective in regions with cooler temperatures, reducing the need for air conditioning systems. Nordic regions are particularly attractive to data center operators for their favorable cooling climate, and some even use free air cooling year-round.

- **Natural hazards and security**: The risk of natural disasters, such as floods or earthquakes, , can disrupt operations or damage infrastructure. Reducing exposure to such risks is important for operational reasons but also for mitigating insurance and recovery costs.

- **Additional cost incentives**: Different types of costs vary immensely between locations. For instance, within the US data centers are frequently build in rural areas where land is more affordable.[32] Also, profit taxes are often a significant factor in the decision-making process for firms' locations (Devereux & Griffith 1998). Ireland has one of the lowest corporate tax rates in Europe, at around 12.5%, and hosts most data centers, after Germany. Also, many US states explicitly offer tax incentives to encourage data centers to establish themselves within their borders.[33] Moreover, also energy price differences in general (independent of carbon pricing) may play a part (Figure 8).

- **Moratoriums:** Regional data center moratoriums are temporary bans or restrictions enacted by local governments to halt the construction of new data centers. These measures are driven by concerns over the environmental impact, high energy consumption, and strain on infrastructure resulting from the rapid expansion of data centers and have been already

[31] Hirschhorn, P. and Brijs, T. (December 17, 2021). Rising to the Challenges of Integrating Solar and Wind at Scale. https://www.bcg.com/publications/2021/addressing-variable-renewable-energy-challenges (Accessed: 2 December 2024).

[32] By Rayome, A. D. Why data centers fail to bring new jobs to small towns. https://www.techrepublic.com/article/why-data-centers-fail-to-bring-new-jobs-to-small-towns/ (Accessed: 2 December 2024).

[33] SDIA (HRSG). US tax incentives for data centers by state. https://knowledge.sdialliance.org/policies/us-tax-incentives-for-data-centers-by-state (Accessed: 2 December 2024).

implemented in Singapore, Netherlands, Ireland, Germany, USA, and the UK (Soares et al. 2024). These measures may spark the relocation of data centers in less regulated areas.

Summary of chapter 6

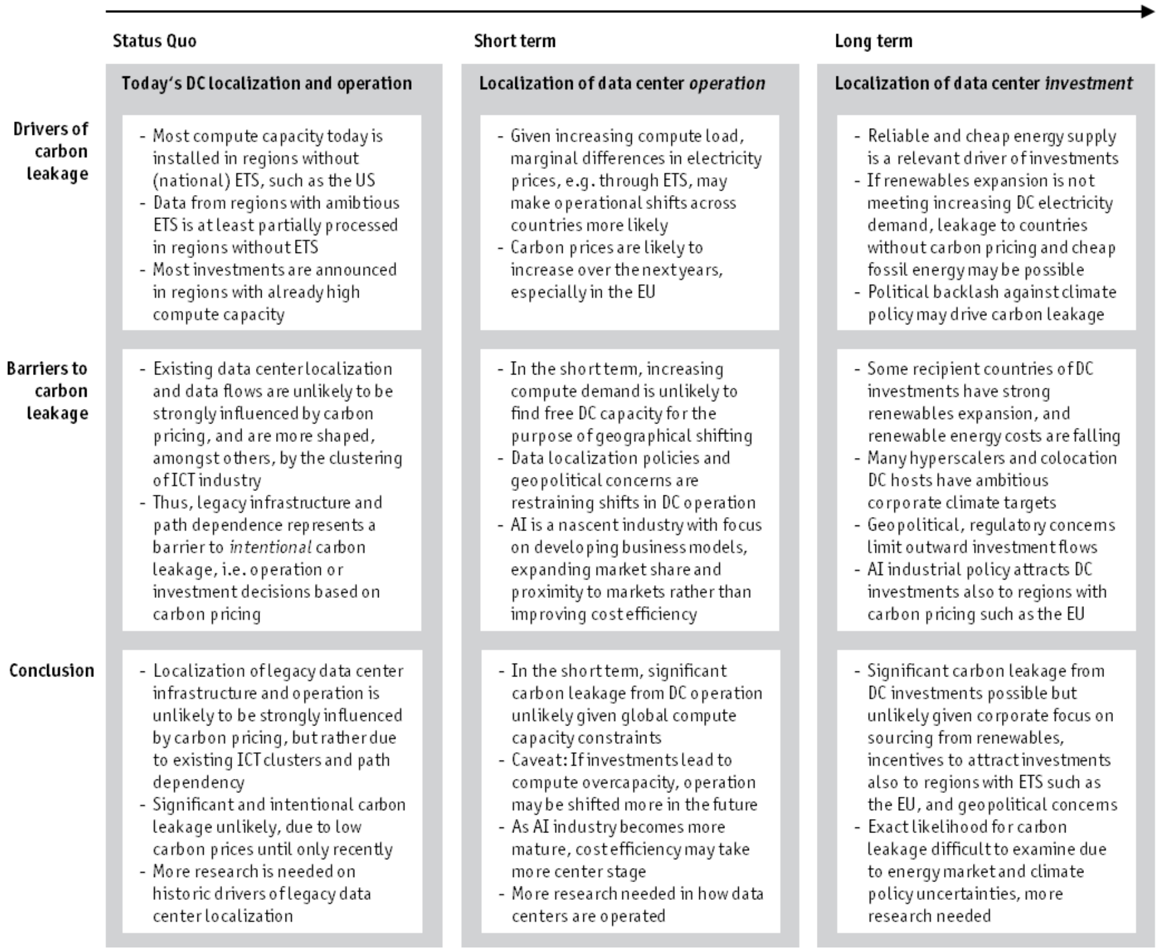
In sum, even if carbon pricing led to significant electricity price increases and thus higher operational expenditure for data centers, the effects on carbon leakage are likely rather limited. Intentional carbon leakage by data center operators and users such as AI developers to lower operating costs seems rather unlikely due to the current market phase. The AI market is, at the moment, in its infancy, with companies focusing on market share expansion and product development rather than on cost efficiency. Also, the market structure that is far from being a perfect market (e.g. with a few dominant players in AI model development) should enable both data center operators and users to pass-through costs along the value chain. Moreover, energy costs only account for a part of overall production costs of final AI products. However, more research is needed to provide a more robust assessment of cost pass-through, market structure, and energy cost shares in the AI-driven data center growth. Also, market structure and cost pass-through may change as the AI market evolves over the coming years. Finally, to fully assess the likelihood of carbon leakage, a wider set of factors beyond climate policy needs to be considered including corporate strategies or data localization policies.

7 Conclusion and outlook

7.1 Summary of the potential, drivers, and barriers to AI carbon leakage

AI-driven data center growth is predominantly projected in regions with already high compute capacity such as the US or China. Most growth is thus happening in regions with none or not sufficiently stringent carbon pricing mechanisms. Europe has implemented an ambitious emissions trading system (ETS), but its installed compute capacity and projected growth remain modest. This uneven climate policy and data center landscape raises concerns about a potentially large amount of emissions in digital trade and carbon leakage, where data center investments and AI compute loads might be geographically shifted to avoid costs imposed by carbon pricing. With respect to shifted emissions, we estimate the theoretical upper limit at hundreds of kilotons CO_2 today and thousands of kilotons CO2 in a few years (see Table 4). However, the report argues that the likelihood for significant carbon leakage from carbon pricing, specifically from ETS, to be moderate, as summarized in Figure 10 below. The actual carbon leakage induced by ETS from AI data center operations should thus be far below this theoretical upper bound. In the short term, the global strain on compute capacity makes intentional geographic shifts due to carbon pricing unlikely, as the top priority for data center users is to secure sufficient compute capacity.

Figure 10: Drivers and barriers to carbon leakage for data center operation and investment.



	Status Quo	Short term	Long term
	Today's DC localization and operation	Localization of data center *operation*	Localization of data center *investment*
Drivers of carbon leakage	- Most compute capacity today is installed in regions without (national) ETS, such as the US - Data from regions with amibtious ETS is at least partially processed in regions without ETS - Most investments are announced in regions with already high compute capacity	- Given increasing compute load, marginal differences in electricity prices, e.g. through ETS, may make operational shifts across countries more likely - Carbon prices are likely to increase over the next years, especially in the EU	- Reliable and cheap energy supply is a relevant driver of investments - If renewables expansion is not meeting increasing DC electricity demand, leakage to countries without carbon pricing and cheap fossil energy may be possible - Political backlash against climate policy may drive carbon leakage
Barriers to carbon leakage	- Existing data center localization and data flows are unlikely to be strongly influenced by carbon pricing, and are more shaped, amongst others, by the clustering of ICT industry - Thus, legacy infrastructure and path dependence represents a barrier to *intentional* carbon leakage, i.e. operation or investment decisions based on carbon pricing	- In the short term, increasing compute demand is unlikely to find free DC capacity for the purpose of geographical shifting - Data localization policies and geopolitical concerns are restraining shifts in DC operation - AI is a nascent industry with focus on developing business models, expanding market share and proximity to markets rather than improving cost efficiency	- Some recipient countries of DC investments have strong renewables expansion, and renewable energy costs are falling - Many hyperscalers and colocation DC hosts have ambitious corporate climate targets - Geopolitical, regulatory concerns limit outward investment flows - AI industrial policy attracts DC investments also to regions with carbon pricing such as the EU
Conclusion	- Localization of legacy data center infrastructure and operation is unlikely to be strongly influenced by carbon pricing, but rather due to existing ICT clusters and path dependency - Significant and intentional carbon leakage unlikely, due to low carbon prices until only recently - More research is needed on historic drivers of legacy data center localization	- In the short term, significant carbon leakage from DC operation unlikely given global compute capacity constraints - Caveat: If investments lead to compute overcapacity, operation may be shifted more in the future - As AI industry becomes more mature, cost efficiency may take more center stage - More research needed in how data centers are operated	- Significant carbon leakage from DC investments possible but unlikely given corporate focus on sourcing from renewables, incentives to attract investments also to regions with ETS such as the EU, and geopolitical concerns - Exact likelihood for carbon leakage difficult to examine due to energy market and climate policy uncertainties, more research needed

Source: own illustration, INFRAS and Roegen Centre for Sustainability.

Over the long term, shifts in investments in new data centers to avoid costs from carbon pricing instruments appear more plausible. In particular, they will become more likely if competition intensity across data centers increases. In contrast, a rapid uptake of renewable energy in countries hosting data centers can prevent carbon leakage (and embodied emissions from digital trade). Also, carbon leakage is likely to remain small if electricity costs become increasingly decoupled from carbon pricing with higher shares of renewables.

Further, factors such as more general energy price differences, data localization and security policies, and broader corporate strategies are likely to currently exert greater influence on data center siting decisions than carbon pricing alone. Many data center operators and users have adopted ambitious climate commitments, prioritizing clean energy for new investments.

Still, major pushbacks against clean energy deployment, such as to be expected under the second Trump administration, and a potential reversal of climate commitments by corporates may provide ground for expecting carbon leakage in the medium to long term. While it is unlikely that this investment decision is primarily driven by cost avoidance from carbon pricing, it may still lead to carbon leakage in the medium term: in case of compute overcapacity, data center users may choose to run their models in data centers with lower operating costs.

While carbon leakage examined in this study—intentional relocation to evade carbon pricing from ETS—is estimated to be unlikely for AI-driven data center growth, the broader climate implications of data center growth warrant attention. Data center growth in regions without robust climate and energy policies may exacerbate emissions, either directly through reliance on fossil fuels or indirectly by displacing other uses of clean electricity to decarbonize transport or industry. Such negative second-order effects on overall decarbonization efforts may be particularly critical in key markets with a slowdown in renewable energy deployment. Beyond carbon pricing, other climate policy instruments such as renewable energy policy and its implications for electricity prices (e.g., potential price increases via network charges or other levies on the electricity price) may also cause carbon leakage. Finally, from a European perspective, the question is how to address the potential of increasing imported emissions from compute load abroad.

7.2 Research agenda on the direct climate effects of data center growth

From our report, several questions emerge that deserve future research and require more granular data and sophisticated methods for a detailed analysis of AI-related carbon leakage.

- **Geographic distribution of AI-related compute growth**: While the proxy used in this study yields a reasonable estimate, more precise data on the actual distribution of AI compute capacity, and related energy consumption, is needed. As soon as such data becomes available, it should substitute the estimates employed here. More granular data should be used for the US as well: With very different data center development, grid carbon intensities, and partly existing ETS, the US is a very heterogeneous landscape. As it also harbors almost half of the worldwide AI compute capacity, a more granular analysis of the US is required.

- **Flexibility and shiftability of AI loads**: The analysis on shiftability and flexibility of AI compute loads in Section 3.3 relies entirely on assumptions. More research is needed on this topic that has only barely been touched upon so far. Research questions may include: How does flexibility of AI loads differ between AI applications? What are technical and other conditions that need to be met for flexible AI loads? How does shiftability of AI loads differ along different temporal resolutions, i.e. how flexible and shiftable are AI loads on a second, minutes, vs. hourly or daily basis?

- **Comparing the evolving climate policy and data center landscape**: Next to assessing the distribution of carbon pricing policies and their overlap with data center growth, other climate policy instruments such as clean energy policy subsidies and regulation or efficiency requirements would be relevant. Novel data allow the systematic mapping of climate policy instrument mixes, e.g. by the OECD or IEA, can help (Steinebach *et al.*, 2024).

- **Influence of carbon prices on electricity prices**: There is only little research on the effect of carbon pricing on electricity prices, and this literature does not provide a clear picture. The following questions deserve further analysis: What is the effect of carbon prices on electricity prices (in specific regions)? How does it compare to other components of electricity prices such as fuel costs, network tariffs, or other taxes?

- **Relevance of electricity prices for AI compute and overall AI production costs**: Building on a more robust insight into the effect of carbon prices on electricity prices, the question remains to what extent electricity prices are important for data center operators and their clients. Here, the concept of cost pass-through is key. While literature has established cost pass-through rates, uncertainties about the extent of carbon leakage remain. The following questions deserve attention: What is the share of electricity costs from data centers in the overall cost structure of a given AI product? How can electricity costs be passed through the AI value chain, i.e. from data center operator to AI company, to end consumer? To answer the latter question, a better understanding of the AI value chain and industry structure is needed: How high are entry barriers and levels of competition in the AI industry? To what extent do cloud service providers and hyperscalers offer locational pricing based on geographic position of costumer, allowing them to pass-through costs of electricity prices? What drives costumers in choosing differently located colocation or cloud services?

- **Relevance of carbon pricing as compared to other climate policy instruments in driving carbon leakage:** This study has focused on carbon pricing instruments, i.e. ETS, as driver of carbon leakage. However, other climate policies such as renewable energy policy or efficiency mandates may also impose costs on data center operators and users, leading to intentional shifts in compute loads or investments. More research is needed on the role these other climate policies have in driving carbon leakage in comparison to carbon pricing.

- **Other factors driving AI-related data center operations and investments**: Even if carbon pricing led to higher electricity prices that could not be passed through the value chain, the question remains how high electricity prices rank as compared to other factors such as corporate strategy, other regulatory requirements or geopolitical concerns: What is the role of carbon pricing – or even climate policy more generally – in driving data center localization as compared to other factors?

7.3 Policy recommendations

AI-driven data center growth presents significant challenges for global climate mitigation (ITU and World Bank, 2024). Next to mitigating the risks of potential carbon leakage and emissions from imported AI-driven data services, two general key areas need to be addressed: ensuring data center growth is powered by clean electricity and addressing the secondary effects of data center expansion on the availability of clean electricity for the decarbonization in other sectors.

- **Enhancing transparency on environmental impacts of AI**: There are currently only rough estimates regarding the environmental impacts of AI, and more precise data is needed to accurately assess these impacts as well as their geographic distribution. While the AI Act is expected to provide more detailed information about energy consumption during AI training, comprehensive information other environmental impacts (e.g., water consumption) remains lacking.

- **Mitigating potential carbon leakage and imported emissions from AI-driven data center operation and investments:** The risk of carbon leakage arises from disparities in carbon pricing across jurisdictions, which could incentivize shifting data center operations and investments to regions with weaker climate policies. To address this potential risk, a more widespread adoption of carbon pricing mechanisms, such as emissions trading systems or carbon taxes, should be encouraged. International agreements or climate clubs such as initiated by the G7 should harmonize carbon pricing levels to reduce discrepancies that undermine climate goals. Additionally, establishing international standards for clean data center operations can enhance transparency and accountability for emissions.

- **Carbon Border Adjustment Mechanisms (CBAM)** such as introduced by the EU could serve as a valuable tool to address emissions associated with imported compute loads by extending its scope to (AI) services. This approach would effectively subject imported AI-driven compute loads to carbon pricing, ensuring their costs align with those of domestically produced services governed by the EU's stringent climate policies. Such measures could help level the playing field and incentivize service providers outside the EU to adopt renewable electricity to remain competitive. However, implementing and enforcing this policy would present significant challenges, particularly in monitoring and verifying emissions from imported services. The fungible and immaterial character of AI products would make monitoring difficult. For example, trained AI models can be copied, which poses questions of how to account for embodied emissions in the copied versions (an AI-adjusted polluter-pays principle may solve the problem). Furthermore, the impact of such a carbon tax is unlikely to offset the persistent electricity price gap with major data trade partners like the US, where electricity costs are 2–3 times lower, meaning data processing is likely to remain cheaper in the US despite any potential CBAM adjustments.

- **Ensuring data center growth is based on clean electricity:** To align data center growth with climate goals, governments must implement policies that prioritize the use of renewable electricity. This could include providing regulatory requirements, such as mandatory use of (additional) renewable electricity in data centers. Mandatory regulations requiring data center operators to procure a significant share of electricity from renewable sources, supported by guarantees of origin, are another critical step. Efforts such as the data center efficiency regulation by the EU could be more ambitious, as the industry is often already surpassing regulatory requirements. Finally, infrastructure investments, including grid upgrades and energy storage, are essential to ensure sufficient clean energy capacity to meet the growing demand from data centers.

- **Addressing second-order effects of data center growth:** Data center expansion can inadvertently slow the decarbonization of other sectors, such as transport and industry, by competing for available clean electricity. To mitigate these second order effects, integrated energy planning is necessary. Policymakers should coordinate renewable energy expansion with the expected demand growth from data centers, transport, and industry. This approach ensures that electrification efforts in sectors with significant emissions reduction potential are not delayed or compromised by competition for (temporarily) limited clean energy resources.

8 List of references

Aatola, P., Ollikainen, M. and Toppinen, A. (2013) 'Impact of the carbon price on the integrating European electricity market', *Energy Policy*, 61, pp. 1236–1251. Available at: https://doi.org/10.1016/j.enpol.2013.06.036.

Alibaba (2024) *Environmental, Social, and Governance Report 2024*. Alibaba, p. 199. Available at: https://www.alibabagroup.com/en-US/esg.

Apergis, N. (2018) 'Electricity and carbon prices: Asymmetric pass-through evidence from New Zealand', *Energy Sources, Part B: Economics, Planning, and Policy*, 13(4), pp. 251–255. Available at: https://doi.org/10.1080/15567249.2014.1004002.

Apple (2024) *Environmental Progress Report*. Apple, p. 113. Available at: https://www.apple.com/environment/pdf/Apple_Environmental_Progress_Report_2024.pdf.

Axenbeck, J., Berner, A. and Kneib, T. (2024) 'What drives the relationship between digitalization and energy demand? Exploring heterogeneity in German manufacturing firms', *Journal of Environmental Management*, 369, p. 122317. Available at: https://doi.org/10.1016/j.jenvman.2024.122317.

Baidu (2023) *Environmental, Social, and Governance Report 2023*. Baidu, p. 55. Available at: https://esg.baidu.com/Uploads/File/2024/05/17/Baidu%202023%20Environmental,%20Social%20and%20Governance%20Report.20240517150706.pdf.

Ben-Nun, T. and Hoefler, T. (2019) 'Demystifying Parallel and Distributed Deep Learning: An In-depth Concurrency Analysis', *ACM Comput. Surv.*, 52(4), p. 65:1-65:43. Available at: https://doi.org/10.1145/3320060.

Bieser, J. C. T., Hintemann, R., Hilty, L. M., & Beucker, S. (2023) 'A review of assessments of the greenhouse gas footprint and abatement potential of information and communication technology', *Environmental Impact Assessment Review*, 99, p. 107033. Available at: https://doi.org/10.1016/j.eiar.2022.107033.

Bremer, C., Kamiya, G., Bergmark, P., Coroamă, V. C., Masanet, E., Lifset, R. (2023) *Assessing Energy and Climate Effects of Digitalization: Methodological Challenges and Key Recommendations*. Research Coordination Network on the Digital Economy and the Environment. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4459526.

Brynjolfsson, E. and McAfee, A. (2017) 'The Business of Artificial Intelligence: What it can — and cannot — do for your organization', *Harvard Business Review*, 18 July. Available at: https://hbr.org/2017/07/the-business-of-artificial-intelligence (Accessed: 21 February 2024).

Central Statistics Office, Ireland (2023) *Key Findings Data Centres Metered Electricity Consumption 2022*. Available at: https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2022/keyfindings/ (Accessed: 21 July 2024).

Renda, A., Arroyo, J., Fanni, R., Laurer, M., Sipiczki, A., Yeung, T., Maridis, G., Fernandes, M., Endrodi, G., Milio, S., Devenyi, V., Georgiev, S., de Pierrefeu, G. (2021) *Study to support an impact assessment of regulatory requirements for Artificial Intelligence in Europe: final report*. Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2759/523404 (Accessed: 26 September 2024).

China Mobile (2023) *2023 Sustainbility Report*. China Mobile, p. 118. Available at: https://www.chinamobileltd.com/en/ir/reports/ar2023/sd2023.pdf.

Chindata (2023) *2022 Environmental, Social, and Governance Report*. Chindata, p. 136. Available at: https://investor.chindatagroup.com/static-files/7dc29e94-bad9-4951-bfc3-eb9e5e27644c.

CIPL (2023) *The 'Real Life Harms' of Data Localization Policies*. Centre for Information Policy Leadership. Available at: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipltls_discussion_paper_paper_i_-_the_real_life_harms_of_data_localization_policies.pdf (Accessed: 18 October 2024).

Coroamă, V. C., Bergmark, P., Höjer, M., Malmodin, J. (2020) 'A Methodology for Assessing the Environmental Effects Induced by ICT Services: Part I: Single Services', in *Proceedings of the 7th International Conference on ICT for Sustainability*. *ICT4S2020: 7th International Conference on ICT for Sustainability*, New York, NY, USA,, pp. 36–45. Available at: https://doi.org/10.1145/3401335.3401716.

Coroamă, V.C. (2021) *Investigating the Inconsistencies among Energy and Energy Intensity Estimates of the Internet – Metrics and Harmonising Values*. 67656. Swiss Federal Office of Energy SFOE, p. 42. Available at: https://www.aramis.admin.ch/Default?DocumentID=67656.

Coroamă, V.C., Hilty, L.M. and Birtel, M. (2012) 'Effects of Internet-based multiple-site conferences on greenhouse gas emissions', *Telematics and Informatics*, 29(4), pp. 362–374. Available at: https://doi.org/10.1016/j.tele.2011.11.006.

Cory, N. and Dascoli, L. (2021) *How Barriers to Cross-Border Data Flows Are Spreading Globally, What They Cost, and How to Address Them*. Information technology & innovation foundation.

Cottier, B., Rahman, R., Fattorini, L., Maslej, N., Owen, D. (2024) 'The rising costs of training frontier AI models'. arXiv. Available at: https://doi.org/10.48550/arXiv.2405.21015.

Cotton, D. and De Mello, L. (2014) 'Econometric analysis of Australian emissions markets and electricity prices', *Energy Policy*, 74, pp. 475–485. Available at: https://doi.org/10.1016/j.enpol.2014.07.024.

CyrusOne (2024) *2024 Sustainbility Report*. CyrusOne, p. 106. Available at: https://documents.cyrusone.com/wp-content/uploads/2024/06/CyrusOne-2024-Sustainability-Report.pdf.

Devereux, M. P., & Griffith, R. (1998). Taxes and the location of production: evidence from a panel of US multinationals. Journal of Public Economics, 68(3), 335–367. doi: 10.1016/S0047-2727(98)00014-0.

De Vries, A. (2023) 'The growing energy footprint of artificial intelligence', *Joule*, 7(10), pp. 2191–2194. Available at: https://doi.org/10.1016/j.joule.2023.09.004.

De-Arteaga, M., Herlands, W., Neill, D. B., & Dubrawski, A. (2018) 'Machine Learning for the Developing World', *ACM Transactions on Management Information Systems*, 9(2), p. 9:1-9:14. Available at: https://doi.org/10.1145/3210548.

Digital Realty (2024) *2023 Environmental, Social and Governance Report*. Digital Realty, p. 58. Available at: https://go2.digitalrealty.com/rs/087-YZJ-646/images/Report_Digital_Realty_2406_2023_ESG_Report.pdf.

Directorate-General for Climate Action (2015) *Ex-post investigation of cost pass-through in the EU ETS: an analysis for six sectors.* LU: Publications Office. Available at: https://data.europa.eu/doi/10.2834/612494 (Accessed: 26 September 2024).

Duan, J., Zhang, S., Wang, Z., Jiang, L., Qu, W., Hu, Q., Wang, G., Weng, Q., Yan, H., Zhang, X., Qiu, X., Lin, D., Wen, Y., Jin, X., Zhang, T., Sun, P. (2024) 'Efficient Training of Large Language Models on Distributed Infrastructures: A Survey'. arXiv. Available at: https://doi.org/10.48550/arXiv.2407.20018.

Ebert, K., Alder, N., Herbrich, R., Hacker, P. (2024) *AI, Climate, and Regulation: From Data Centers to the AI Act*. arXiv:2410.06681v1. Available at: https://arxiv.org/html/2410.06681v1 (Accessed: 12 November 2024).

Ember (2024) *European Electricity Review 2024*. Brussels: Ember. Available at: https://ember-climate.org/insights/research/european-electricity-review-2024/ (Accessed: 15 October 2024).

Equinix (2024) *Sustainbility Report FY2023*. Equinix, p. 26. Available at: https://sustainability.equinix.com/wp-content/uploads/2024/07/Equinix-Inc_2023-Sustainability-Report.pdf.

Ernst & Young (2024) *Independent Accountants' Review Report*, p. 3. Available at: https://sustainability.aboutamazon.com/2023-renewable-energy-assurance.pdf.

European Commission (2021) *Carbon leakage*. Available at: https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/free-allocation/carbon-leakage_en (Accessed: 20 November 2024).

European Commission (2024) *Report on energy prices and costs in Europe*. Brussels: European Commission. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A52024DC0136 (Accessed: 30 September 2024).

Fan, F. (2021) 'Green data centers in focus', *China Daily*. Available at: https://www.chinadaily.com.cn/a/202112/09/WS61b13913a310cdd39bc7a2ee.html (Accessed: 28 May 2023).

Freitas, C.J.P. and Silva, P.P. da (2015) 'European Union emissions trading scheme impact on the Spanish electricity price during phase II and phase III implementation', *Utilities Policy*, 33, pp. 54–62. Available at: https://doi.org/10.1016/j.jup.2015.01.004.

Gao, P. *et al.* (2018) 'Low latency RNN inference with cellular batching', in *Proceedings of the Thirteenth EuroSys Conference*. New York, NY, USA: Association for Computing Machinery (EuroSys '18), pp. 1–15. Available at: https://doi.org/10.1145/3190508.3190541.

GDS (2023) *2023 Environmental, Social, and Governance (ESG) Report*. GDS, p. 82. Available at: https://c.gds-services.com/esg2023/docs/2023_ESG_Report_EN.pdf.

Goldman Sachs (2024) *AI is poised to drive 160% increase in data center power demand*, *Goldman Sachs*. Available at: https://www.goldmansachs.com/intelligence/pages/AI-poised-to-drive-160-increase-in-power-demand.html (Accessed: 27 May 2024).

Google (2024) *Environmental Report 2024*, *Google*. Available at: https://blog.google/outreach-initiatives/sustainability/2024-environmental-report/ (Accessed: 21 July 2024).

Grubb, M. *et al.* (2022) 'Carbon Leakage, Consumption, and Trade', *Annual Review of Environment and Resources*, 47(Volume 47, 2022), pp. 753–795. Available at: https://doi.org/10.1146/annurev-environ-120820-053625.

Hager, G. D., Drobnis, A., Fang, F., Ghani, R., Greenwald, A., Lyons, T., Parkes, D. C., Schultz, J., Saria, S., Smith, S. F., Tambe, M. (2019) *Artificial Intelligence for Social Good*. arXiv:1901.05406. arXiv. Available at: https://doi.org/10.48550/arXiv.1901.05406.

Hintemann, R., Hinterholzer, S. and Seibel, H. (2023) *Rechenzentren in Deutschland: Aktuelle Marktentwicklungen (Update 2023)*. Available at: https://www.borderstep.de/publikation/hintemann-r-hinterholzer-s-seibel-h-2023-rechenzentren-in-deutschland-aktuelle-marktentwicklungen-update-2023-berlin-bitkom/ (Accessed: 21 July 2024).

Huawei (2023) *Sustainability Addendum to Huawei 2023 Annual Report*. Huawei, p. 26. Available at: https://www-file.huawei.com//media/corp2020/pdf/sustainability/huawei_sustainability_addendum_en_240828.pdf?la=en.

IEA (2023) *Data Centres and Data Transmission Networks*, *IEA*. Available at: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks (Accessed: 15 January 2024).

IEA (2024a) *Electricity 2024 - Analysis and forecast to 2026*. Paris: International Energy Agency. Available at: https://www.iea.org/reports/electricity-2024

IEA (2024b) *World Energy Outlook 2023*. Paris: International Energy Agency. Available at: https://www.iea.org/reports/world-energy-outlook-2024 (Accessed: 4 February 2025).

ITU and World Bank (2024) *Measuring the Emissions and Energy Footprint of the ICT Sector*. © World Bank and International Telecommunication Union. Available at:

http://documents.worldbank.org/curated/en/099121223165540890/P17859712a98880541a4b71d57876048abb (Accessed: 4 February 2025).

Jabłońska, M., Viljainen, S., Partanen, J., Kauranne, T. (2012) 'The impact of emissions trading on electricity spot market price behavior', *International Journal of Energy Sector Management*, 6(3), pp. 343–364. Available at: https://doi.org/10.1108/17506221211259664.

Joltreau, E. and Sommerfeld, K. (2019) 'Why does emissions trading under the EU Emissions Trading System (ETS) not affect firms' competitiveness? Empirical findings from the literature', *Climate Policy*, 19(4), pp. 453–471. Available at: https://doi.org/10.1080/14693062.2018.1502145.

Jouvet, P.-A. and Solier, B. (2013) 'An overview of CO2 cost pass-through to electricity prices in Europe', *Energy Policy*, 61(C), pp. 1370–1376. Available at: https://doi.org/10.1016/j.enpol.2013.05.090.

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., Rolnick, D. (2022) 'Aligning artificial intelligence with climate change mitigation', *Nature Climate Change*, 12(6), pp. 518–527. Available at: https://doi.org/10.1038/s41558-022-01377-7.

Kamiya, G. and Bertoldi, P. (2024) *Energy Consumption in Data Centres and Broadband Communication Networks in the EU*, *JRC Publications Repository*. Available at: https://publications.jrc.ec.europa.eu/repository/handle/JRC135926 (Accessed: 31 March 2024).

KDDI (2024) *KDDI_ESGdata*. KDDI, p. 16. Available at: https://www.kddi.com/extlib/corporate/sustainability/report/esg-data/pdf/KDDI_ESGdata.pdf.

Kosch, M., Blech, K. and Abrell, J. (2022) 'Rising Electricity Prices in Europe: The Impact of Fuel and Carbon Prices'. Rochester, NY. Available at: https://doi.org/10.2139/ssrn.4259209.

Lacoste, A., Luccioni, A., Schmidt, V., Dandres, T. (2019) 'Quantifying the Carbon Emissions of Machine Learning'. arXiv. Available at: https://doi.org/10.48550/arXiv.1910.09700.

Luccioni, A.S., Viguier, S. and Ligozat, A.-L. (2024) 'Estimating the carbon footprint of BLOOM, a 176B parameter language model', *Journal of Machine Learning Research*, 24(1), pp. 11990–12004. Available at: https://doi.org/10.48550/arXiv.2211.02001.

Luccioni, S., Jernite, Y. and Strubell, E. (2024) 'Power Hungry Processing: Watts Driving the Cost of AI Deployment?', in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY, USA: Association for Computing Machinery (FAccT '24), pp. 85–99. Available at: https://doi.org/10.1145/3630106.3658542.

Malmodin, J., Lövehagen, N., Bergmark, P., Lundén, D. (2024) 'ICT sector electricity consumption and greenhouse gas emissions – 2020 outcome', *Telecommunications Policy*, p. 102701. Available at: https://doi.org/10.1016/j.telpol.2023.102701.

McKinsey (2024) *The state of AI in early 2024 | McKinsey*. Washington: McKinsey Consulting. Available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/ (Accessed: 15 October 2024).

Meta (2024) *For a better reality 2024 Sustainbility Report*. Meta, p. 94. Available at: https://sustainability.atmeta.com/wp-content/uploads/2024/08/Meta-2024-Sustainability-Report.pdf.

Microsoft (2024a) *How can we advance sustainability? 2024 Environmental Sustainability Report*. Microsoft, p. 88. Available at: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW1lMjE.

Microsoft (2024b) *Data residency in Azure*. Microsoft, p. 24. Available at: https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/1_data-residency-in-azure-2024-06-20.pdf.

Morgan Stanley (2024) *Powering the AI Revolution*. Available at: https://www.morganstanley.com/ideas/ai-energy-demand-infrastructure (Accessed: 6 May 2024).

Mosquera-López, S. and Nursimulu, A. (2019) 'Drivers of electricity price dynamics: Comparative analysis of spot and futures markets', *Energy Policy*, 126, pp. 76–87. Available at: https://doi.org/10.1016/j.enpol.2018.11.020.

Ni, W., Hu, X., Du, H., Kang, Y., Ju, Y., Wang, Q. (2024) 'CO2 emission-mitigation pathways for China's data centers', *Resources, Conservation and Recycling*, 202, p. 107383. Available at: https://doi.org/10.1016/j.resconrec.2023.107383.

NTT DATA (2023) *Sustainbility Report 2023*. NTT DATA, p. 126. Available at: https://www.nttdata.com/global/en/-/media/nttdataglobal/1_files/sustainability/susatainability-report/2023/sr2023.pdf?rev=75c606edf94d412d9f853c8d7beeb3e8.

OECD (2022a) *Artificial Intelligence and international trade*. Paris: OECD. Available at: https://www.oecd.org/en/publications/artificial-intelligence-and-international-trade_13212d3e-en.html (Accessed: 11 November 2024).

OECD (2022b) *Measuring the value of data and data flows*. Paris: OECD. Available at: https://www.oecd.org/en/publications/measuring-the-value-of-data-and-data-flows_923230a6-en.html (Accessed: 11 November 2024).

OECD (2022c) *The Evolving Concept of Market Power in the Digital Economy*. Paris: OECD. Available at: https://one.oecd.org/document/DAF/COMP/WD(2022)34/en/pdf (Accessed: 26 September 2024).

OECD (2024) *Carbon prices, emissions and international trade in sectors at risk of carbon leakage: Evidence from 140 countries*. OECD Economics Department Working Papers 1813. Available at: https://doi.org/10.1787/116248f5-en.

Pahle, M., Sitarz, J., Osorio, S. *et al.* (2022) *The EU-ETS price through 2030 and beyond: A closer look at drivers, models and assumptions The EU-ETS price through 2030 and beyond: A closer look at drivers, models and assumptions - Input material and takeaways from a workshop in Brussels*. Documentation Kopernikus Projekt Ariadne. Berlin.

Palmer, K., Paul, A. and Keyes, A. (2018) 'Changing baselines, shifting margins: How predicted impacts of pricing carbon in the electricity sector have evolved over time', *Energy Economics*, 73, pp. 371–379. Available at: https://doi.org/10.1016/j.eneco.2018.03.023.

Patterson, D., Gonzalez, J., Hölzle, U., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D. R., Texier, M., Dean, J. (2022) 'The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink', *Computer*, 55(7), pp. 18–28. Available at: https://doi.org/10.1109/MC.2022.3148714.

Pietzcker, R.C., Osorio, S. and Rodrigues, R. (2021) 'Tightening EU ETS targets in line with the European Green Deal: Impacts on the decarbonization of the EU power sector', *Applied Energy*, 293, p. 116914. Available at: https://doi.org/10.1016/j.apenergy.2021.116914.

QTS (2023) *2023 Sustainbility Report*. QTS, p. 61. Available at: https://mma.prnewswire.com/media/2519339/QTS_Sustainability_Report_2023_FINAL_compressed.pdf?p=pdf.

Ritz, R. A. (2024). Does competition increase pass-through? RAND Journal of Economics, 55(1), 140–165. doi: 10.1111/1756-2171.12461.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J. C., Creutzig, F. (2022) 'Tackling Climate Change with Machine Learning', *ACM Computing Surveys*, 55(2), p. 42:1-42:96. Available at: https://doi.org/10.1145/3485128.

Rostelecom (2022) *Sustainbility Report 2022*. Rostelecom, p. 129. Available at: https://www.company.rt.ru/en/social/report/Rostelecom_ESG_report_2022_eng.pdf.

Sato, M., Neuhoff, K., Graichen, V., Schumacher, K., Matthes, F. 2015) 'Sectors Under Scrutiny: Evaluation of Indicators to Assess the Risk of Carbon Leakage in the UK and Germany', *Environmental and Resource Economics*, 60(1), pp. 99–124. Available at: https://doi.org/10.1007/s10640-014-9759-y.

Saussay, A. and Sato, M. (2024) 'The impact of energy prices on industrial investment location: Evidence from global firm level data', *Journal of Environmental Economics and Management*, 127, p. 102992. Available at: https://doi.org/10.1016/j.jeem.2024.102992.

Schneider Electric (2023) *The AI Disruption: Challenges and Guidance for Data Center Design*. Available at: https://www.se.com/ww/en/download/document/SPD_WP110_EN/?ssr=true.

Schwartz, R., Dodge, J., Smith, N. A., Etzioni, O. (2020) 'Green AI', *Communications of the ACM* [Preprint]. Available at: https://doi.org/10.1145/3381831.

Selby, J. (2017) 'Data localization laws: trade barriers or legitimate responses to cybersecurity risks, or both?', *International Journal of Law and Information Technology*, 25(3), pp. 213–232. Available at: https://doi.org/10.1093/ijlit/eax010.

Semianalysis (2024) *AI Datacenter Energy Dilemma - Race for AI Datacenter Space*. Available at: https://www.semianalysis.com/p/ai-datacenter-energy-dilemma-race (Accessed: 6 May 2024).

Sergeev, A. and Balso, M.D. (2018) 'Horovod: fast and easy distributed deep learning in TensorFlow'. arXiv. Available at: https://doi.org/10.48550/arXiv.1802.05799.

Soares, I. V., Yarime, M., & Klemun, M. (2024). Balancing the trade-off between data center development and its environmental impacts: A comparative analysis of Data Center Policymaking in Singapore, Netherlands, Ireland, Germany, USA, and the UK. Environmental Science & Policy, 157, 103769. doi: 10.1016/j.envsci.2024.103769.

Statistics Netherlands (2022) *Elektriciteit geleverd aan datacenters, 2017-2021*, *Centraal Bureau voor de Statistiek*. Available at: https://www.cbs.nl/nl-nl/maatwerk/2022/49/elektriciteit-geleverd-aan-datacenters-2017-2021 (Accessed: 21 July 2024).

Steinebach, Y., Hinterleitner, M., Knill, C., Fernández-i-Marín, X. (2024) 'A review of national climate policies via existing databases', *npj Climate Action*, 3(1), pp. 1–9. Available at: https://doi.org/10.1038/s44168-024-00160-y.

Strubell, E., Ganesh, A. and McCallum, A. (2019) 'Energy and Policy Considerations for Deep Learning in NLP'. arXiv. Available at: https://doi.org/10.48550/arXiv.1906.02243.

Strubell, E., Ganesh, A. and McCallum, A. (2020) 'Energy and Policy Considerations for Modern Deep Learning Research', *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(09), pp. 13693–13696. Available at: https://doi.org/10.1609/aaai.v34i09.7123.

Swedish Energy Agency (2023) *Energianvändning i datacenter och digitala system*. Swedish Energy Agency. Available at: https://www.ri.se/en/the-status-of-data-center-and-crypto-mining-energy-use-in-sweden.

Tencent (2023) *Environmental, Social, and Governance Report 2023*. Tencent, p. 116. Available at: https://static.www.tencent.com/uploads/2024/04/08/a041eba55b96c7952e26180a2c7cdd28.pdf.

Görlach, B., Duwe, M., Germeshausen, R., Ostwald, R., Riedel, A., Velten, E. K., Voigt, S., Voß, P., Wölfing, N., Zelljadt, E. (2020) *Analysen zum direkten und indirekten Carbon Leakage-Risiko europäischer Industrieunternehmen*. Berlin: Umweltbundesamt. Available at: https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2020_10_20_climate_change_32_2020_analysen_carbon-leakage-risiko.pdf (Accessed: 26 September 2024).

Vantage (2023) *2023 Esg Report*. Vantage, p. 53. Available at: https://vantage-dc.com/wp-content/uploads/2024/07/2023-ESG-Report_Vantage-Data-Centers.pdf.

Verde, S.F. (2020) ‘The Impact of the Eu Emissions Trading System on Competitiveness and Carbon Leakage: The Econometric Evidence’, *Journal of Economic Surveys*, 34(2), pp. 320–343. Available at: https://doi.org/10.1111/joes.12356.

Verdecchia, R., Sallou, J. and Cruz, L. (2023) ‘A systematic review of Green AI’, *WIREs Data Mining and Knowledge Discovery*, 13(4), p. e1507. Available at: https://doi.org/10.1002/widm.1507.

VNET (2023) *2023 Esg Report*. VNET, p. 76. Available at: https://www.vnet.com/upload/portal/PDF/2023esgreport.pdf.

Wang, M., & Kuusi, T. (2024). Trade flows, carbon leakage, and the EU Emissions Trading System. Energy Economics, 134, 107556. doi: 10.1016/j.eneco.2024.107556.

Wolff, G. and Feuerriegel, S. (2019) ‘Emissions Trading System of the European Union: Emission Allowances and EPEX Electricity Prices in Phase III’, *Energies*, 12(15), p. 2894. Available at: https://doi.org/10.3390/en12152894.

Woo, C.K. & Olson, A. & Chen, Y. & Moore, J. & Schlag, N. & Ong, A. & Ho, T., 2017. "Does California's CO2 price affect wholesale electricity prices in the Western U.S.A.?" Energy Policy, Elsevier, vol. 110(C), pages 9-19. DOI: 10.1016/j.enpol.2017.07.059WTO (2023) *Handbook on Measuring Digital Trade: Second edition*. Washington, D.C: World Trade Organizations in collaboration with others. Available at: https://doi.org/10.5089/9789287073600.071.

Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Behram, F. A., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H.-H. S., Akyildiz, B. (2022) ‘Sustainable AI: Environmental Implications, Challenges and Opportunities’. arXiv. Available at: https://doi.org/10.48550/arXiv.2111.00364.

Ye, R., Wang, W., Chai, J., Li, D., Li, Z., Xu, Y., Du, Y., Wang, Y., Chen, S. (2024) ‘OpenFedLLM: Training Large Language Models on Decentralized Private Data via Federated Learning’. arXiv. Available at: https://doi.org/10.48550/arXiv.2402.06954.

Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., Gao, Y. (2021) ‘A survey on federated learning’, Knowledge-Based Systems, 216, p. 106775. Available at: https://doi.org/10.1016/j.knosys.2021.106775.

A Appendix: Expert interviews

Table 8: Expert interviews for this study.

Expert	Organization	Role or qualification
Jan Abrell	University of Basel	Environmental economist
Carlos Alves	Digital Realty	DC capacity & energy manager
Johannes Leon Kirnberger	OECD	AI & sustainability expert
Dieter Kranzlmüller	Leibniz-Rechenzentrum (LRZ)	Director

Source: own overview, INFRAS and Roegen Centre for Sustainability.

B Appendix: Important AI data center developments announced in 2024

While we cannot provide strong evidence in support of this assumption, this year's announcements on the development of new AI data centers do correlate with the general DC focus on the US, China, Europe, and some countries in the Middle East and Asia.

Data on the Chinese development plans are hard to come by, but it can be safely assumed that accelerated AI development will also take place in China. Google- and Gemini-supported searches on AI data center development announced in 2024 by the 22 companies analyzed here include:

- US: Large US hyperscalers, which are so power-hungry for their new AI-oriented DCs that within a few months, they all announced backing up the refurbishment or new development of nuclear reactors, Microsoft supporting the restart of Three Mile Island,[34] while Google and Amazon support the development of new small modular reactors.[35] Oracle has announced a US-based AI data center with a capacity of 800 MW.[36] And the already sprawling DC ecosystem in Virginia might soon double in size.[37]
- Europe: Microsoft alone plans to invest more than 3 billion € in two new AI data centers in Germany, [38] each of which will be in the 100-200 MW range (which is very large for European standards), while the same North Rhine-Westphalia region expects by 2026 further colocation DCs in the hundreds of MWs.[39] Around Frankfurt and Berlin, further hundreds of MWs of DC capacity are already being built, with further yet in planning.[40] Meanwhile, QTS plans the development of 300 MW in Spain and 1.1 GW in the UK.[41]
- Middle East: Both the UAE and Saudi Arabia invest billions, perhaps even dozens of billions, in the development of data center capacity, mainly aimed at AI.[42]

[34] Martucci, B. (20 September 2024). Constellation plans 2028 restart of Three Mile Island unit 1, spurred by Microsoft PPA. See https://www.utilitydive.com/news/constellation-three-mile-island-nuclear-power-plant-microsoft-data-center-ppa/727652/ (Accessed: 2 December 2024).

[35] See Terrell, M. (14 October 2024). New nuclear clean energy agreement with Kairos Power. https://blog.google/outreach-initiatives/sustainability/google-kairos-power-nuclear-energy-agreement/ and Amazon (16, October 2024). Amazon signs agreements for innovative nuclear energy projects to address growing energy demands https://www.aboutamazon.com/news/sustainability/amazon-nuclear-small-modular-reactor-net-carbon-zero, respectively (Accessed: 2 December 2024).

[36] See Butler, G. (10 September 2024). Oracle to build nuclear SMR-powered gigawatt data center. https://www.datacenterdynamics.com/en/news/oracle-to-build-nuclear-smr-powered-gigawatt-data-center/ (Accessed: 2 December 2024).

[37] See Schlotterback V., Pasqualichio J., and Bond, J. (18 September 2024). The Data Center Balancing Act: Powering Sustainable AI Growth. https://www.brownadvisory.com/us/insights/data-center-balancing-act-powering-sustainable-ai-growth (Accessed: 2 December 2024).

[38] See Butler, G. (16 February 2024). Microsoft to invest €3.2bn in doubling AI infrastructure and cloud capacity in Germany. https://www.datacenterdynamics.com/en/news/microsoft-to-invest-32bn-in-doubling-ai-infrastructure-and-cloud-capacity-in-germany/ (Accessed: 2 December 2024).

[39] See Thomeczek, H. (19. Februar 2024) Microsoft baut Rechenzentren in NRW und sucht Mietflächen in Frankfurt. https://www.iz.de/projekte/news/-microsoft-baut-rechenzentren-in-nrw-und-sucht-mietflaechen-in-frankfurt-2000023361 (Accessed: 2 December 2024).

[40] See Williams, M. (23 September 2024). German Data Center Market. https://www.jll.de/en/trends-and-insights/investor/german-data-center-market-addressing-rising-demand (Accessed: 2 December 2024).

[41] See Swinhoe, D. (15 October 2024). Blackstone to develop 300MW data center campus in Aragon, Spain. https://www.datacenterdynamics.com/en/news/blackstone-to-develop-300mw-data-center-campus-in-aragon-spain/ and Swinhoe, D. (3 September 2024). QTS files to build 1.1GW data center campus in Northumberland, UK. https://www.datacenterdynamics.com/en/news/qts-files-to-build-11gw-data-center-campus-in-northumberland-uk/ (Accessed: 2 December 2024), respectively.

[42] See Martini, E. P. (30 July 2024). Saudi Arabia and UAE's race for AI, data center dominance https://www.diplomaticourier.com/posts/saudi-arabia-and-uaes-race-for-ai-data-center-dominance (Accessed: 2 December 2024).

C Appendix: Assessing the global distribution of general DC energy consumption

C.1 Methodology

Data on the exact number, placement, and energy consumption of general-purpose DCs is itself patchy (Bremer *et al.*, 2023). There is, however, plenty of information that can be leveraged to achieve country-wide DC energy consumption estimates. Two main possibilities exist: i) using national statistics on DC energy consumption, and ii) company reports of hyperscale and colocation operators in conjunction with the global distribution of their DCs:

- More straightforward are country-wide estimates. Unfortunately, there are few estimates available at the country level. The most credible are based on reported electricity consumption and metered data – e.g. for Ireland (Central Statistics Office, Ireland, 2023) or the Netherlands (Statistics Netherlands, 2022). Estimates based on available reports, expert interviews and other statistical data are also quite confident, as in Sweden (Swedish Energy Agency, 2023) or on bottom-up estimates with robust models developed over many years, as is the case in Germany (Hintemann, Hinterholzer and Seibel, 2023). For other countries, such data is more difficult to find and of lower quality – this is particularly true for China (Fan, 2021; Ni *et al.*, 2024) but also for the USA when it comes to very recent data.
- For a company-based assessment, data from the US "big four" (Google, Microsoft, Amazon, Meta) but also from important Chinese operators (such as Chinese telecom operators, Alibaba, and Baidu), further important players from all continents as well as content distribution networks (CDN) are required. In particular the "big four", but also Chinese and some of the other AI actors, have an international presence. It does thus not suffice to assume that all company-owned DCs are located in (and thus the entire energy consumption takes place in) their country of residence; complementary data on the geographic distribution of their DCs and subsequent geographic aggregation are also required.

The analysis needs to focus on hyperscale and colocation DCs, as traditional in-house enterprise DCs are less relevant for AI, even though the inference of smaller ML models can also take place here. Hyperscalers such as the US "big four" (Amazon, Google, Microsoft, and Meta), however, together with a few further Chinese and international players and (to a lesser extent) colocation DCs are those who either develop in-house (in the case of Google) or buy the vast majority of AI accelerators on the market to train and run ML models.

Unlike the geographic distribution of cryptocurrency mining equipment, assuming that the large worldwide DC operators are also the large AI operators is thus a reasonable assumption, and one that has been validated in several interviewees as well as by members of the scientific accompanying team. Prof. Kranzlmüller, director of the Leibniz supercomputing centre, said in his interview that for various reasons such as more available power, faster approval processes, and more favorable societal views of large data centers and supercomputing, he expects the trend to intensify and compute capacities of DCs in general and AI specifically to grow mainly in the US as well as Asia (specifically China and Japan).

The large hyperscale and colocation operators used in this study to assess the geographic distribution of general data center energy consumption are presented in the section below.

C.2 Hyperscale and colocation operators used to assess the geographic distribution of general-purpose DC energy consumption

Table 9: Hyperscale and colocation operators used in this study, together with the sources used to extract their 2023 overall energy consumption and the location of their data centers, respectively.

Company	Energy source	DC distribution source
Google	Google (2024)	https://www.google.com/about/datacenters/locations/ (Accessed: 2 December 2024).
Microsoft	Microsoft (2024a)	(Microsoft, 2024b)
Amazon	Reverse engineered from Amazon's location-based Scope 2 emissions of 15.67 Mt CO_2eq./year (Ernst & Young, 2024), using the US average mix of 0.369 kg CO_2/kWh	https://www.datacentermap.com/c/amazon-aws/ (Accessed: 2 December 2024).
Meta	Meta (2024)	https://datacenters.atmeta.com/all-locations/ (Accessed: 2 December 2024).
Apple	Apple (2024)	(Apple, 2024), https://dgtlinfra.com/apple-data-center-locations/ (Accessed: 2 December 2024).
Equinix	Equinix (2024)	https://www.equinix.com/data-centers (Accessed: 2 December 2024).
DigitalRealty	Digital Realty (2024)	https://www.digitalrealty.com/data-centers (Accessed: 2 December 2024).
Chindata	Chindata (2023)	https://www.chindatagroup.com/global.html (Accessed: 2 December 2024).
GDS	GDS (2023)	https://c.gds-services.com/esg2023/docs/2023_ESG_Report_EN.pdf (Accessed: 2 December 2024).
Alibaba	Alibaba (2024)	https://www.alibabacloud.com/en/global-locations?_p_lc=1 (Accessed: 2 December 2024).
CyrusOne	CyrusOne(2024)	https://www.cyrusone.com/data-centers/ (Accessed: 2 December 2024).
NTT Data	NTT DATA (2023)	https://services.global.ntt/en-us/services-and-products/global-data-centers/global-locations (Accessed: 2 December 2024).
QTS	QTS (2023)	https://qtsdatacenters.com/data-centers (Accessed: 2 December 2024).
KDDI	KDDI (2024)	https://www.eu.kddi.com/en/services/datacenter/global-datacenter/ (Accessed: 2 December 2024)
VNET	VNET (2023)	https://www.vnet.com/en/resource.html (Accessed: 2 December 2024)
Chinamobile	China Mobile (2023)	https://www.chinamobileltd.com/en/ir/reports/ar2023/sd2023.pdf (Accessed: 2 December 2024)

Company	Energy source	DC distribution source
Baidu	Baidu (2023)	https://intl.cloud.baidu.com/doc/Reference/s/2jwvz23xx-en (Accessed: 2 December 2024)
Huawei	Huawei (2023)	https://www.huaweicloud.com/intl/en-us/about/global-infrastructure.html (Accessed: 2 December 2024)
Tencent	Tencent (2023)	https://www.tencentcloud.com/global-infrastructure (Accessed: 2 December 2024)
Akamai	https://www.akamaisustainability.com/indicators/	https://techdocs.akamai.com/cloud-computing/docs/how-to-choose-a-data-center (Accessed: 2 December 2024)
Vantage	Vantage (2023)	https://vantage-dc.com/wp-content/uploads/2024/07/2023-ESG-Report_Vantage-Data-Centers.pdf (Accessed: 2 December 2024)
Rostelecom	Rostelecom (2022)	https://baxtel.com/data-centers/rostelecom (Accessed: 2 December 2024)

C.3 Corrections for Germany and France

The results of the analysis yielded 4.6 TWh/year for Germany and 1.6 TWh/year for France. These numbers, however, are most likely underestimates, in particular for France: While variability in their assessment exists, the most likely estimates for the energy consumption of all DCs in these countries are 17.9 TWh/year for Germany (Hintemann, Hinterholzer and Seibel, 2023) and 9 TWh/year for France (Kamiya and Bertoldi, 2024).

These values from the literature refer to all DCs and not just hyperscalers and large colocation providers. These sub-categories, which are the ones relevant to this study as they are indicative of AI energy, currently represent about 50% of all DC energy consumption around the world (Malmodin *et al.*, 2024); in Europe perhaps slightly less (Hintemann, Hinterholzer and Seibel, 2023). Using 45% of the values above yields 8.1 TWh/year for Germany and 4.1 TWh/year for France. Fortunately, the large discrepancy between the initial and corrected values (factor of 1.76 for Germany and 2.56 for France) are not characteristic for the results as a whole. According to both the literature and interviewees, the data center landscape in most of Europe is not characterized by the presence of a few major global DC operators. Instead, it has a greater emphasis on colocation data centers, serving a diverse range of customers, with many more mid-sized operators. This is for a variety of reasons, among which the comparative difficulty to source the quantities of power required by large DCs (one of the interviewees said "I envy the power ratings of hundreds of megawatts they achieve in Texas; at our DC [in Europe], I am afraid we will not be able to grow from 15 to 40 MW), stricter environmental and privacy regulations than elsewhere, and the comparative unavailability and high cost of land.[43]

By contrast, the US market is dominated by hyperscale data centers built by the "big four" (Amazon, Google, Microsoft, and Meta) and other large operators. China's DC landscape is also heavily influenced by domestic hyperscalers such as Alibaba, Tencent, and Baidu, reflecting the country's focus on developing its own digital ecosystem.[44] These are easily caught by the method

[43] See KPMG (2024) Data centres in Europe: A strategic approach. https://kpmg.com/ie/en/home/insights/2024/09/data-centres-in-europe-strategy.html (Accessed: 2 December 2024).

[44] See Triolo P. and Schaefer K. (27 June 2024). China's Generative AI Ecosystem in 2024: Rising Investment and Expectations. https://www.nbr.org/publication/chinas-generative-ai-ecosystem-in-2024-rising-investment-and-expectations/ (Accessed: 2 December 2024).

deployed in this study, while the more numerous but smaller DC operators in Europe are more likely to evade it.

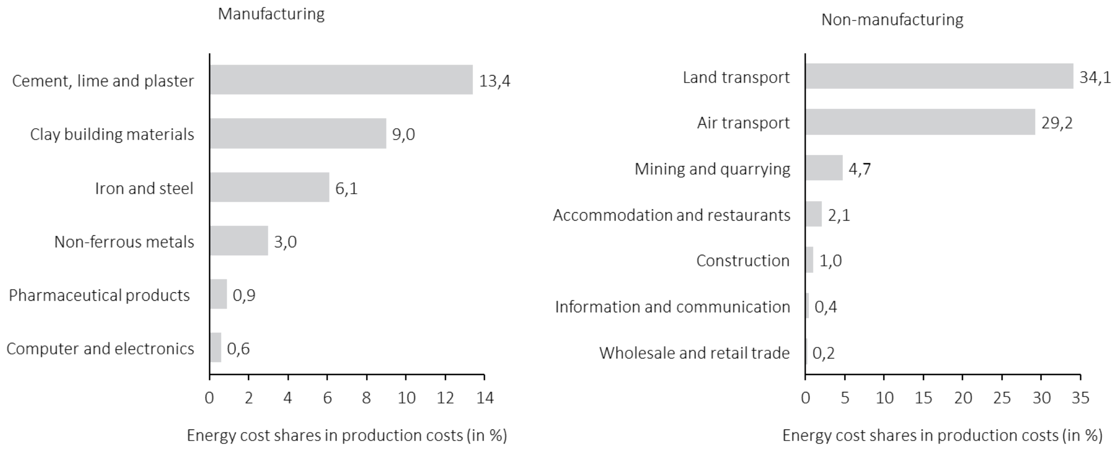
As further proof that the method is overall sound stand the results for Ireland. The country, whose DC landscape more closely resembles that of the US, has attracted significant investment from hyperscalers due to factors like favorable tax policies, reliable power infrastructure, and a supportive government. This has made Ireland a major hub for data centers serving not just the Irish market but also the broader European region. The result according to the methodology deployed here (6.39 TWh in 2023 for Ireland) is very close to the 6.3 TWh reported by the Irish Central Statistics Office.[45]

[45] See Irish Central Statistics Office (23 July 2024). Data Centres Metered Electricity Consumption 2023. https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2023/keyfindings/ (Accessed: 2 December 2024).

D Appendix: Energy intensity of different economic sectors

Figure 11 shows the energy cost shares as a % of production costs in manufacturing and non-manufacturing sectors.

Figure 11: Energy cost shares in production costs of different economic sectors.



Source: own illustration, INFRAS and Roegen Centre for Sustainability. Based on data from (European Commission, 2024).

E Appendix: Policies affecting data center localization

Table 10: Assessment of non-binding public initiatives and private sector initiatives.

Name	Short description	Potential impact on carbon leakage
AI Code of Conduct	The International Code of Conduct for Organizations Developing Advanced AI Systems aims to promote safe, secure, and trustworthy AI worldwide and will provide voluntary guidance for actions by organizations developing the most advanced AI systems, including the most advanced foundation models and generative AI systems.	Non-binding, harmonizes certain principles for I usage between G7 / OECD countries
European Code of Conduct on Data Centre Energy Efficiency	The European Code of Conduct for Data Centres (EU DC CoC) is a voluntary initiative set up by the Joint Research Centre (JRC) in response to the increasing energy consumption in data centers […]. It encourages and guides data center operators and owners in cost-effectively reducing energy consumption without compromising the mission-critical function of these facilities. Since its launch in 2008, more than 500 data centers have joined the EU DC CoC to improve their energy efficiency.	It offers extensive guidelines on all kinds of technical best practices, but makes no reference to specific reduction targets for energy (or GHG emissions). Non-binding. May mitigate incentives for carbon leakage due to lower energy costs.
German AI Strategy and German AI Action Plan	German AI Strategy has the goals of making Germany a leading center for AI, responsible usage of AI and the integration of AI into society in ethical, legal, cultural, and institutional terms. Based on this, the AI Action Plan names concrete measures to achieve the goals in the strategy.	The AI strategy has the main goal of increasing Germany's competitiveness and attractiveness for AI. It shows that Germany (and likely other EU countries) recognize the importance of AI and data centers and want to build an attractive environment for it. Thus, in principle, it should have a preventive impact on carbon leakage. However, it remains open how effective the measures taken will be.
Climate Neutral Data Centre Pact	The focus and span of control of the Pact is on achieving sustainable data center facilities, more specifically to meet commonly accepted goals for energy efficiency, carbon-free or renewable energy, water conservation, circular economy, and heat recovery and reuse.	Similar ambition level as regulation, but already older. It shows that (many) data center operators follow environmental targets with similar or higher ambition than current DC regulations. May mitigate incentives for carbon leakage.
RE100	RE100 is the global corporate renewable energy initiative bringing together hundreds of large and ambitious businesses committed to 100% renewable electricity	Only focus on renewable energy (including certificates and power purchase agreements).. Shows that at least large corporations like Google, Microsoft, Meta or Adobe have taken measures for their data centers since a long time..

Source: own illustration, INFRAS and Roegen Centre for Sustainability.

Table 11: Assessment of EU policies/regulations and their implications for carbon leakage.

Name	Short description	Potential impact on carbon leakage
EU Carbon Border Adjustment Mechanism (CBAM)	CBAM is EU's tool to put a fair price on the carbon emitted during the production of carbon intensive goods that are entering the EU, and to encourage cleaner industrial production in non-EU countries. It complements the EU ETS.	CBAM leads to higher prices to a number of imported goods, including electricity, which could potentially affect the profitability of data centers. However, since consumed electricity is mainly produced in the EU itself, the EU ETS seems much more relevant and the impact of CBAM on data centres seems very low.
Electricity price compensation mechanisms in the EU	The purpose of communication 2020/C 317/04 is to ensure that electricity intensive activities are compensated for high electricity prices to maintain competitiveness and prevent carbon leakage.	In principle, the goal of these EU guidelines is exactly to prevent carbon leakage and to compensate for higher electricity prices through EU ETS (and CBAM). For this purpose, the EU has also defined some (sub-)sectors that are particularly prone to the risk of carbon leakage. However, data centres (or any AI activity) are not part of this and thus are currently not impacted by this measure.

Name	Short description	Potential impact on carbon leakage
EU Energy Efficiency Directive (EED) and country-wise implementations.	The main focus of the EED are the energy savings targets of countries and the EU overall. Countries need to implement concrete targets themselves (incl. Specific measures on data centers)	The impact of this regulation on data centers is manifold: 1. it could disincentivise countries to host data centers, as this might impede their own energy saving targets. However, however, this seems very indirect and speculative. 2. Countries should take measures specifically on data centers, e.g. through energy efficiency targets (as done e.g. by Germany). However, many large IT companies (e.g. Google and Microsoft) are already below the (German) thresholds for the year 2030 as of now. 3. Further regulatory obligations for data centers such as environmental management systems and disclosure obligations including auditing of key performance indicators, leading to additional costs. Overall, impact on carbon leakage seems presumably small.
EU AI Act	These rules establish obligations for providers and users of AI depending on the level of risk. While many AI systems pose minimal risk, they need to be assessed. The AI act differentiates between 3 levels of risks: unacceptable risks, high risks and others.	Most affected by this regulation are AI applications categorized under unacceptable (completely forbidden) or high risk (special approval processes). As such, this regulation has considerable impact on AI companies. However, it affects all parts of the value chain of AI, i.e. the AI services would neither be offered nor processed in the EU. Therefore, carbon leakage is strongly limited to companies developing potentially high-risk applications of AI that could at the same time completely process and offer the final product/service in non-EU countries.
EU General Data Protection Regulation (GDPR)	The GDPR is a regulation on information privacy. It is an important component of EU privacy law and human rights law, in particular Article 8(1) of the Charter of Fundamental Rights of the European Union. It also governs the transfer of personal data outside the EU and EEA.	Affects AI data centers mainly through a number of restrictions regarding data usage. However, similar regulations already in place in other countries and similar problems as with AI act (whole value chain). Furthermore, usage of personal data of EU citizens would even restrict them from processing it outside the EU, which could rather have a preventive effect on carbon leakage, particularly with AI applications using personal and private data.
EU Corporate Sustainability Due Diligence Directive (CSDDD)	The CSDDD lays down rules on: (a) obligations for companies regarding actual and potential human rights adverse impacts […] with respect to their own operations, the operations of their subsidiaries, and the operations carried out by their business partners in the chains of activities of those companies; (b) liability for violations of the obligations as referred to in point (a).	There is no apparent direct link to AI and data centers (at least not more direct than for any other business).

Source: own illustration, INFRAS and Roegen Centre for Sustainability.