

[AI and Tax] Data-Center Taxation and the Geography of AI Value

Keith Lee

Swiss Institute of Artificial Intelligence, Chaltenbodenstrasse 26, 8834 Schindellegi, Schwyz, Switzerland

Abstract

Data centres are increasingly presented as an attractive tax base, given their capital intensity, immobility and local demands for electricity, land, water and other infrastructure. The article argues that the physical location of servers cannot automatically identify the wider economic rents generated by AI intrinsic to the wider AI value chain. These value streams are dispersed across geographies and layers in an international value chain—from semiconductors, hardware and energy, through cloud orchestration, foundation models and application software, to proprietary data and ultimately to adopting entities. These locations may differ substantially from the physical location of the server and are linked through contractual, licensing, transfer-pricing and corporate ownership arrangements which, together, separate compute geography from the location of accrued economic rent and associated profit. The article contends that host jurisdictions do have a legitimate claim to recover for identifiable land, grid, water, environmental and other infrastructure costs, but do not automatically acquire taxing rights over all model, cloud, software and market-based rents. It proceeds to differentiate the locational characteristics of training and inference workloads and to note how mixed workloads weaken the case for a broad facility levy. Concluding that Europe's key regulatory dilemma is to balance legitimate cost-reflective charges with an interest in growth and innovation, the article advocates taxing site-specific infrastructure and environmental burdens locally while addressing wider rents through separate profit-allocation and jurisdictional rules.

1 Introduction - Why Data Centres Are an Attractive but Incomplete Tax Base for AI

The attraction of a city-based and physical tax base for data centers is straightforward. If AI diminishes aspects of the tax base that have traditionally fallen on labor income, its employment, its plant and equipment or its labor-intensive processes, then the tangible, capital-intensive, polluting, geographically fixed over the short to medium term and state-plannable nature of the data center infrastructure seems to offer a recognizable alternative handle-one that is easier to notice, measure and charge than less tangible digital profits. In a digital, and indeed AI, age, where value appears to be drifting partly across borders via codified lines of code in the cloud, the data center tower seems to have *prima facie* appeal as an analog solution. However, that intuition is incomplete. The tangible placement of data center infrastructure can justly underpin certain local taxation and fee-recovery rights, but it does not secure a rightful claim in law or economy as to the entire AI-generated economic rent.^[1]

The trade-off is that a data center is just one node in a layered and globally distributed value chain.^[2] OECD's recent benchmarking of AI supply shows chips, data centers, clouds, models, connectivity and end users form part of the same supply chain, which has a structure characterized by large fixed costs, scale economies, vertical integration and bottlenecks not exactly aligned with taxes across national frontiers.^[3] OECD analysis shows that leading AI firms increasingly operate across several layers of the infrastructure stack, while countries differ markedly in their specialization across compute, cloud, models and applications. A host country can thus earn property taxes, construction-related activity, operator and contractor wages, electricity payments and utility revenues from hosting the data center, while much larger rents go elsewhere to chip design, semiconductor fabrication, hardware suppliers, cloud hyperscalers, model vendors, integration software, proprietary data owners, adopting firms and final investors.

A single charge against the coal in the mines, or those contracts, or those silicon wafers, involves other geographies of rent. For computing, geography has to determine where the servers actually run. For energy geography, the question is where electricity, capacity, storage, cooling and transmission lines are supplied from and reinforced. For labor geography, there is the question of where the engineers, technicians, managers, AI researchers, software engineers, facility technicians, clients and end-users are. For intangibles, it is where legally protected models, patents, the software program, or the data rights are owned or exercised. For corporate, it is where the subsidiary posts its contract, invoices and reports. For markets, it is where the paying users or consumers are. Each separate geography may imply a participant in value creation, a legal interface and a political policy base. Entry into the data center, as if it encapsulated all those six geographies simultaneously, erodes distinctions maintained for inextricable reasons by international tax law, by utilities regulation and by industrial economics.

This differentiation is clearly observable in tax doctrine.^[4] According to the OECD Model Tax Convention, a server may qualify as a fixed place of business under certain conditions, but a website by itself has no taxable presence and ordinary hosting arrangements do not usually bring the server into use for the business tenant at the hosting.^[5] Where a permanent establishment does exist, the location of the hosting service does not alone

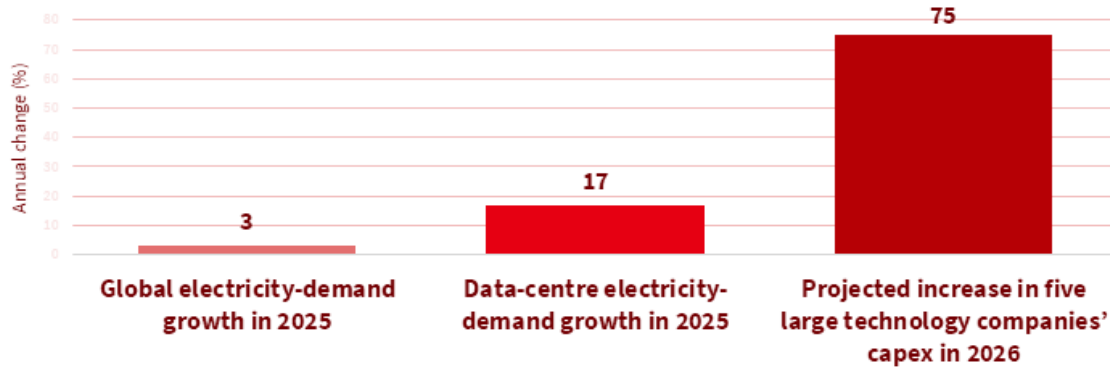
tell how much profit is attributable to it. Transfer-pricing rules draw another difference; the legal owner of an intangible property does not by itself receive the residual return if the development, enhancement, maintenance, protection and exploitation of that intangible property are carried out or controlled elsewhere.^[6] Put otherwise, neither the server nor the nominal owner of the source code is sufficient to define the locus of AI rent.

The truth is therefore more subtle and the mainstream rhetoric is simply too coarse to capture what is realistic. Data-center taxation is feasible when it recovers identifiable local costs, when it accounts for local externalities and when it only taxes the incremental operating income of the data center that is genuinely attributable to it. It ceases to be analytically sound and may even prove distortionary when it is believed to be a surrogate for the broader economic rent created along the entire AI value chain. The server has a place in the matrix, but indispensability should not be confused with platform dominance. Sound policy must distinguish between taxing local environmental and physical infrastructure externalities and taxing facility profits and claiming taxing rights over rent extracted from cloud, model, software, data and IP ownership and design and capturing the tax domain of the jurisdiction in which paying users live. The rest of the paper explores this nuance in terms of the AI value stack, the multiple geographies of AI, the justifiable claims of host jurisdictions, the shortcomings of data facility taxation as a proxy for AI rent and an application to the particular strategic dilemma now facing Europe.

2 The AI Value Stack: Seven Layers of Value Creation and Rent Capture

The first is that an incomplete tax proxy for the data center exists because AI value is not generated on a lone asset but on an entire stack. Closest to the value chain are semiconductors. The need for powerful AI capabilities pushes for GPU and AI accelerators, high-bandwidth memory, foundry services such as TSMC's and manufacturing capacity, all of which are characterized by high capital intensity and high concentration and differentiation.^[7] Previous OECD research on AI infrastructure already demonstrated the fact that the GPU market is highly concentrated, that complementary software (e.g., CUDA) is integral and that specialized upstream suppliers, including TSMC, ASML, SK Hynix and Samsung, occupy critical positions without necessarily participating in consumer cloud application markets.. Whenever compute becomes limiting, or advanced-generation chips are constrained due to supply problems, then a large share of rent is specific to this layer, rather than the later data centre.^[8]

Growth in Electricity Demand and Technology-Company Capital Expenditure; 2025–2026



Note: The first two bars show observed electricity-demand growth in 2025; the third shows projected capital-expenditure growth in 2026. The indicators have different bases.

Source: IEA; Data centre electricity use surged in 2025; 2026

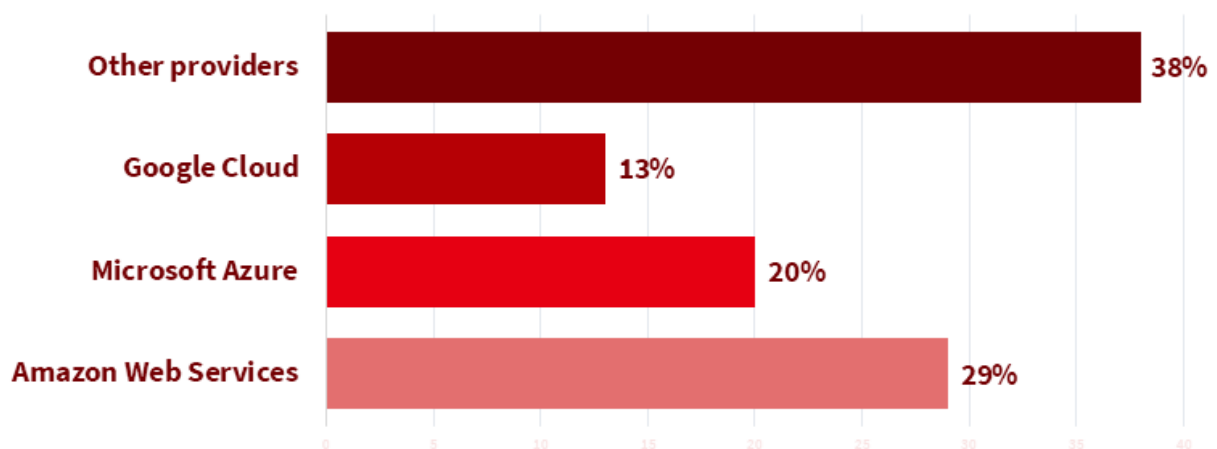
Figure 1: Infrastructure pressure is spreading across electricity, chips and capital expenditure rather than concentrating solely in server facilities.

The second layer includes servers, networking, storage and the construction of the data center.^[9] This is the layer most visible to local decision makers since it encompasses the buildings, racks, cables, switchgear, site preparation and complex engineering works. It is also the layer most often called upon in the eyes of many by public debates about the locus of AI. Yet even here, the economic footprint is quite equivocal. Some value is gained on-site in terms of civil engineering works, construction services and routine operations. Much of the value in the provision of advanced AI flows to globally active equipment vendors, engineering consulting firms and integrated infrastructure suppliers, whose earnings do not remain local. OECD studies emphasize that modern AI systems are reliant not just on an anonymous warehouse of servers but on advanced interconnects, resilient networks and digital backbone infrastructure -a good proportion of which is owned or coordinated by the world's most major technology players with a global presence. The physical plant thus remains a requirement, but is often not the primary owner of the space of strategic value, which renders the shell itself valuable.

The third layer includes electricity, the grid, access, cooling, land and water.^[10] Here, the argument for local public benefits is strongest. Power is almost always the largest operational expense of running a data center; heating or cooling loads and power densities are rising dramatically for AI-centric data centers; and according to the International Energy Agency (IEA), it may not be long before a single AI server rack consumes maximum electrical power similar to that of dozens of households.^[11] OECD analysis concludes that these data centers consume enormous amounts of electricity, often excessively consume resources like water and require large, costly cooling infrastructure. Electricity supply and water supplies and discharge systems are geographically sticky because the location of grid connection points, substations and water will be explicitly local. This lets hosts claim the value for the environmental costs and burdens of the local infrastructure and resources, even if the rent flows up the stack to many global owners.

The fourth layer is the cloud-computing and orchestration services.^[12] Economics is now no longer about physical raw materials. Rather, they are about having control over the shared capacity, interfaces, middleware, scheduling, managed services and ecosystem access. Since the OECD shows that the largest hyperscalers are especially well positioned to claim large proportions of the AI cloud market, that combined hyperscaler market shares are consistently very large across national and regional studies and that a growing proportion of large AI firms go to market primarily via partnerships through which cloud providers provide both capital and compute, then it follows that much of the product is in fact the contract, not the server hall in Finland, Ireland, or Virginia; while large proportions of the quasi-rent can be claimed where the racks are in another country and the customer never actually sees its servers.

Global Cloud-Infrastructure Services Market Share; Q3 2025



Note: Shares are rounded; other providers is calculated as the residual to 100.

Source: Synergy Research Group; Q3 2025 cloud-infrastructure services market shares; reported by TechRadar; 2025

Figure 2: Cloud concentration shows how control over contracts and orchestration can capture value above the physical facility layer.

The fifth layer is the foundation models and model APIs.^[13] Frontier-model developers can capture rents through scarce expertise, safety fine-tuning, brand, switching costs, first-mover benefits, proprietary evaluation pipelines and the ability to commercialize a pretrained model over numerous downstream applications. However, these companies are not necessarily completely vertically self-reliant. BIS evidence demonstrates that private frontier model companies like OpenAI and Anthropic have limited direct participation in compute and cloud markets and utilize close partner and infrastructure arrangements. This matters for tax analysis because the model rent may not sit with the data-center operator or with the customer-facing user, but with a third party who licenses or provides access to the model via cross-border contracts, which are themselves possibly routed by different connected firms.

The sixth layer comprises data, software, applications and distribution. This is where narrow model capacity can translate into industry-specific commercial value. Firms that own or maintain proprietary industrial data, customer workflows, embedded software, enterprise integration capacity and distribution channels may be able

to claim economic returns that aren't readily reducible to the raw compute or frontier model. Evidence on national AI ecosystems indicates that economies without leading positions in compute or frontier models may nevertheless specialize in downstream applications, software and distribution. This is an important antidote to data-center-tax narratives. Even if the most intensive compute activities are thousands of kilometers away, a significant portion of the value that can be effectively appropriated may be claimed by those firms that translate the models to the relevant markets, own the relevant workflow, or maintain the route to the customer. The seventh layer is the adopting firm and its final customers.^[14] Economically, not all of the value of AI is captured as taxable profit by upstream digital companies. Some of the value is transferred down as lower prices, better/cheaper products, better quality, less downtime, faster design cycles, higher yields, or newer output in the adopting company. Some accrue to end consumers. Some is earned by low-cost producers of an imported model, on an imported cloud, who build a factory, hospital, supply chain, or professional service that is more productive. So the largest social or private gain from AI adoption can potentially be realized outside the data center and outside the model vendor company. The AI value chain, thus, does not end when an individual facility powers up a GPU cluster; economically, it ends where the adoption shifts output, margins, quality and market power.

In principle, these seven levels can also see their rents shift over time.^[15] During times of chip shortage, rents may accrue to upstream suppliers of design, foundries and memory. When power-rather than efficiency-becomes the critical bottleneck, tremendous rents could go to the few rare grid access providers or the handful of players capable of quick energization. When cloud monopolies tighten their grip in the coordination layer, tremendous rents might eventually go to those few providers who link platforms together and those firms that lock in entire ecosystems of users. Where model interface measures become commoditized and where open alternatives flourish, these rents may migrate to all but the earliest stages of industrial applications, to industrial data, or to the distribution layer. Where adoption diffuses to all firms, the distribution of rents may shift again toward those firms that best reorganize production itself. The available evidence therefore, suggests that policy should not assume that current rent-holders will necessarily hold power forever because their facility is eventually heavily taxed and therefore fixed in place.

3 Six Geographies of AI Value: Compute, Energy, Labor, Intellectual Property, Corporate Structure and Markets

Since the AI value stack is layered, the production of AI services and products also occurs across multiple geographies that are only partially overlapping.^[16] Compute geography specifies site of server and accelerator installation and job execution; energy geography specifies site of electric power and grid capacity, backup, cooling and transmission reinforcement; labor geography identifies where model engineers, site of facility operators, where associated professionals work, site of people working with the output; intellectual-property geography identifies where ownership of models, software, patents, datasets, or licenses; corporate geography specifies the location of the parent and subsidiary contracting, invoicing, funding and profit-reporting site; market geography

specifies the location of consumers and end users of AI products and services. The location of the economic value stems not from any one of these, but from the interaction of all six. Compute geography and energy geography are not the same thing, even on the same site. The IEA finds that the siting and expansion of data centers is highly dependent upon local grid variables, queue times and capacity to bring large loads online.^[17] The OECD finds that the availability of water and cooling can influence siting decisions. A Finnish or Swedish site may be selected because its temperature, electrical mix, water availability and permitting process will enable it to support inference or training loads efficiently. Such a choice says a lot about local infrastructure scarcity and local system costs, but it does not mean that Finland and Sweden become the sole source of the firm's model, customer base, or reported profits.

Labor geography and market geography often tend to differ even more from compute geography. The team working on the initial development of the model might be based in California, London, Paris, or Berlin. The operations team administering and updating the site might be based in Finland. The customer success team might be based in Dublin or Warsaw. The engineers actually using the system might come from Bavaria and the head of production might come from Poland. The organization paying for the service might be a German manufacturer, but its customers will be spread across the rest of the European Union. In terms of taxation, these differences matter because production, consumption and business activity can all stay within the jurisdiction without substantial local compute capacity. In terms of economics, they matter because the productivity effects of artificial intelligence are powered by organizing a firm once it adopts the technology – not by the servers in the data center being cooled down.

Another level of stratification was created between intellectual-property geography and corporate geography. In many cases where commercial AI services are sold into the EEA, the services may be contracted through Irish entities with respect to the service, even if the technology is created elsewhere. OpenAI's European terms state that when European EEA and Switzerland residents seek to contract with OpenAI, customers that reside within the EEA or Switzerland shall enter into the terms of the agreement with OpenAI Ireland Ltd; whereas the contractor-entity schedule for Google Cloud states that for much of EMEA, the contracting-entity name is Google Cloud EMEA Limited, stated in Dublin.^[18] The mere presence of such foreign locations (for services, contracting entities, or even ultimate parent companies does not necessarily reflect where profits are ultimately taxed and potentially where profits will be taxed, even if no activity is directly linked to holding that specific location), but the location of an Irish contracting entity remains a possibility for a European customer where the model is created in the US, run on servers within another Member State, and at the same time, under an ultimately U.S. parent company. For representative factors such as who is legally invoiced, where the server is located, where the ultimate parent company is located, and who is the beneficial owner, the distinctions are difficult to discern.

The hypothetical commissioned by a German manufacturer, using an American model, via a contracting Irish entity, running inference in a Finnish data center, on chips designed in the United States, fabricated in Taiwan, applying German industrial data to improve the output of a Polish factory but still owned by a parent corporation in Germany, is economically and legally plausible. Each element is a described pattern elsewhere in the current AI economy. Google publicizes Hamina in Finland as one of its live data-center locations.^[19]

OpenAI and Google both contract some of their European work through Irish entities. Publications by the OECD demonstrate that chip design, fabrication, cloud infrastructure, models, and applications are themselves geographically distinct and frequently located in separate jurisdictions. The real legal point is not to claim the example as exhaustive or universal, but rather, to demonstrate that it is theoretically, legally, and economically feasible for the one unit of AI-enabled value to pass through no fewer than six jurisdictions before taxing rights must be allocated among jurisdictions.

Once that multi-jurisdictional chain is recognized, the boundaries of a facility tax are clear.^[20] An Irish or German tax on the value generated in a data center in Ireland or Germany might be justified to the extent that the data centers are the means by which land-use authorizations, access to the grid, cooling infrastructure, ecological regulation, connected transport and communication links, and the public services and streets of a neighborhood are supplied. An Irish, German, or other national taxing the profits generated by servers within a particular permanent establishment or facility business is probably justified to the extent that they are collecting from the operation of the servers taxes on the resources used by the servers, be they the services of a Polish, U.S., Taiwanese, or other chip designer or the running costs of an Irish contract manufacturer. But the home country of the server is not in and of itself in a position to determine what part of the overall AI surplus should go to the Taiwanese foundry, the Irish contractor, the German consortium, the American chip home, and the European end user.

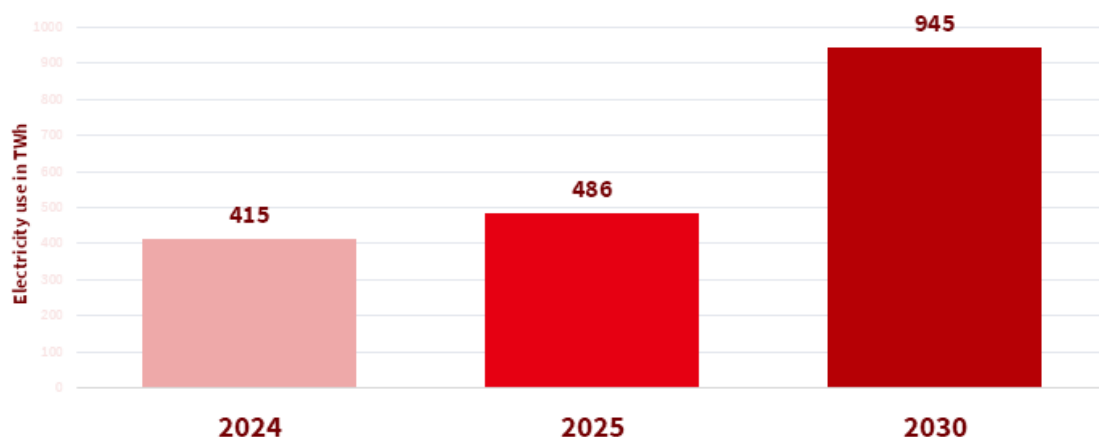
4 What Data-Centre Host Jurisdictions Can Legitimately Tax and Charge

By untangling the broader geography of value in AI, the fiscal claim of host jurisdictions is not any weaker but rather clearer. There is a compelling normative argument for a data-center host to charge for what it supplies in practice, and for adjusting for the additional burdens that the facility places on the local systems.^[21] These norms are familiar: benefit taxation, user charging, recovering costs, and Pigovian correction. If a data center occupies scarce land, necessitates rezoning, consumes public infrastructure, and its operation strains constrained grid capacity, the host jurisdiction can legitimately charge to recoup the costs. This claim carries its greatest force when it is implementation-neutral, can be relatively quantified, and is defined in relation to the immediate, local facility rather than certain purported rights over the rents of the distant AI stack. That logic clearly leans towards natural land-use controls, property taxation, building- and civil engineering-related charges, and local authority permitting fees. A server hall is a building on land. It may necessitate road access, public safety supervision, flood or fire planning, and utility cooperation. The IEA points out that data centers can become operational in between 2 and 3 years, which on its own indicates a significant and concentrated development process, even if the broader energy system required to support them proceeds at a slower rate. There is therefore nothing conceptually unusual about considering the facility in its initial form to be no different than other large industrial or warehousing assets when it comes to land, building, and property taxation. These are not levies on AI rent in the abstract; they are levies on land-use and built environment infrastructure on a site-specific

capital asset.

Grid-connection costs, transmission and distribution upgrades, and capacity-reservation charges are defensible if they correspond to system costs.^[22] One must make a special case here because data centers are capable of creating system costs that extend well beyond their fenced perimeter. For instance, IEA analysis of Europe shows waiting times for grid connections averaging over two years, core European hubs experiencing seven-to-ten-year queues, and the European Union needing some policy coordination simply to match project pipelines with existing electricity infrastructure.^[23] ACER reports that in 2024 EU transmission system operators spent 4.3 billion on various remedial actions to address grid congestion.^[24] If a facility needs a dedicated network line, a new substation, flexible connection arrangements, or network reinforcement, then the host system is not just hosting a digital service but securing limited grid capacity. Costing that scarcity with market forces is internally consistent and can help guard against other ratepayers being burdened by implicit cross-subsidy.

Global Data-Centre Electricity Use; 2024, 2025 and 2030



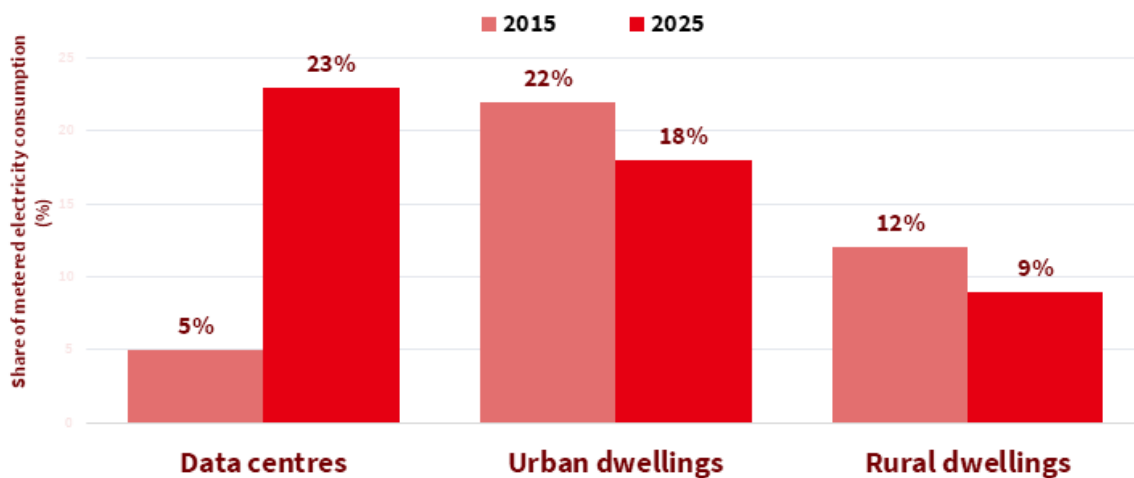
Note: 2024 is an estimate; 2025 is calculated by applying the IEA's reported 17% annual growth rate; 2030 is a projection.
 Source: European Commission; Energy performance of data centres; IEA; Energy and AI; 2025; IEA; Key Questions on Energy and AI; 2026

Figure 3: Rising electricity demand strengthens the case for local infrastructure charges, but not for assigning the entire AI rent to the host jurisdiction.

Water use, cooling, carbon emissions, externalities of backup generation, and environmental monitoring are other justifiable local compensations. OECD analysis emphasizes that many AI data centers are water-dependent in cooling requirements and that water access is a cause of location; it also concludes that cooling technologies are changing, so local policy should focus on actual resource use rather than a blanket anti-data-center approach. The EU's second-generation data-center energy-performance framework has explicitly included monitoring of water usage once the facility becomes operational. Carbon or local air-pollution taxes are equally justifiable if on-site generation or backup-generator testing, or local particulate matter emissions impinge on the inflowing community. Again, the relevant tax or charging base is not the global AI but the physically present load and its measurable externalities. Ireland is an example of how such a claim can be meaningful. Central Statistics Office Ireland has stated that data centers were responsible for 5 percent of metered electricity

consumption in 2015, 22 percent in 2024 and 23 percent in 2025.^[25] At these levels, the host-country validity of concerns over the adequacy of the grid, capacity planning, limited locations, and fair distribution should be clear. Such concerns are clearly not merely rhetorical; they are indicative of how a significant proportion of an actual national electricity system should be allocated. Slightly smaller-scale data-center concentrations could cause similar host country concerns in other parts of the world. This justifies the prediction that the charging of infrastructure and pollution costs is not merely a pretext for taxing intangible profit flows, but an actual attempt to deal with the substantial, geographically distributed consequences of digital loads that require substantial infrastructure.

Ireland Metered Electricity Consumption by User Group, 2015 and 2025



Source: Central Statistics Office Ireland; Data Centres Metered Electricity Consumption 2025; 2026

Figure 4: Ireland's rising data-center share of metered electricity demonstrates the scale of a legitimate host-country infrastructure claim without locating the entire AI surplus there.

European law and policy now increasingly mirror this logic of host jurisdiction precisely. The recast Energy Efficiency Directive requires Member States to cover reporting by significant data centers, and the Commission's 2024 delegated regulation established the database and KPIs for that reporting.^[26] The Commission's data-center performance page states that the database collects information on energy performance and water footprint for sites with significant energy consumption, and earlier Commission documents established the reporting threshold at built-in (IT) power consumption of at least 500 kW.^[27] The very same framework makes clear that the legal regime in place, not yet evolved into an EU tax instrument, is essentially a transparency and monitoring one. It is not a mature EU tax instrument even now, and no attempt should be made to categorize it as such.

As of mid-2026, the Commission is preparing a Data Center Energy Efficiency Package that would include baseline data collection, a rating scheme, and an electronic label, followed by work on minimum performance standards.^[28] The Commission's March 2026 call for evidence outlined the proposed rating regulation as a follow-up to the 2024 delegated regulation and the Commission's broader energy efficiency roadmap, which identified

minimum standards in a consultation phase.^[29] In other words, EU reporting and collecting obligations are in place, but the rating-label regime and any mandatory minimum performance standards are a work in progress, not a finalized, fully functioning tariff and regulatory scheme. This distinction is central to accurate analysis. Existing standards and labels could influence costs and investment signals, but they should not be conflated with implemented fiscal measures or an EU decision to extract rents from global AI by taxing server venues.

Development of the EU Data-Centre Reporting and Performance Framework; 2023–2026



Note: The reporting framework is operational; the rating scheme and minimum performance standards remain under development.

Source: European Commission; Energy performance of data centres; 2026

Figure 5: The EU framework has advanced from reporting toward rating and prospective standards, not toward an implemented tax on global AI rent.

The correct policy position, therefore, becomes the disciplined rather than anti-host one. Data-center hosts could legitimately charge for land, property, building, and connection, congestion, environment, monitoring, and facility-flexibility profits, impose locational conditions, and negotiate obligations of system-integration (such as use of waste-heat or flexible loads) where appropriate. However, these charges should be aimed at as site-specific public financing tools, not policy rhetoric used as an improper substitute for taxing the whole of AI. After the financial purpose is changed from recovering local costs to capturing the entire surplus of global AI, this data center is no longer an appropriate tax base and a comparatively inappropriate substitute.

5 Why Physical Server Location Is an Incomplete Proxy for AI Rent

The most significant logical error in the data-center tax proposal is a category error. A data centre is a physical input into AI, but it cannot be the only residual claimant on the model, be the sole owner of a customer relationship in AI, or stand to gain from the returns to AI adoption. In many commercial arrangements, it is a cost center, an infrastructure asset, or an inception node. The rents that really do matter may not be sitting on a facility but on a cloud platform distributing finite compute time among users, on a model supplier turning a boundary API into revenue, on a software company leveraging AI in the enterprise's processes, or on an in-house company sharpening its margins and output.

This is further underscored by demand heterogeneity. Data centers are not solely for AI use, and even the abrupt current spate of activity in AI does not change the multipurpose character of the infrastructure it is built upon.^[30] The IEA estimates that accelerators (of which the majority are AI accelerators) constitute nearly half of the net growth in global data-center electricity use by 2030; conventional data-center servers still account for just under a fifth, and the remaining part is accounted for by data-center infrastructure and other IT equipment.^[31] All three main types of data centers: enterprise, colocation, and server provider, and hyperscale (including AI-driven workloads of any type) contribute to the overall demand increase. Any facility-level tax base attracted by AI rent will therefore almost certainly appear as some combination of AI and non-AI workload within a mixed-use facility (absent a transparent way of individualizing them in meters, which the public debate does not seem to assume); the potential end result could be a tax on non-labor digital infrastructure with only partial links to frontier-AI business phenomena.

Training and inference also differ in ways that weaken the argument for a simple facility proxy.^[32] OECD analysis describes training as extremely expensive, requiring high-density clusters, specialized accelerators, high-bandwidth networking, power, and cooling. Inference inevitably also consumes hardware optimized for energy efficiency (or latency). But as services reach the point of commercial scale, they harness device-to-edge latency effects, and cloud providers also benefit from developing optimum local or regional infrastructure, for example, for security reasons. Carnegie's analysis of data-center competitiveness shows that for major compute investments, time to power, project speed, and the availability of infrastructure may dominate over even huge differentials in local taxes or electricity prices.^[33] In fact, these two facts point in different local directions. Frontier-model training may often be best concentrated where power, chips, financing, and permitting are all readily available. Inference, by contrast, needs to be assembled along the geographies where latency, customer proximity, security, and institutional conditions are optimum. A single, uniform facility tax may therefore not be adequate to capture the value spread across all AI workloads.

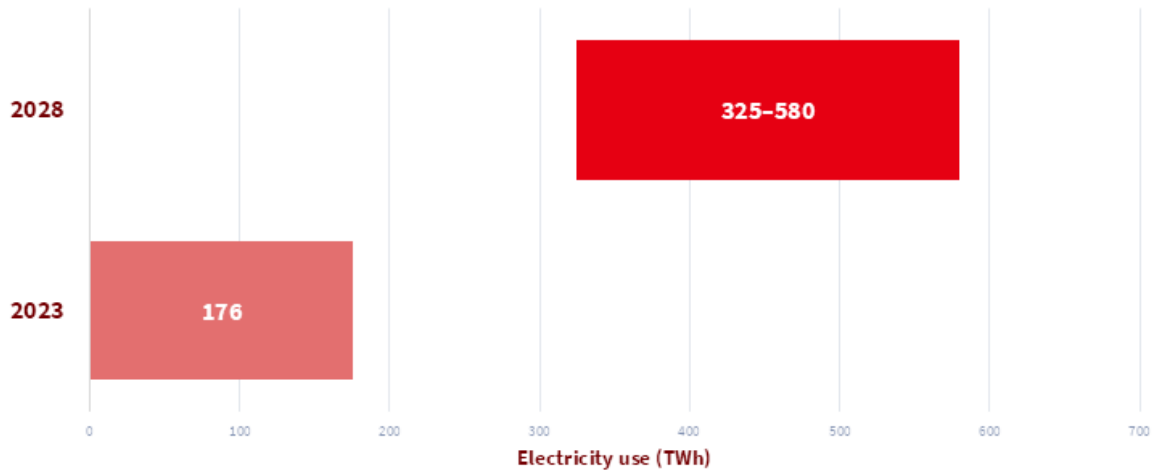
Legal complications extend the physical/taxable location divide. The OECD Model Tax Convention supports three propositions that are relevant here: a website by itself is not a place of business, a hosting arrangement will generally not involve a server being made available to an enterprise, and even where a server or related facility does constitute a permanent establishment, that alone will not determine the division of profit attributable to it. Further, the business conducted on the server must be analyzed on a case-by-case basis; attribution is a separate determination. Modern cloud architectures heighten this importance because a cloud client accessing AI may never own, lease, or have managerial control over the specific server involved in generating a model response. The transfer-pricing rules introduce a second form of incompleteness. The 2022 Transfer Pricing Guidelines issued by the OECD are explicit that legal ownership of an intangible asset does not confer the right to a 100 percent share of its returns: what counts is who does, supervises, and bears the risks for development, enhancement, maintenance, protection, and exploitation. If the foreign-owned affiliate holding the legal title to a platform, portfolio of patents, or software stack doesn't control or perform those functions, that affiliate shouldn't be guaranteed to get all the income flowing from those activities. The other side of this coin is equally significant for the facility argument: a data-center company that supplies power, space and operations but doesn't own or develop the intangible shouldn't automatically be where the model rent is located. All manner

of revenue flows-royalties, license fees, cloud charges, cost-sharing arrangements, services agreements-can be used to redistribute income from the physical location, subject to arm's-length and treaty limitations.

A blunt facility charge can be distortionary because, rather than falling on the parties best able to bear or pass on it, it targets a plausible global bottleneck. OECD evidence demonstrates that the most substantial AI and cloud companies have scale, scope, vertical integration, and multi-layer presence. Firms active across compute, infrastructure, models and applications are generally those best able to push workloads to others, repackaging contracts, securing high-powered deals or bearing higher facility costs for long periods. These advantages, which are unequally distributed within and across communities of local and global research users, universities, start-ups, and incumbent providers, will come into play in the event of a facility tax and may prevent it from falling solely on the incumbent rent recipient. Private providers, domestic research organizations, universities, and local start-up providers will be less able to pass that fee through to their own constituents. So targeting the rent through a facility tax may be costly even if the tax is invoked rhetorically.

The difference between operational and announced capacity aggravates the problem.^[34] Many public conversations about AI data-center booms often move from project announcements into assumptions about realized rent. However, the IEA Europe analysis distinguishes between pipeline capacity and installed capacity. It reports that the project pipeline in Europe implies a capacity of 130 percent of the current installed capacity, but that installed capacity by 2030 will increase by only about 70 percent in relation to 2024, due to delays and local restrictions (for example, on emissions or controls on end-user access due to grid congestion). In other words, levies or obligations related to announced capacity, memoranda of understanding, or speculative AI-hub announcements could be put in place long before the manufacturing activity is capital-intensive, operational, profitable, or yet to be built. For an already tenuous attempt to tax emerging AI rents, such a timing mismatch makes the proxy even less robust. On the whole, the takeaway is not that taxation of facilities is misguided-it is just a more limited tool than most supporters believe. It is effective at valuing local externalities and taxing the operations of facilities. It is ineffective at embodying the putative makeup of AI-generated rent when the chips, the models, the contracts, the data, the producers, the adopters, and the shareholders are frequently offsite. The farther away from local cost return a policy goes, and the closer it gets to claiming a share of AI's total economic surplus, the more likely it is to be invalid analytically and economically distortionary.

U.S. Data-Centre Electricity Use; 2023 Estimate and 2028 Projection Range



Note: The 2023 value is an estimate; the 2028 values represent the lower and upper bounds of the projection range.
 Source: Lawrence Berkeley National Laboratory; 2024 United States Data Center Energy Usage Report; 2024

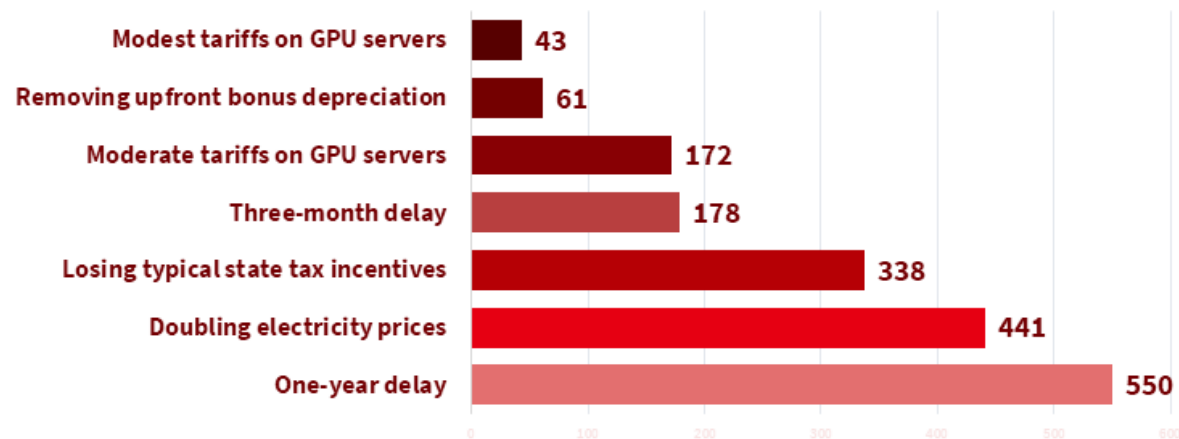
Figure 6: The broad projection range demonstrates why prospective capacity should not be treated as already operational or profitable.

6 Europe’s Policy Trade-Off: Cost Recovery, Compute Expansion and the Risk of Taxing the Wrong Base

The European policy risk results directly from the above analysis.^[35] Europe is now pursuing two objectives. On one side, it is tightening up data centers’ energy efficiency, use, and integration into systems. On the other hand, it is explicitly trying to grow its domestic cloud and compute capacity, as part of a new overall agenda for AI and technology-sovereignty. The Commission’s AI Continent plan envisages tripling EU data-center capacity at least within five to seven years, and its proposal for a Cloud and AI Development Act aims at easing deployment and access to energy, land, water, and finance, towards building up strategic capacity for AI, cloud, and computing-intensive applications.^[36] Such a policy mix does not sound inconsistent at face value. It can only sound inconsistent if Europe is going to regard the data center equally as a limited factor of production that must be grown and as a fictional tax base convenient for extracting the entire rent of AI.

The pain is genuine because the pain of Europe’s energy constraints is genuine. The IEA finds EU grid-connection waits that can range from two to ten years, and concludes that Europe will not realize the full volume signaled by its project pipeline under current constraints. Data centers are also expected to contribute 10 percent to EU electricity-demand growth by 2030 in current policy settings.^[37] In that setting, cost-reflective local levies are not a policy indulgence; they are a necessity both for economic management and public support. Host communities and network users cannot reasonably be expected to freely carry additional congestion, capacity shortages, water use, or system-upgrade costs. The mistake is elsewhere: in stacking on non-cost-reflective fiscal loads to imitate an AI rent tax, and simultaneously expecting a faster rollout of private digital infrastructure in a region already constrained in the squeeze between time to power.

Estimated Life-Cycle Value Losses for a 100 MW U.S. AI Data Centre; USD Millions



Note: Modelled reductions in life-cycle value for an illustrative 100 MW U.S. AI data centre; figures are not observed project costs.

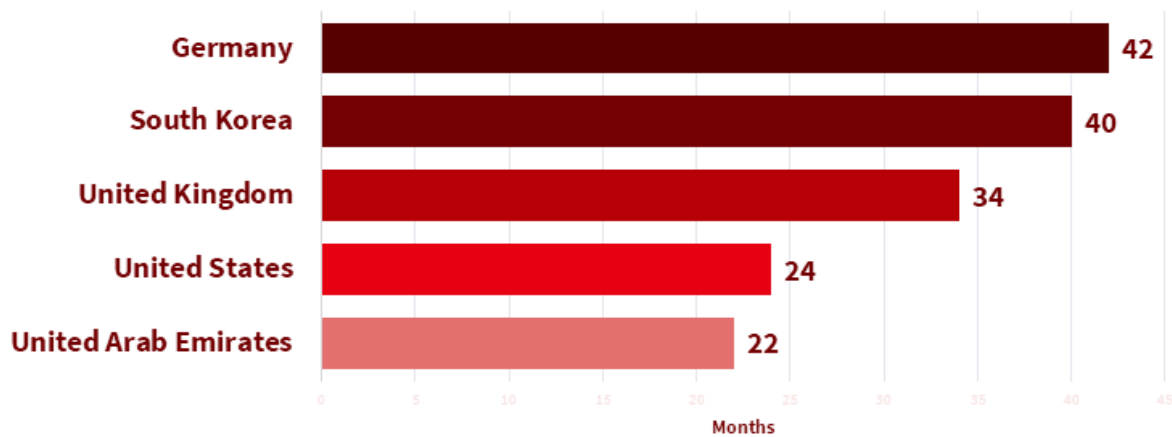
Source: Carnegie Endowment for International Peace; The Compute Coalition; 2026

Figure 7: Differences in deployment time can materially affect where new compute capacity is built.

These goals can, however, be partially reconciled. The first step is an honest assessment of the mission of each instrument. Cost-reflective connection fees, capacity-reservation charges, water charges, carbon tax, and environmental reporting covered local costs and externalities, whereas easier permitting, queue management, flexible interconnection design, and better planning addressed deployment bottlenecks. The IEA states that non-firm connections and queue management would provide a way to upgrade ready-to-build projects; and the Commission is developing so-called tripartite models between data center operators, other energy players, and public authorities precisely to make data centers an integrated part of the local energy system and heat networks. A Europe that makes data centers pay for the costs they impose and also reduces unnecessary permitting and grid frictions is not contradicting itself; a Europe that seeks to morph facility charges into a new, lower global AI profit surcharge does contradict itself.

The idea that certain frontier-model training might take place in the Middle East and other resource-plentiful regions, while inference occurs nearer to European endpoints, should thus be put to the test rather than simply accepted or dogmatically thrown out. Carnegie's cross-national competitiveness index places the UAE high in part because of project speed, and because prolonged delay dominates other cost factors; but it also illustrates how political, economic, and security shocks, or conflict-driven delay, can cause rapid oscillations in such rankings. The economics of locating some training in power-providing and fast-permitting economies and geographies is sound. Large training loads are capital-dependent, schedule-sensitive, and more easily relocated than many policymakers appreciate. But that alone is not necessarily enough of a cure to Europe's need to bring certain inference workloads closer to EU users, or to preserve some strategically necessary compute capacity, either within or beyond the EU's borders, for resilience, security, and public sector control; or to realize the EU digital-ecosystem objectives that cannot be achieved simply by externalizing certain capability-creating activities.

Estimated time to First Operation for Major AI Data Centres



Note: Modelled median estimates derived from completed and announced projects; actual timelines vary by site. Estimates exclude initial land planning and pre-construction permitting.
 Source: Carnegie Endowment for International Peace; The Compute Coalition; 2026

Figure 8: For the illustrative facility, deployment delay produces a larger modeled value loss than higher electricity prices or the removal of tax incentives.

Both of the strongest potential counterarguments to the paper's thesis, therefore, warrant serious consideration. Data centers have the potential to produce significant investments, local construction expenditure, utility requirements, local government receipts, resilience gains, and ecosystem externalities; none of these arguments necessitates the assumption that the data center captures all of the rent accruing to AI, and none of them would be diminished if the Commission moved to regulate, tax, or subsidize facilities at the least-cost jurisdiction. The danger would come if one slides from those sensible host-jurisdiction arguments into the further conclusion that a tax on the host facility is therefore the best way of taxing AI.

This is also where the conceptual difference between facility taxation and market-jurisdiction taxation becomes essential. The 2026 briefing document on cross-border services taxation produced by the European Parliament states that digitalization means that services can be supplied from afar at near-zero additional costs, and describes one policy lever as trying to reverse the balance of taxation taxing rights – nations where the services are supplied to global markets instead of where the services come from.^[38] But whether or not one favors any digital-levy measure, this particular debate involves a different kind of tax entity, a different tax object entirely from a local charge on a data-centre facility. One involves global customers and globally sourced services. The other involves a locatable infrastructure asset. Mixing the two leads to bad policy in compelling a local utility-and-environment measure to do the job that a global profit allocation inquiry should.

The simplest and most justifiable European stance is therefore a tiered one. Taxes and charges on AI facility use should be narrowly correlated with local costs, congestion externalities, and environmental impact, and levied universally among all workloads, inclusive of AI, proportionately to their demands. Any policy to expand AI compute capacity should be targeted at shortening power-up times, integrating more additions to national grids, siting capacity where there is genuinely existing headroom, and protecting critical inference and

public-interest AI computing within the Union and allied safe havens. Where broader AI profits and rents are taxed, the taxation of broader AI profits and rents should be directed at taxing those profits and rents rather than relying on the mistaken assumption that the full equation of taxable surpluses resides wherever a server rack is sited, as long as it happens to be getting the power from a server rack. Europe's mistake would not be if data centers were made to pay their way locally. It would be treating that local payment as a substitute for taxing AI value where it is actually earned.

7 Conclusion - Tax the Facility for Local Costs, Not the Entire AI Value Chain

The case for taxing data centers starts from a solid intuition and would end up in a narrower conclusion than many of its proponents seem to desire. Data centers are fixed, resource-use-heavy sites that impose tangible burdens on land, grids, cooling, water, and local public services; the home jurisdiction can thus fairly seek to recover local expenditures, impose local externality prices, and tax the profits that are truly generated there by data center operation. It does not contain an unearned claim to the full AI rent in the form of the full economic rent, simply on the grounds of having a few servers stacked within its geographical jurisdiction. AI's value is spread across chips, clouds, model-builders, software system-stacks, proprietary data, user firms, and final customers: each with different geographies, jurisdictions, and legal considerations. A server hall is a single essential point on an AI value chain, but by no means is it the natural geographical location of that value.

The strategic implication is unambiguous. By using facility taxation as an approximation of AI rent taxation, Europe (or another state) may be collecting too much tax on a relatively immobile taxable input, and too little on more mobile, legally complicated forms of ability-to-pay and value creation. The right way forward is institutional separation. Local facility-specific charges to reflect relative resource costs, taxed in a tightly calibrated manner at the facility level, and a separate global-policy debate to attribute other profits, intangible assets, and jurisdictional rights. The priority is not the abolition of data center taxation, but the restoration of its proper function.

References

- [1, 2, 3, 7, 12, 15, 16, 32] OECD (2025) *Competition in Artificial Intelligence Infrastructure*. OECD Roundtables on Competition Policy Papers, No. 330. Paris: OECD Publishing.
- [4, 5, 20] OECD (2017) *Model Tax Convention on Income and on Capital: Condensed Version 2017*. Paris: OECD Publishing.
- [6] OECD (2022) *OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations 2022*. Paris: OECD Publishing.
- [8] Sastry, G., Heim, L., Belfield, H., Anderljung, M. et al. (2024) 'Computing Power and the Governance of

- Artificial Intelligence’. *arXiv*, 2402.08797.
- [9] Pilz, K. and Heim, L. (2023) ‘Compute at Scale: A Broad Investigation into the Data Center Industry’. *arXiv*, 2311.02651.
- [10, 27, 28, 30] European Commission (2026) *Energy Performance of Data Centres*. Brussels: Directorate-General for Energy.
- [11, 22, 23, 31, 34, 37] International Energy Agency (2026) *Key Questions on Energy and AI*. Paris: IEA.
- [13] Stanford Institute for Human-Centered Artificial Intelligence (2026) *Artificial Intelligence Index Report 2026*. Stanford, CA: Stanford University.
- [14] International Monetary Fund (2024) *Broadening the Gains from Generative AI: The Role of Fiscal Policies*. Washington, DC: International Monetary Fund.
- [17] International Energy Agency (2026) ‘Data Centre Electricity Use Surged in 2025, Even with Tightening Bottlenecks Driving a Scramble for Solutions’. Press release, 16 April.
- [18] OpenAI (2026) *Europe Terms of Use*. Updated 16 January 2026.
- [18] Google Cloud (2026) *Google Contracting Entity*. Google Cloud.
- [19] Google (n.d.) *Hamina, Finland — Google Data Center Location*. Google Data Centers.
- [21] OECD (2019) *Taxing Energy Use 2019: Using Taxes for Climate Action*. Paris: OECD Publishing.
- [24] European Union Agency for the Cooperation of Energy Regulators (2025) *Market Monitoring Report: Electricity Wholesale Markets*. Ljubljana: ACER.
- [25] Central Statistics Office Ireland (2026) *Data Centres Metered Electricity Consumption 2025*. Cork: Central Statistics Office.
- [26] European Parliament and Council (2023) ‘Directive (EU) 2023/1791 of 13 September 2023 on Energy Efficiency and Amending Regulation (EU) 2023/955’. *Official Journal of the European Union*, L 231, pp. 1–111.
- [26] European Commission (2024) ‘Commission Delegated Regulation (EU) 2024/1364 of 14 March 2024 on the First Phase of the Establishment of a Common Union Rating Scheme for Data Centres’. *Official Journal of the European Union*, L 2024/1364.
- [29] European Commission (2026) ‘Rating Scheme for Data Centres in the EU — Commission Launches Call for Feedback’. Directorate-General for Energy, 27 March.
- [33] Phillips-Robins, A., Tawil, T. and Winter-Levy, S. (2026) *The Compute Coalition: How to Build the Future of AI in the Free World*. Washington, DC: Carnegie Endowment for International Peace.
- [35, 36] European Commission (2025) *The AI Continent Action Plan*. Brussels: Directorate-General for Communications Networks, Content and Technology.

- [36] European Commission (2026) *Cloud and AI Development Act*. Brussels: Directorate-General for Communications Networks, Content and Technology.
- [38] Amaro, F. and Picciotto, S. (2026) *Possible EU Own Resource Based on a Digital Levy: Cross-Border Services Trade, Digital Transformation and Tax Implications*. Brussels: European Parliament.