

# DATA CENTERS, INTELIGÊNCIA ARTIFICIAL E IMPACTOS NA TRANSIÇÃO ENERGÉTICA

Anibal Traça de Almeida, Pedro Moura e Nuno Quaresma



Instituto de Sistemas e Robótica

Departamento de Engenharia Eletrotécnica e de Computadores

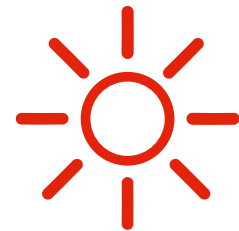
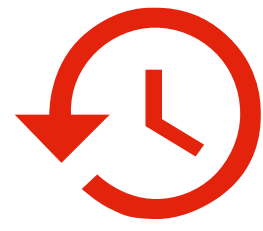
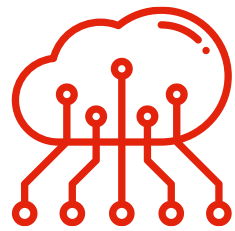
Universidade de Coimbra

# DATA CENTERS & ARTIFICIAL INTELLIGENCE

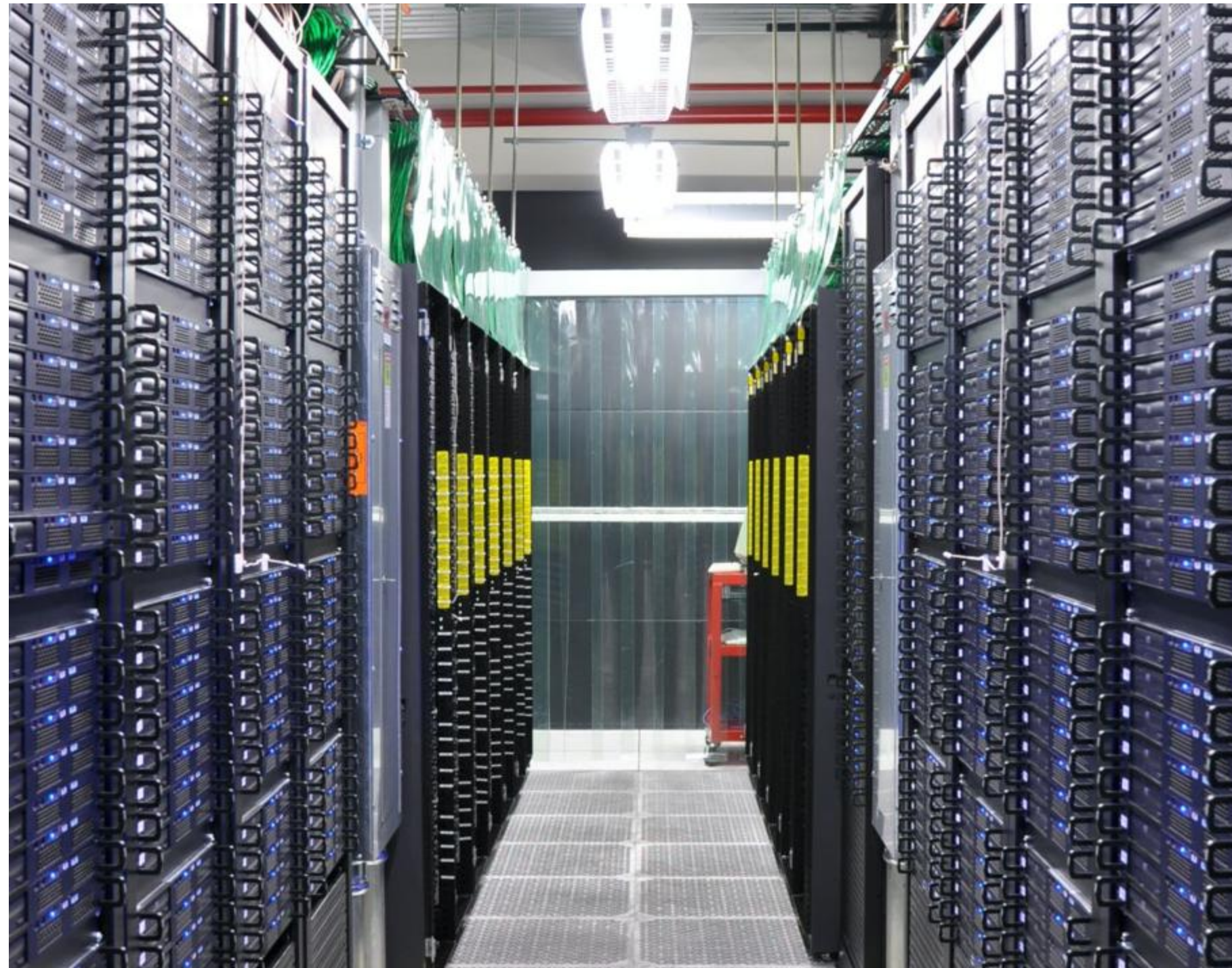


## Potential impacts in the clean energy transition

- Energy
- Grid integration
- Demand response mitigation strategies



# Inside of a data center, showing vertical racks with each rack containing many servers



**Source:** Nadel, Steven. 2025. Opportunities to Use Energy Efficiency and Demand Flexibility to Reduce Data Center Energy Use and Peak Demand. Washington, DC: ACEEE.

# Executive summary



- Data centres already use ~415 TWh/year (~1.5% of global electricity = 8 x Portugal consumption) and are projected to more than double by 2030 (~945 TWh).
- Local grid impacts can be outsized because capacity is geographically concentrated and often constrained at peak hours.
- If operated flexibly (load shifting + backup activation), data centres can help integrate new capacity—potentially needing flexibility for only ~0.1–1% of the time in some scenarios.
- Best mitigation stack: (1) efficiency, (2) flexible compute & demand response, (3) storage/microgrids, (4) clean procurement & co-location, (5) use free cooling (e.g. sea water)
- Policies matter: flexibility can cut system costs, but emissions outcomes depend on the grid mix—pair flexibility with strong clean energy policy.

# Scale & trajectory of electricity demand



Global data centre electricity (2024)

**415 TWh**

~1.5% of global

Projected (2030)

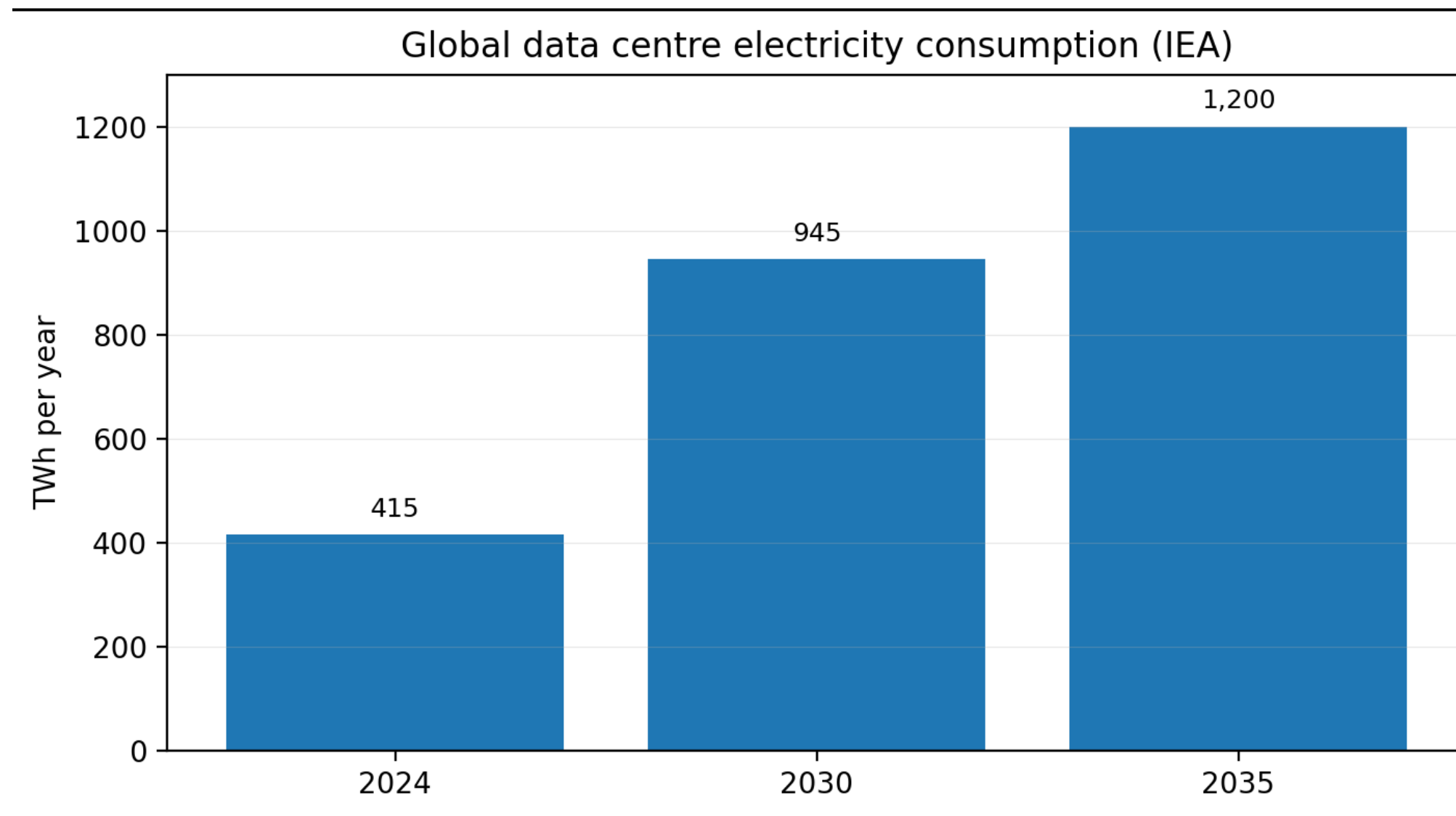
**≈945 TWh**

More than double

Projected (2035)

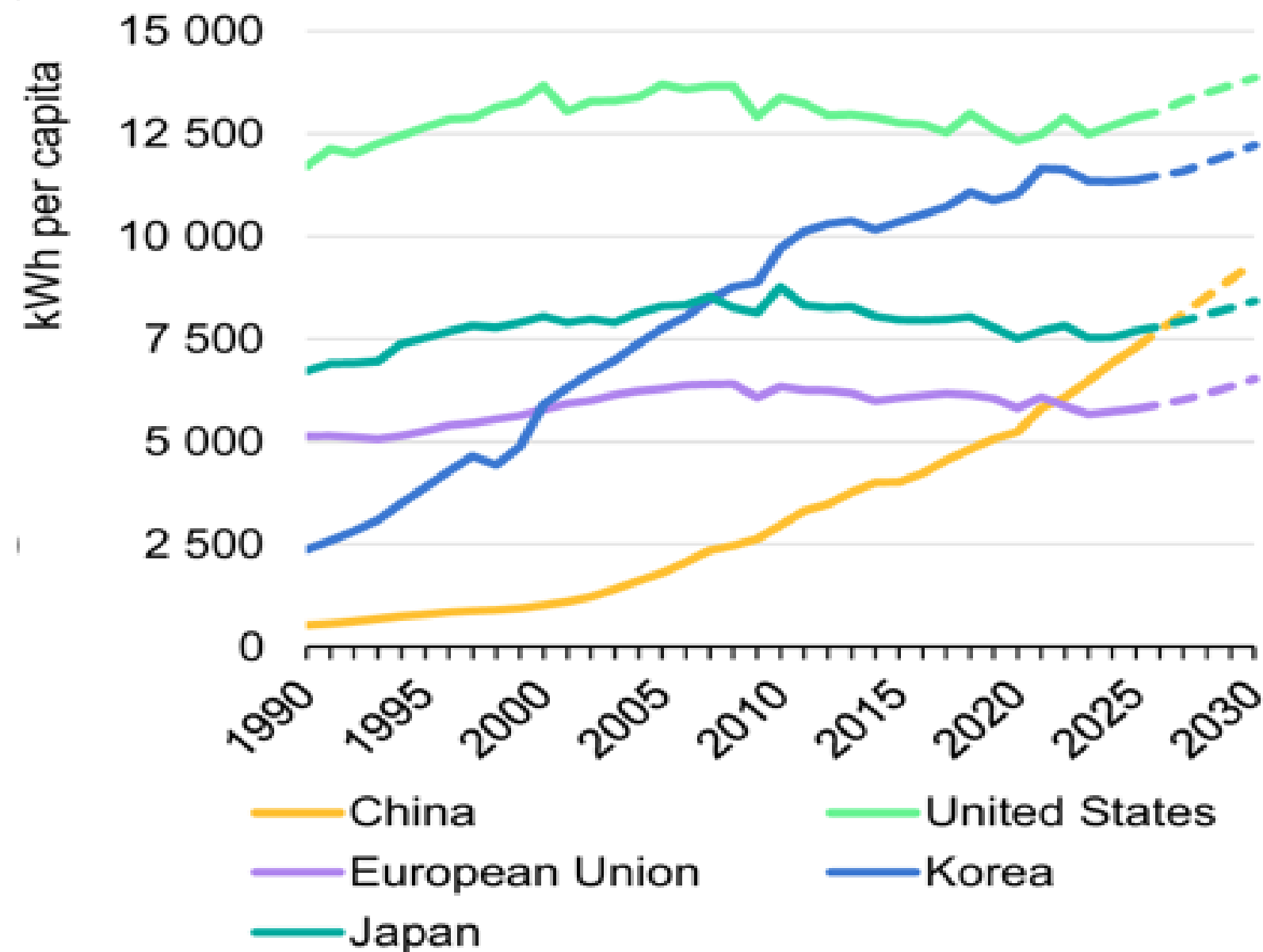
**≈ 1,200 TWh**

Base case



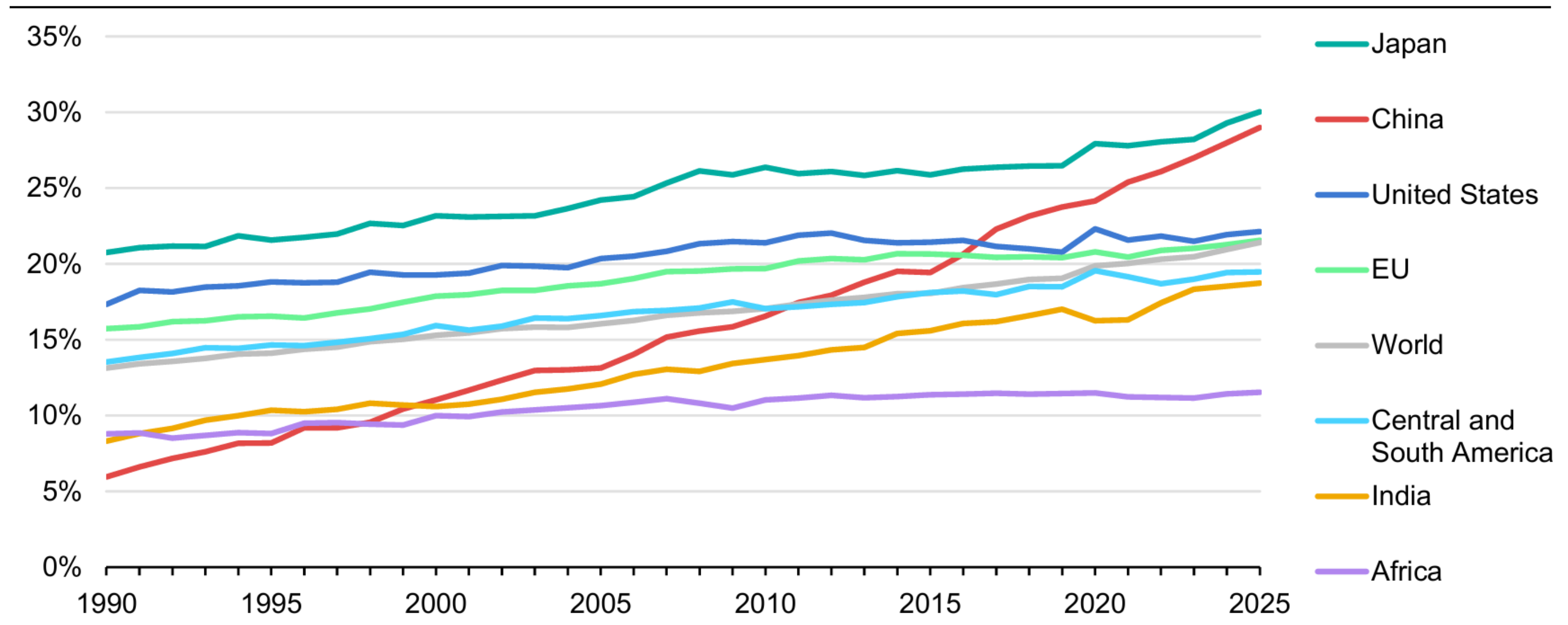
IEA projection (global): rapid growth to 2030, continued rise to 2035

# Electricity consumption per capita in selected countries and regions, 1990-2030



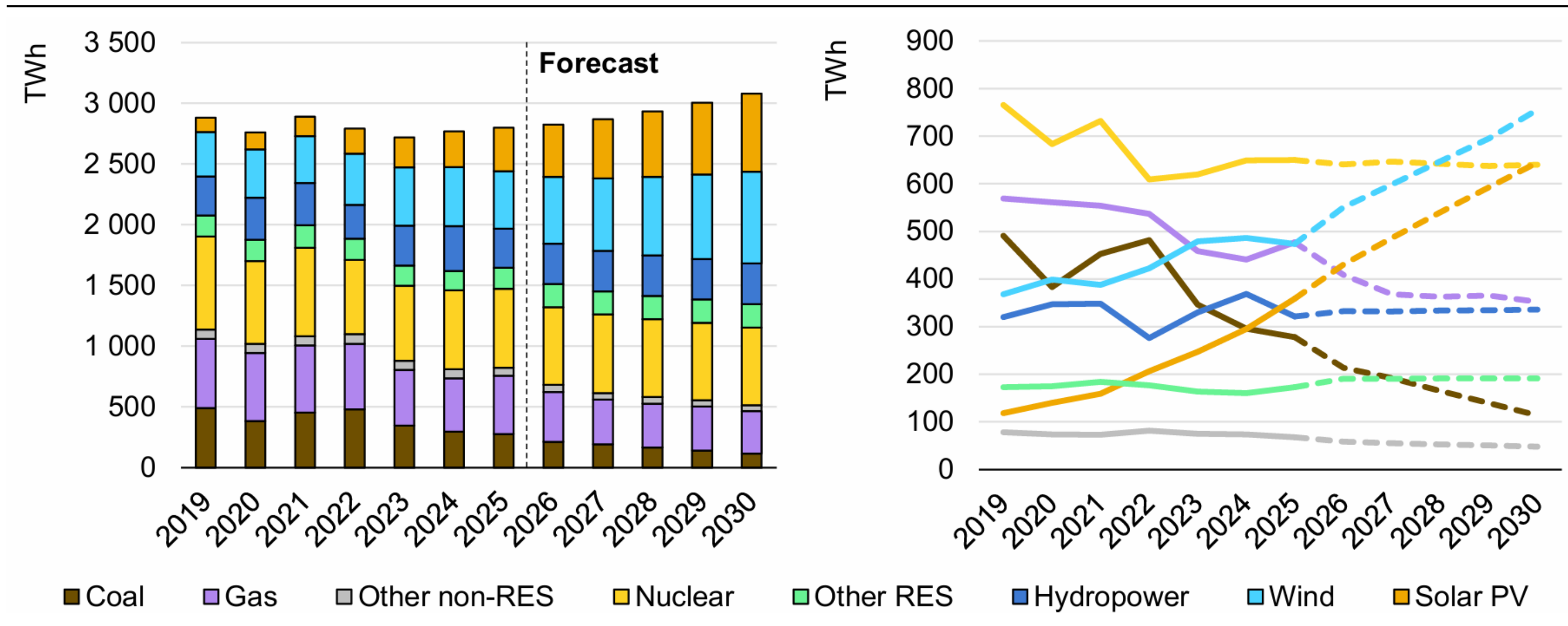
Source: IEA (2026), Electricity 2026, IEA, Paris

# Share of electricity in total final consumption in select countries and regions, 1990-2025



Source: IEA (2026), Electricity 2026, IEA, Paris <https://www.iea.org/reports/electricity-2026>

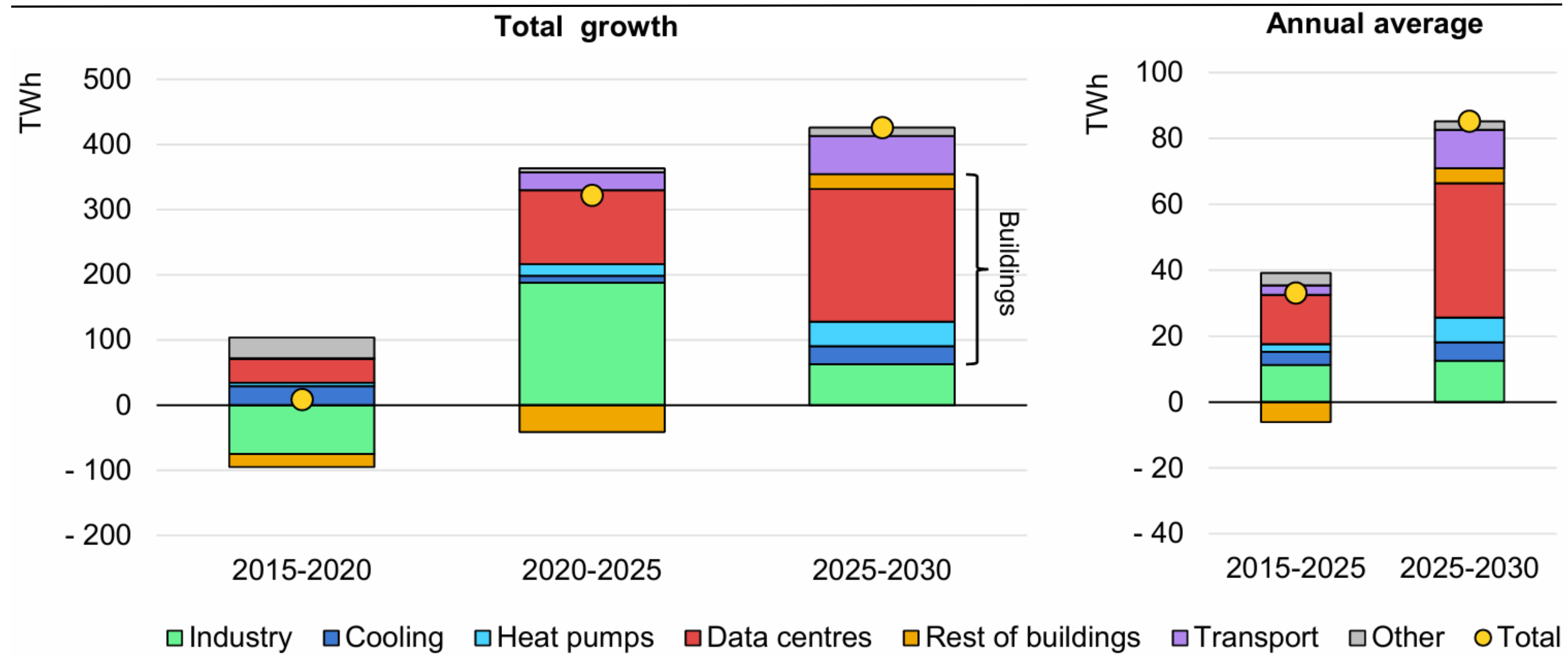
# Electricity generation by source in the European Union, 2019-2030



Source: IEA (2026), Electricity 2026, IEA, Paris <https://www.iea.org/reports/electricity-2026>

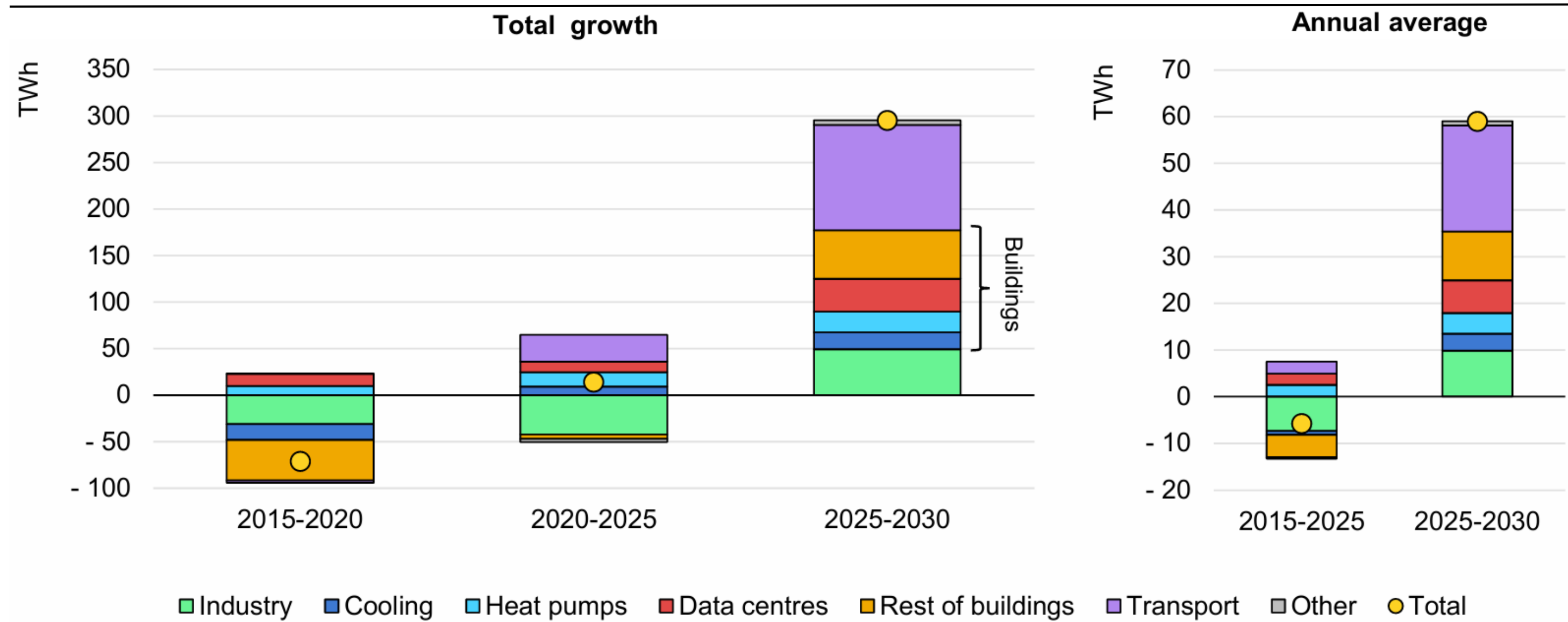
Notes: RES = renewable energy sources. 'Other non-RES' includes oil, waste and other non-renewable sources. 'Other RES' includes geothermal, bioenergy, concentrated solar power (CSP), and ocean energy. Data for 2026-2030 are forecast values.

# Electricity demand growth by sector and end-use in the United States, 2015-2030 - 60% of USA Load Growth



Source: IEA (2026), Electricity 2026, IEA, Paris <https://www.iea.org/reports/electricity-2026>

# Electricity demand growth by sector and end-use in the European Union, 2015-2030 - 12% of EU Load Growth



Source: IEA (2026), Electricity 2026, IEA, Paris <https://www.iea.org/reports/electricity-2026>

# There are four types of Data Centers: hyperscale, colocation, enterprise, and edge



## Hyperscale facilities

owned by companies such as Google, Microsoft, Meta, and Amazon that deploy Internet services and platforms at massive scale.

## Colocation companies

provide wholesale and retail colocation leasing, typically deploying large and very large data centers. Many companies will share server space and capacity in a colocation facility.

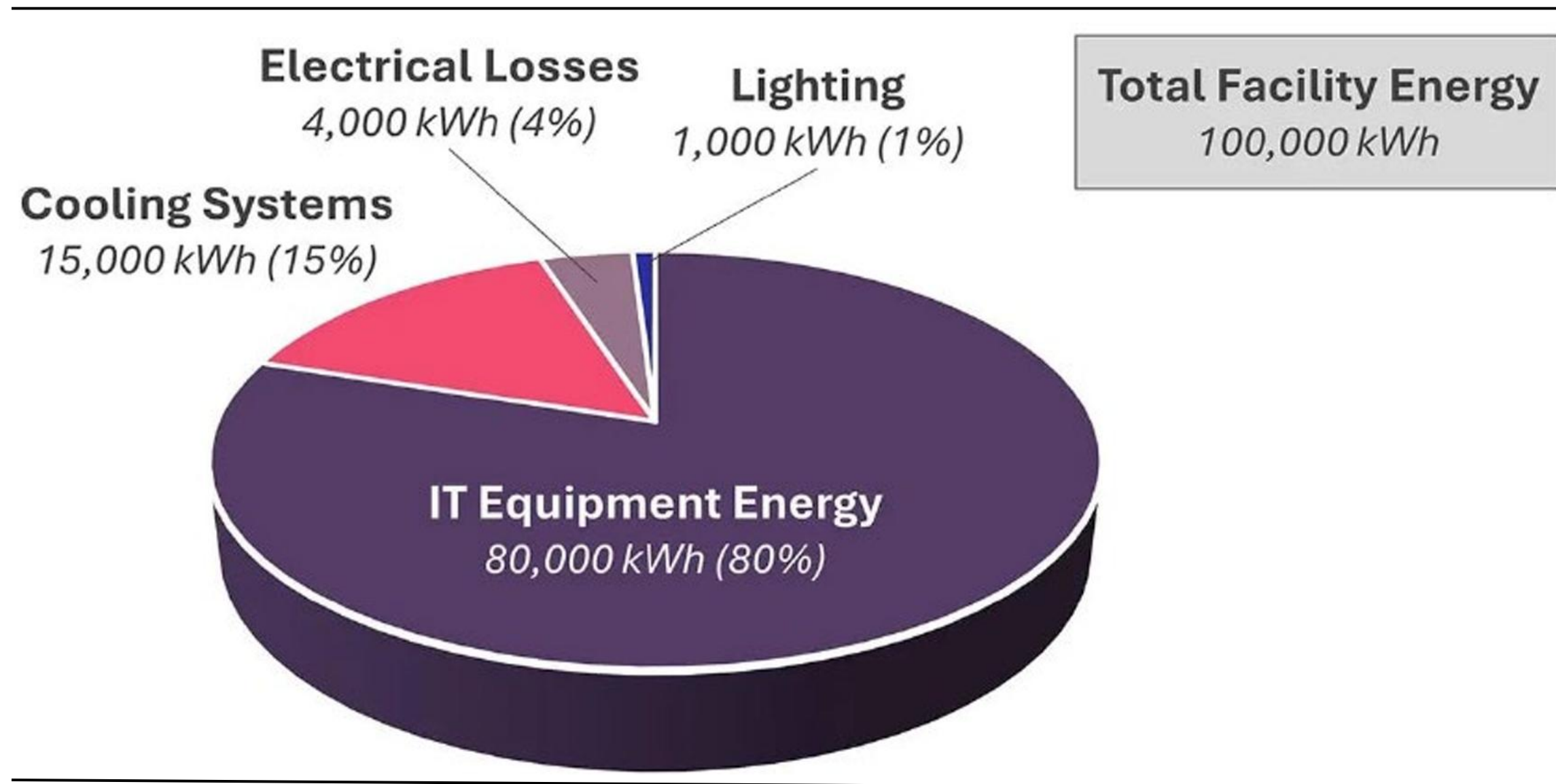
## Enterprise sites

are owned and operated by firms such as banks and phone companies for their internal operations. Their advantage is security and control, but they may have low utilization rates, rigid designs, and long upgrade cycles. Some enterprises are moving less sensitive data to colocation sites.

## Edge centers

serve local Internet of things providers such as Netflix and 5G cell services.

# Distribution of data center energy consumption by component



Source: Nadel, Steven. 2025. Opportunities to Use Energy Efficiency and Demand Flexibility to Reduce Data Center Energy Use and Peak Demand. Washington, DC: ACEEE.

# Savings and useful life for IT equipment and systems in data centres



**Table:** Energy savings by type of efficiency measure (ESm) and lifetime of savings for servers' loads

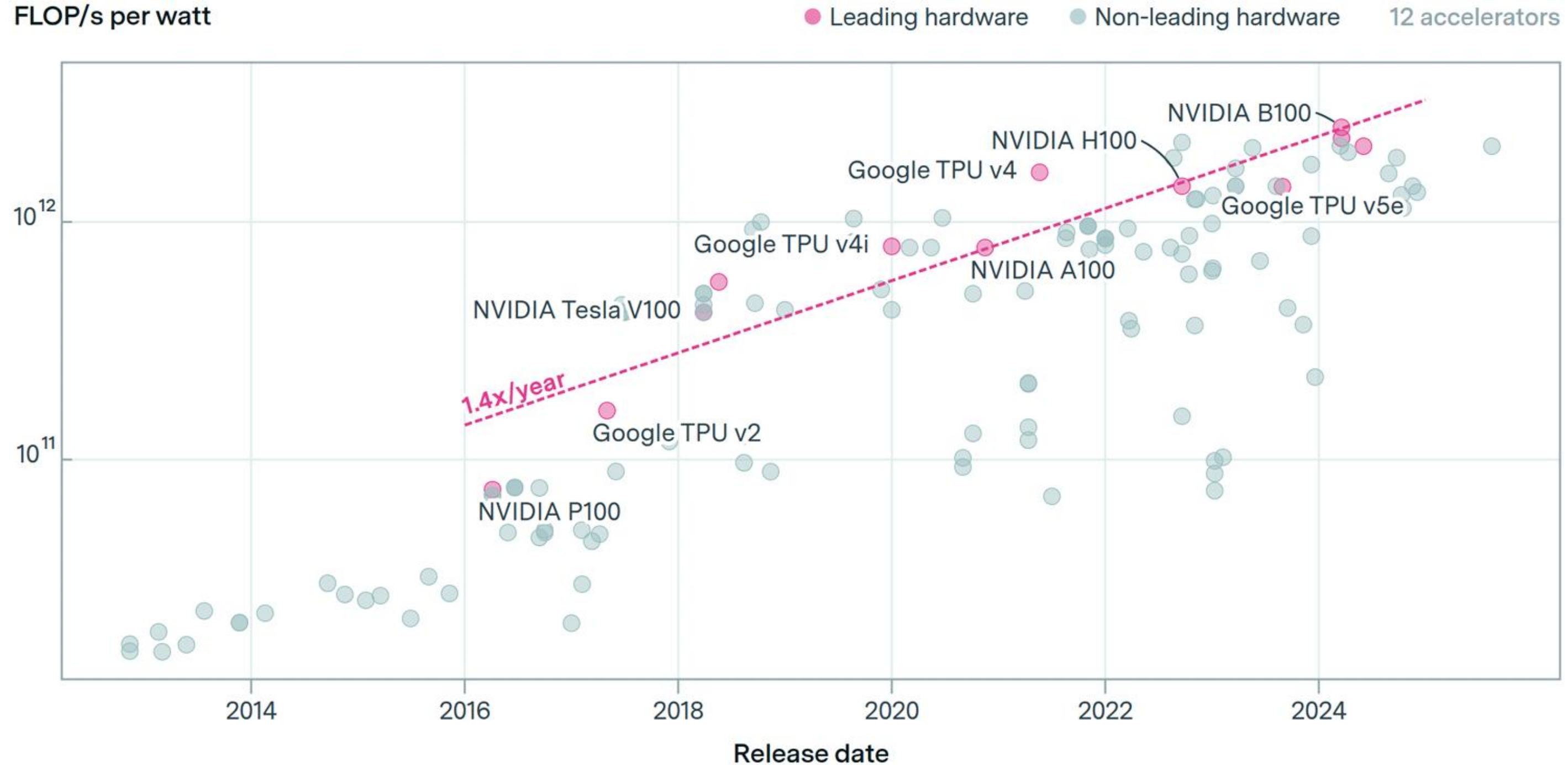
Measure	Load (%)	Lifetime (years)
Server virtualisation and consolidation	20 – 40%	6
Decommissioning obsolete servers	5 – 15%	3
Deployment of energy-efficient server hardware	10 – 25%	5
Intelligent workload scheduling	10 – 30%	4
Activation of power management features	5 – 20%	4
Efficient virtualisation/container platforms	10 – 20%	5
Monitoring and analytics for server energy use	0 – 5%	2

**Source:** Table 23 of the Extended guidance for standardized savings methodologies & indicative values (Deliverable D2.2). StreamSAVE+ Project, Grant No. 101167618.

# Energy efficiency (FLOPs per watt) of leading chips



FLOP - Floating-point Operations Per Second – Log Scale



CC-BY

epoch.ai

# Savings and useful life for IT equipment and systems in data centres



Table: Energy savings by type of efficiency measure (ESm) and lifetime of savings for data storage loads

Measure	Load (%)	Lifetime (years)
Data management optimisation	5 – 15%	4
Storage tiering and energy-aware systems	10 – 20%	5
Modernisation of storage hardware	10 – 20%	5

Source: Table 24 of the Extended guidance for standardized savings methodologies & indicative values (Deliverable D2.2). StreamSAVE+ Project, Grant No. 101167618.

# Savings and useful life for IT equipment and systems in data centres



**Table: Energy savings by type of efficiency measure (ESm) and lifetime of savings for network loads**

Measure	Load (%)	Lifetime (years)
Efficient network design and topology optimisation	5 – 15%	4
Energy-efficient network equipment	5 – 15%	5
Intelligent port and link management	5 – 10%	4
Monitoring network device consumption	0 – 5%	2

**Source:** Table 25 of the Extended guidance for standardized savings methodologies & indicative values (Deliverable D2.2). StreamSAVE+ Project, Grant No. 101167618.

# Savings and useful life for IT equipment and systems in data centres - Power Usage Effectiveness (PUE)



$$\text{PUE} = \frac{\text{Total Facility Energy}}{\text{IT Equipment Energy}}$$

**Table:** Reference values for Power Usage Effectiveness for different cooling technologies of data centres (PUE)

Technology	Baseline PUE	Improved PUE	Lifetime [years]
Upgrade CRAC/CRAH units to variable-speed systems	> 1.8	1.5 – 1.6	15
Transition to chilled water system with air-side economizers	> 1.6	1.3 – 1.5	15
Implement free cooling (air-side, water-side, TES etc.)	1.6 – 1.8	1.2 – 1.4	15
Deploy liquid cooling (direct-to-chip or immersion)	1.6 – 1.8	1.02 – 1.1	15
Optimize two-phase/passive cooling (e.g., thermosiphons)	1.5 – 1.7	1.1 – 1.3	15
Integrate thermal energy storage (TES) for peak shaving and free cooling	1.6 – 1.8	1.2 – 1.4	15

# Scale & trajectory of electricity demand



## Concentration + peak coincidence

- **AI-focused facilities can draw power comparable to heavy industry**, but are geographically clustered.
- **Grid constraints and connection queues are often driven by a limited number of peak hours** each year.
- In the U.S., data centres' share of system-wide peak demand is projected to rise from ~6% today to ~13% by 2030 (IEA).

## Illustrative stress points

- Interconnection & transmission upgrades take years; compute demand can ramp much faster.
- Reliability requirements (N+1, backup power) shape site selection and resource mix.
- **Some jurisdictions have imposed limits or paused new connections where demand threatens system adequacy.**

# Impacts on the clean energy transition



## Upside potential

- **Catalyses new clean generation** via long-term contracts (PPAs) and co-located projects.
- **Flexible computing** can help absorb variable renewables (shift to low-carbon hours/regions).
- Waste-heat recovery can displace fossil heating in district energy systems.

## Downside risks

- **Tight timelines and reliability needs** can trigger new **fossil capacity** if grids are constrained.
- **If flexibility is unmanaged, loads can coincide with peaks and raise system costs (All people pay!)**
- Water stress and local opposition can slow projects and shift them to less optimal sites.

# Data centres as flexible grid assets



**Flexibility for ~0.1–1% of hours can be enough to integrate planned additions (to 2035) in some systems.**

- **Levers:** workload management (shift/curtail non-urgent jobs), storage dispatch, backup activation, on-site generation.
- **Goal:** reduce grid draw during the small number of “top peak” hours that drive infrastructure needs.
- **Design principle:** keep mission-critical SLAs intact while making a subset of load dispatchable.

## Flexibility stack (concept)

- 1) Measure & forecast (PUE/WUE, load, carbon intensity)
- 2) Software control (shift, throttle, reroute)
- 3) Hardware support (BESS, UPS, cooling controls)
- 4) Market participation (DR, ancillary services, dynamic tariffs)
- 5) Verify outcomes (cost, CO<sub>2</sub>, water, reliability)

# Demand response via flexible computing (example)

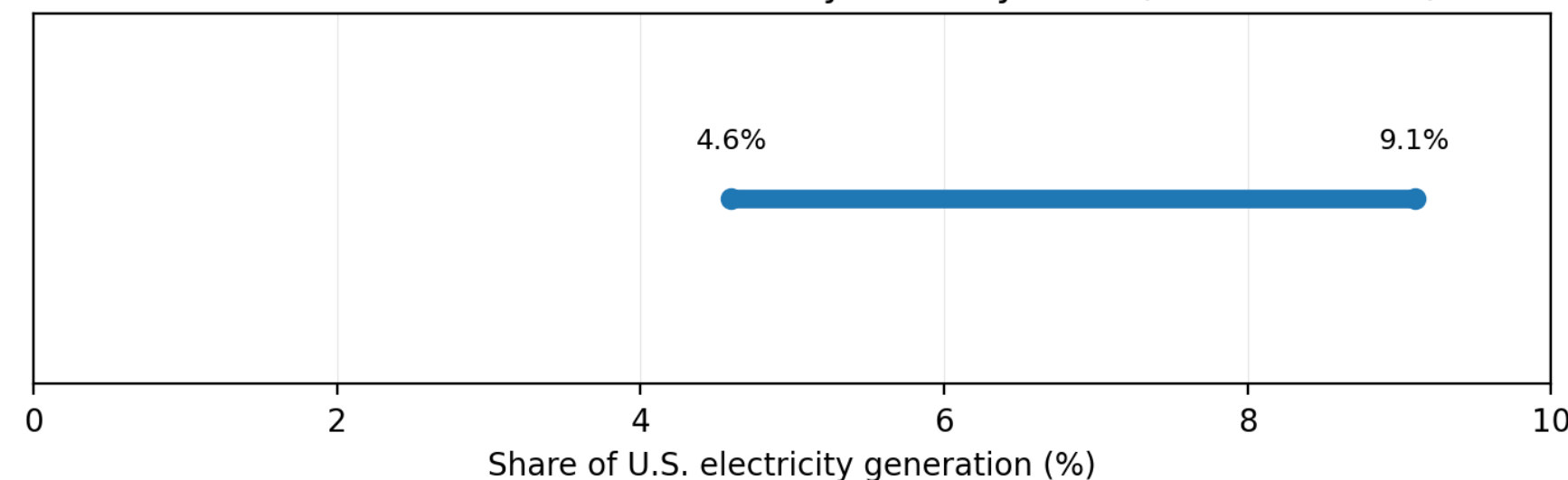


## How a “software-defined” DR event can work

- Grid operator forecasts a local constraint (e.g., extreme weather) and notifies the data centre/utility partner.
- A compute planning system generates hour-by-hour instructions to reduce non-urgent compute at impacted sites.
- Tasks can be rescheduled later or rerouted to other regions when feasible.
- Target: lower local grid stress without adding new hardware or breaking service-level guarantees.

## Flexibility can also be quantified and contracted:

Estimated U.S. data centre electricity share by 2030 (EPRI scenarios)



Range shows uncertainty in future load share — a reason to value flexibility

## Design cues for DR programs

- Clear event triggers & baselines
- Verified curtailment
- Fail-safe opt-outs for critical services

# Demand response via flexible computing (example)



## Why batteries matter

- BESS can reduce peak grid draw and respond faster than diesel-based backup in many scenarios.
- A “grid-interactive” design can deliver curtailment or ancillary services while maintaining uptime.
- Integrating BESS with workload management expands the controllable portion of the load.

Case study (simulation + HIL):

NREL + Verrus validated a 70-MW grid-interactive data centre platform (Vulcan) to test control strategies and grid services.

## Mitigation architecture (typical)



Compute (IT) load



Cooling & facilities



UPS + BESS



Grid interface (DR/ancillary)

# Policy & market design for flexibility



## What makes demand response & load shifting work at scale

Incentives & tariffs	Rules & guardrails	Planning integration
<ul style="list-style-type: none"><li>• <b>Dynamic pricing</b> that reflects congestion/peaks</li><li>• <b>Curtailment payments</b> with verified baselines</li><li>• <b>Performance bonuses</b> for fast response</li></ul>	<ul style="list-style-type: none"><li>• Define eligible flexible workloads</li><li>• Reliability fail-safes and opt-out clauses</li><li>• Emissions accounting to avoid “shifting into coal”</li></ul>	<ul style="list-style-type: none"><li>• <b>Treat dispatchable load as a resource in adequacy planning</b></li><li>• Coordinate with interconnection studies</li><li>• <b>Standardised telemetry &amp; M&amp;V</b></li></ul>

# Clean electricity procurement & co-location



## Approaches

- Long-term PPAs to add new renewables and storage.
- 24/7 carbon-free matching (time- and location-based) to reduce marginal emissions.
- Co-locate data centres with new clean generation to reduce grid congestion and speed delivery.
- Contract for flexibility (curtailment windows) alongside energy supply.

## Example: co-located “clean power + compute” parks

Google-backed Intersect Power projects aim to co-locate gigawatts of data centre capacity with new carbon-free generation to ease grid constraints and reduce timelines.

## Key design questions

- Who pays for transmission upgrades?
- How to share clean output with the wider grid?
- How to ensure additionality & deliverability?

# Efficiency first: reduce watts per unit of compute



## Track the right KPIs

- PUE (Power Usage Effectiveness) — industry average remains ~1.56 in Uptime's 2024 survey.
- WUE (Water Usage Effectiveness) — liters per kWh of facility energy.
- CUE (Carbon Usage Effectiveness) — emissions per IT energy.
- ERF (Energy Reuse Factor) — recovered heat used outside the facility.
- Compute efficiency — utilization, scheduling, and model/hardware efficiency (e.g., performance per watt).

## High-impact levers

- Raise IT utilization (consolidation, autoscaling)
- Reduce overprovisioning (right-size power & cooling)
- Adopt liquid cooling where appropriate for high-density racks
- Tune models (smaller models, quantization, batching, inference optimization)
- Improve fleet efficiency (refresh cycles, circularity, heat reuse)



**Obrigado pela vossa atenção**

**Nuno Quaresma**  
**nquaresma@isr.uc.pt**