European Commission

ENTEC2

Energy Transition Expertise Centre



Support for the preparations of a strategic roadmap for digitalisation and AI in the energy sector

Prepared by **EUREC**, **EURAC**, **RISE**, **AEE Intec**, **WIP**

Authors:

Greg Arrowsmith, arrowsmith@eurec.be

Marco Cozzini, marco.cozzini@eurac.edu

Benedetta Di Costanzo, benedetta.dicostanzo@wip-munich.de

Jannis Haack, jannis.haack@wip-munich.de

Jenny Holgersson, jenny.holgersson@ri.se

Martin Larsson, martin.larsson@ri.se

Mehrzad Lavassani, mehrzad.lavassani@ri.se

Markus Lindahl, markus.lindahl@ri.se

Jeanette Petersson, jeanette.petersson@ri.se

Jonas Petzschmann, jonas.petzschmann@zsw-bw.de

Anna Spoden, spoden@eurec.be

Katharina Strecker, katharina.strecker@zsw-bw.de

Bharath Varsh Rao, bharath-varsh.rao@ait.ac.at

Marcus Vendt, marcus.vendt@ri.se

Mattias Vesterlund, mattias.vesterlund@ri.se

Daniel Weiss, daniel.weiss@dlr.de

Gerald Zotter, g.zotter@aee.at

Reviewers:

Matteo Giacomo Prina, matteogiacomo.prina@eurac.edu

Christoph Rohringer, c.rohringer@aee.at

Steffi Misconel Schreiber, steffi.misconelschreiber@eurac.edu

Mark Stefan, Mark.Stefan@ait.ac.at

Stefan Übermasser, Stefan.Uebermasser@ait.ac.at

Bharath Varsh Rao, bharath-varsh.rao@ait.ac.at

Daniel Weiss, daniel.weiss@dlr.de

Prepared for **the European Commission, Directorate-General for Energy** under contract N° ENER/B4/2024-459/SI2.930521-ENTEC2

EUROPEAN COMMISSION

Directorate-General for Energy
Directorate B - Just Transition, Consumers, Efficiency and Innovation
Unit B.4 - Digitalisation, Competitiveness, Research and Innovation

Contact: Szilvia Bozsoki

E-mail: szilvia.bozsoki@ec.europa.eu

European Commission
B-1049 Brussels

Manuscript completed in May 2026

0.1 Edition

This document has been prepared for the European Commission under contract n° ENER/B4/2024-459/SI2.930521, however, it reflects the views only of the authors, and the European Commission is not liable for any consequence stemming from the reuse of this publication. The European Commission does not guarantee the accuracy of the data included in this document. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

Luxembourg: Publications Office of the European Union, 2026

© European Union, 2026



The reuse policy of European Commission documents is implemented by Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Unless otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

PDF ISBN 978-92-68-39960-6 doi:10.10.2833/3572353 MJ-01-26-034-EN-N

Content

1. Task 1: Analysis of the benefits and costs of digitalisation and use of AI in energy 6

1.1. Introduction 6

1.2. Comparative Cost-Benefit Analysis and Use Case Prioritisation 7

1.3. Policy Recommendations for AI and Digitalisation in Europe's Energy Sector 19

1.4. AI Use Case Summary in the Energy Sector 21

2. Task 2: Market analysis of AI in energy 22

2.1. Use Case 1: AI-Assisted Grid Operation 22

2.1.1. Status Quo & Market Potential 22
2.1.2. Barriers to Adoption 24

2.2. Use Case 2: Demand-Side Response and Flexibility Activation 25

2.2.1. Status Quo & Market Potential 25
2.2.2. Barriers to Adoption 27

2.3. Use Case 3: Generation and Load Forecasting 29

2.3.1. Status Quo & Market Potential 29
2.3.2. Barriers to Adoption 31

2.4. Use Case 4: Anomaly Detection & Predictive Maintenance 32

2.4.1. Status Quo & Market Potential 32
2.4.2. Barriers to Adoption 33

2.5. Use Case 5: AI-Supported Permitting of Energy Infrastructure 34

2.5.1. Status Quo & Market Potential 34
2.5.2. Barriers to Adoption 35

2.6. Policy recommendations 36

2.6.1. Establish a SME-Friendly AI Governance Framework for the Energy Sector 37
2.6.2. Strengthening the Societal Acceptance of AI in Energy through Transparency and Inclusion 38
2.6.3. Launch an EU-wide Benchmark Suite for Evaluating Energy-AI Systems 39
2.6.4. Ensuring Technological Sovereignty in the Adoption of AI in Europe's Energy Sector 40
2.6.5. Towards a Trusted, Interoperable and Privacy-Preserving Data Foundation for AI in Energy 42

3. Task 3: Foundation Models for the Energy Sector 44

3.1. Introduction 44

3.2. Foundation Models and the Energy Sector Use Cases 45

3.3. Foundation Models and Development for the Energy Sector 49

3.3.1. Technical Characteristics of FMs 49
3.3.2. Prerequisites and Resource Considerations 50
3.3.3. Approaches to Developing FM Applications 50
3.3.4. Evaluating FM-Based Energy Applications 53

3.4. Cybersecurity Risks in Foundation Models 54

3.5. Foundation Models Adoption in The Energy Sector 56

3.5.1. Key Areas to Address for FM Integration in the Energy Sector 57
3.5.2. Technical FM Integration Roadmap 59

3.6. Foundation Model Stakeholders 61

3.6.1. Collaboration of Foundation Model Stakeholders 62

3.7. Core Recommendations 64

4. Task 4: Flexibility and Heat Reuse Potential of Data Centre 65

4.1. Executive summary 65

4.1.1. Data centre sector overview 65
4.1.2. Cooling technologies 65
4.1.3. Heat pumps 66
4.1.4. Integration with district heating 66
4.1.5. Economic aspects 66
4.1.6. Flexibility and sector coupling 67
4.1.7. Examples of business cases 67
4.1.8. Recommendations 67

4.2. Introduction 68

4.2.1. Background 68
4.2.2. Objectives 68
4.2.3. Methodology 69

4.3. Data centre sector overview 71

4.3.1. Introduction to data centres 71
4.3.2. Data centre energy use 73
4.3.3. Data centre trends 75
4.3.4. Data centre sector summary data 77

4.4. Overview of data centre cooling technologies 77

4.4.1. Introduction to cooling technologies 78
4.4.2. Air-based cooling 80
4.4.3. Liquid-based cooling 82

4.5. Relevant heat pump technologies 84

4.5.1. Heat pump performance metrics and qualitative dependence on temperature 85

4.6. Overview of commercially available heat pumps 90

4.7. Heat pump types and configurations 91

4.7.1. Heat pump refrigerants 91
4.7.2. Heat pump compressors 94
4.7.3. Heat pump configurations and data centre PUE 94

4.8. Connecting data centres and district heating 97

4.9. Overview of data centres and district heating 97

4.10. Waste heat reuse in the proximity of the data centre 97

4.11. Waste heat reuse for district heating 98

4.12. Overall estimates for data centre waste heat recovery to district heating 101

4.13. Flexibility and sector coupling 102

4.13.1. Flexible operation to support the power grid 102
4.13.2. Sector coupling for data centres 104
4.13.3. Key takeaways from flexibility and sector coupling 106

4.14. Overall techno-economic estimates 107

4.14.1. General considerations on cost estimates 107
4.14.2. Waste heat recovery substation costs 111
4.14.3. Pipeline costs 114
4.14.4. Summary on cost estimates and possible revenue streams 117

4.15. Examples of business cases and best practices 118

4.15.1. Business cases for district heating and other sector coupling initiatives: opportunities and barriers 119
4.15.2. Business cases for flexibility services supporting the power grid 126

4.16. Recommendations 130

4.16.1. Policy background: Heat recovery and flexibility measures from data centres in the EU legislation 130
4.16.2. Recommendations for future actions 132

5. Task 5: Analysis of the Open Public Consultation Results for the Strategic Roadmap 138

1. Task 1: Analysis of the benefits and costs of digitalisation and use of AI in energy

1.1. Introduction

The energy sector, specifically electricity energy vector, is undergoing a transformation driven by decarbonisation, decentralisation, and digitalisation ([1]). Meeting climate and energy security goals, such as Europe's net zero targets, requires integrating large volumes of variable renewable generation and managing supply and demand in complex, data-rich power systems ([2]). Advanced digital technologies, particularly artificial intelligence (AI), are key enablers ([3]). AI can process high-volume data (for example, weather, smart meters, grid sensors) to generate insights and control signals beyond traditional tools ([4]). Utilities can use AI for real-time forecasting and balancing of electricity supply and demand, reducing costs, emissions, and improving grid efficiency and reliability ([5]). On the user side, AI enhances energy efficiency and flexibility, empowering consumers and lowering system peaks ([6]).

Recognising the strategic importance of digitalisation, the European Commission's 2022 Action Plan on "Digitalising the Energy System" emphasises technologies such as IoT, cloud-edge computing, digital twins, and AI ([7]) Major funding initiatives support this trend, including the EU's €200 billion digital and AI investment programme and Horizon Europe projects like TwinEU (€25 million, 75 partners) which develops a digital twin of the European grid ([8]). These reflect strong policy backing for AI-driven energy innovation ([9]). The AI Act, adopted in June 2024, establishes a risk-based regulatory framework to ensure that AI systems in the EU are safe, transparent, and respect fundamental rights. It includes specific requirements for high-risk AI applications, such as those used

([1]) European Technology and Innovation Platform for Smart Networks for Energy Transition (ETIP SNET). Unlocking the Potential of AI and Generative AI in European Smart Grids. 2025. See Executive Summary and Chapter 1, pp. 6–8

([2]) International Energy Agency (IEA). Energy and AI – World Energy Outlook Special Report. 2025. See Chapter 1.4, "Energy for AI and AI for Energy", pp. 37–49.

([3]) U.S. Department of Energy. AI for Energy: Opportunities for a Modern Grid and Clean Energy Economy. April 2024. See Executive Summary and Section 2.

([4]) IEA. *Energy and AI*, Chapter 3, "AI for Energy Optimisation", pp. 110–134

([5]) Luís Vale Cunha. Workshop on Opportunities and Challenges of AI/GenAI for Energy Grids, ETIP SNET / European Commission, 20 September 2024. See "Prominent operational AI/GenAI use cases from a DSO perspective", Slide 4.

([6]) IEA. *Energy and AI*, Chapter 3.5, "AI for Energy End Uses – Buildings", pp. 150–159

([7]) European Commission. *Digitalising the Energy System – EU Action Plan*. 2022. See Policy Overview, Section 1

([8]) ETIP SNET. Unlocking the Potential of AI and Generative AI in European Smart Grids, Chapter 5, "Action List and Roadmap", pp. 47–56

([9]) AI for Energy Report. U.S. Department of Energy and European Commission. 2024. See Sections 12.

in energy infrastructure, covering areas like data quality, human oversight, and cybersecurity to ensure responsible deployment aligned with the EU's sustainability and digital goals ([10]).

However, the digital shift brings challenges: data governance, interoperability, AI transparency, skill shortages in both the energy and AI sectors, and cybersecurity risks. These issues also highlight the need for new competencies to effectively monitor, evaluate and govern AI systems, both within organisations and at the regulatory level. Additionally, AI development and deployment increase ICT energy use, raising power system demand. Policymakers must balance innovation with responsible deployment. The EU's approach, including the AI Act and funding for testbeds and sandboxes, aims to ensure safe and sustainable AI use in the clean energy transition.

The report identifies five key areas where digitalisation and AI are transforming the energy sector. They were selected because they represent the most critical functions where digitalisation and AI can deliver immediate, measurable improvements in performance, reliability and cost efficiency across the energy system. Each addresses a unique and high impact challenge, ranging from managing variability in renewable generation to enhancing asset longevity and empowering consumers, while collectively supporting the transition to a smarter, cleaner and more resilient grid. Their strategic importance is widely recognised in industry and policy roadmaps, making them central to the digital transformation of the energy sector.

Use Case 01: Forecasting Energy Demand, Renewable Generation, and Grid Conditions

Use Case 02: Grid Planning, Operations, and Real Time Control Optimisation

Use Case 03: Predictive Maintenance for Energy Assets.

Use Case 04: Demand Side Management and Intelligent Flexibility Activation

Use Case 05: Tacit and Explicit Knowledge Management

1.2. Comparative Cost-Benefit Analysis and Use Case Prioritisation

The following table presents a comprehensive, comparative analysis of five high-impact AI use cases within the energy sector. It is structured to support technically informed decision-making by outlining domain applications, traditional versus AI-driven methodologies, quantifiable benefits (economic, operational, environmental), strategic prioritisation and key adoption barriers. This table aims to inform infrastructure planners, regulators and digital transformation strategists

([10]) European Parliament & Council. (2024, July 12). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union L 1689.

on the tangible value of deploying AI across the energy value chain. Barriers to adoption vary significantly depending on the context, such as market structure, regulatory environment, data availability and institutional capacity. This section focuses on the most relevant and commonly observed barriers across use cases, rather than attempting to provide an exhaustive list.

Use Case 01	Forecasting of Energy Demand, Renewable Generation and Grid Conditions
Domains of Application	**Electricity demand forecasting** (from building-level to system-wide load predictions). **Renewable generation forecasting** (wind speed & solar irradiance forecasts for wind/PV output). **Grid conditions forecasting** (anticipating congestion, voltage/frequency deviations, ([11],[12]), weather impacts on grid assets).
Traditional vs. AI Methods	**Traditional:** Statistical time-series models (e.g. ARIMA), simple regressions, and physics-based simulations (numerical weather prediction, load profiles) often used with expert rules. **AI:** Machine learning models (recurrent networks and transformer neural networks, ensemble methods like gradient-boosted trees) trained on big data (weather, grid parameters and meta data...) for pattern recognition; hybrid physics-AI models that ingest weather data to improve accuracy beyond rule-based methods ([13]).
Impact and Benefits	Higher forecast accuracy yields cost savings and efficiency gains. e.g. improved wind power forecasts saved utilities to the order of $95 million per year and saved consumers ~$380 million by reducing reserve fuel costs ([14]). Better predictions reduce renewable **curtailment** and emissions by enabling grid operators to optimise scheduling of resources (AI-enhanced forecasts help integrate more wind/solar without over-reliance on fossil backup) ([15]). For instance, AI-based forecasting and integration of renewables can directly cut curtailment and CO2 emissions ([16]). **Reliability gains:** Allows proactive measures for peak demand or extreme events, improving resiliency. Accurate load forecasts reduce blackout risks by ensuring adequate
Prioritisation	This is a high-priority use case. Forecasting underpins planning and real-time operations in a renewable-rich grid. It directly supports energy transition goals by handling variability. Accurate forecasts are foundational for balancing supply and

([11]) E. Heylen, F. Teng, and G. Strbac, “Challenges and opportunities of inertia estimation and forecasting in low-inertia power systems,” *Renewable and Sustainable Energy Reviews*, vol. 147, Art. no. 111176, Sept. 2021, doi: 10.1016/j.rser.2021.111176

([12]) M. Tuo and X. Li, "Machine Learning Assisted Inertia Estimation Using Ambient Measurements," in *IEEE Transactions on Industry Applications*, vol. 59, no. 4, pp. 4893-4903, July-Aug. 2023, doi: 10.1109/TIA.2023.3269732.

([13]) Das, Puja, August Posch, Nathan Barber, Michael Hicks, Kate Duffy, Thomas Vandal, Debjani Singh, Katie van Werkhoven, and Auroop R. Ganguly. “Hybrid physics-AI outperforms numerical weather prediction for extreme precipitation nowcasting.” *npj Climate and Atmospheric Science* 7, no. 1 (18 November 2024): 282. https://doi.org/10.1038/s41612-024-00834-8.

([14]) Krishnamurthy et al., *Bulletin of the American Meteorological Society*, 2022 (as referenced in U.S. DOE Wind Forecast Improvement Project report, 2023).

([15]) IEA, *Energy and AI* Report, p.17 (International Energy Agency, 2023).

([16]) IEA, *Energy and AI* Report, p.17 (International Energy Agency, 2023).

	demand with high-RES penetration, which in turn maintains stability and reduces operating costs ([17]). Grid operators could prioritise AI in forecasting because even modest error reductions translate to significant cost and reliability benefits.
Barriers to Adoption	**Data limitations:** An example of a barrier would be that forecast models need granular weather and consumption data, in many regions, data quality and sensor coverage are insufficient or siloed, hindering AI performance. Rare events (extreme weather) are hard to train for, leaving some uncertainty. Additionally, cybersecurity is better because a potential attack path would be to insert fake data into the forecast models. **Model trust and integration:** Operators may be wary of "black box" AI forecasts. Without transparency or explainability, there is reluctance to replace tried-and-true methods for critical decisions. AI forecasts must integrate with existing energy management systems (EMS) and market rules, some of which presuppose certain deterministic methods. **Regulatory and skill barriers:** In some markets, regulations determine how forecasts are produced or penalise deviations, limiting flexibility to use new AI methods. Utilities also face skill gaps in data science, making it challenging to implement and maintain AI forecasting systems. A regulatory issue is also the access of data that is now becoming more and more relevant for critical systems. Another barrier is that the personnel in regulatory agencies may not be adequately skilled for effective regulation and implementation.
Overcoming barriers	To improve forecasting, data integration and standards are key. European experts recommend developing harmonised data formats and energy data spaces to pool information from weather services, smart meters and grid sensors ([18]). This ensures AI models have the rich, real-time data they need. Hybrid modelling approaches can also build trust, and combining machine learning with physical grid models provides forecasts that are both accurate and explainable. For example, advanced AI-driven weather and demand prediction models allow grid operators to anticipate fluctuations more precisely, reducing renewable curtailment through timely demand shifting or storage use ([19],[20]). Regulators can encourage innovation by aligning incentives: in electricity markets, this might mean rewarding actors (suppliers, aggregators, TSOs) for forecast accuracy or providing sandboxes to test AI forecasting tools without penalty. Building in-house AI capability is equally important; several utilities have created teams blending data scientists and power engineers to tackle data quality issues and tailor AI tools to operations.

Use Case 02	Optimisation for Grid Planning, Operations, and Real-Time Control
Domains of Application	**Grid planning:** Long-term capacity expansion and grid design (transmission expansion, siting of resources) using AI to handle complex scenarios and multi-criteria optimisation. **Grid operations:** Optimal power flow (OPF), state estimation and economic dispatch in daily operations; dynamic grid reconfiguration, voltage control, congestion management with AI support.

([17]) IEA, *Energy and AI* Report – Executive Summary, p.3-4 (International Energy Agency, 2023).

([18]) Energy and AI – World Energy Outlook Special Report, p. 124–131, International Energy Agency, 2025 – provides analysis of AI applications in energy (e.g. forecasting, grid operations, maintenance) and discusses adoption barriers and opportunities.

([19]) Artificial intelligence unlocking a smarter, greener energy future, Report, European Commission DG CNECT, June 2025 – strategic study highlighting AI's role in digital twins, data spaces, and recommendations on standards, infrastructure, and regulation to foster AI in the energy sector.

([20]) *Generative AI and Foundation Models in the EU: Uptake, Opportunities, Challenges, and a Way Forward*, p. 40–41, CEPS (Centre for European Policy Studies), 2025 – includes case studies like Schneider Electric's AI-powered predictive maintenance for renewable assets

	Real-time control: Fast control actions for frequency regulation, inverter controls, demand-response activation, anomaly detection, fault diagnosis and outage management and grid stability (using AI/ML for decision support or autonomous control in microgrids).
Traditional vs. AI Methods	**Traditional:** Mathematical optimisation (linear/non-linear programming for OPF), heuristic and rule-based control algorithms (e.g. droop control, preset operating rules), and human operator expertise for contingency management. These methods often struggle as system complexity grows (combinatorial scenarios, stochastic inputs). **AI:** Advanced optimisation assisted by ML (e.g. deep neural network "surrogate models" to speed up OPF calculations, reinforcement learning for control policies, graph neural networks for grid simulations). AI can learn from vast scenario data to recommend actions faster than iterative solvers, enabling **real-time adaptive control** that traditional tools cannot easily do. Foundational models, pre-trained on diverse and extensive datasets, offer a scaleable way to generalise across tasks and domains, enhancing AI's adaptability in power system applications.
Impact and Benefits	**Efficiency gains:** AI-aided grid optimisation can reduce losses and operational costs. *Globally, grid inefficiencies (losses, suboptimal dispatch) led to ~1 gigaton CO_2 emissions in 2018*, improved optimisation directly translates to huge emissions cuts ([21]). AI can find solutions that minimise transmission losses and generation costs more effectively than static heuristics. In power generation, AI-based operations & maintenance optimisation could save up to $110 billion annually by 2035 in fuel and O&M costs (Widespread AI adoption scenario) ([22]). AI-optimised dispatch also reduces curtailment of renewables by better matching supply-demand in real time ([23]). **Reliability and grid stability:** AI can run contingency analyses and grid reconfigurations much faster, helping prevent outages. For example, AI-based grid control can react to faults or fluctuating renewables in milliseconds, maintaining frequency and voltage within safe limits better than manual control. This **enhances resilience** to disturbances and extreme events (faster restoration, fewer cascading failures). **Integration of new resources:** By handling the complexity of distributed energy resources (DER), electric vehicle (EV) charging, and storage coordination, AI optimisers enable higher renewable and distributed energy uptake without sacrificing stability. This supports climate goals by allowing the grid to accommodate more clean energy and flexibility resources than it could with manual or static controls.
Prioritisation	**Highest priority:** AI based grid planing is central to managing a decarbonised, complex grid. Optimising grid operations in real-time is crucial for grid stability and efficient use of assets. With rising renewable penetration and an ageing infrastructure, operators see AI-based optimisation as key to avoiding bottlenecks and blackouts. In the EU, digital and AI solutions for grid operations are deemed *crucial* to achieve 2050 climate neutrality while keeping the lights on ([24]). The ability of AI to solve problems faster than human operators or traditional solvers makes this a strategic focus area for investment (to handle the "duck curve", EV charging surges, etc.).
Barriers to Adoption	**Safety-critical trust:** Grid control is high stakes; regulators and operators require assurance that AI will not introduce instability. The power system must deliver reliable power 24/7, so any AI decision-making is subject to rigorous validation and often a "human-in-the-loop" requirement. There is reluctance to cede control to

([21]) Puech et al., Optimal Power Grid Operations with AI Foundation Models (IBM Research preprint), p.1 (2024).

([22]) IEA, *Energy and AI* Special Report, Chapter 3 (International Energy Agency, 2023).

([23]) IEA, *Energy and AI* Special Report, Chapter 3 (International Energy Agency, 2023).

([24]) European Commission, *Role of AI in Transforming Energy Grids* (presentation, 2024); Puech et al. (IBM), 2024.

	algorithms without extensive testing and explainability (a black-box optimisation that could threaten reliability will face approval hurdles) ([25]). **Legacy systems and interoperability:** Utilities often run decades-old Supervisory Control and Data Acquisition (SCADA) and EMS systems, integrating modern AI tools with these can be technically complex. Legacy IT and siloed data systems impede seamless AI deployment ([26]). Upgrading infrastructure and ensuring data interoperability (across transmission, distribution, market platforms) is costly and slow. First, it is essential to standardise data formats, communication protocols and digital interfaces to enable seamless interoperability across platforms and actors. EU-level initiatives such as the Common European Energy Data Space ([27]) should be expanded and adequately resourced to support secure, efficient and transparent data sharing. Initiatives such as the Interoperability Network for the Energy Transition (int:net) ([28]) are essential to bring together all relevant stakeholders from across the European energy sector to jointly develop, test and deploy interoperable energy services. **Regulatory and liability issues:** There may be a lack of clear regulatory frameworks for AI-driven decision-making in grid operations (for example: who is liable if an AI-caused action leads to an outage?). Utility commissions can be conservative, requiring proven benefit and risk mitigation. Compliance standards, such as N-1 security, must be met by any AI solution, sometimes requiring extensive safety margins that diminish AI's theoretical efficiency gains. These safety margins, combined with complex regulatory requirements, can also discourage participation and experimentation by small and medium-sized enterprises (SMEs), which often lack the resources to navigate heavy compliance burdens. **Cybersecurity:** Adding AI and digital control increases potential cyber-attack surfaces. AI systems themselves could be targets for manipulation, so robust cyber protection and fail-safes are needed, which is an additional hurdle and cost for adoption. **Compatibility with physics-based constraints:** One key barrier to adopting AI in grid stability applications is ensuring compatibility with physical laws and system constraints. Traditional grid operations are governed by well-established physical models (e.g. power flow equations, inertia dynamics), and AI systems must follow these principles to be safe and trustworthy. However, many AI models, especially black-box or general-purpose ones, lack explicit integration of these physical rules, risking unrealistic or unsafe outputs. Addressing this requires hybrid approaches that embed domain knowledge into AI models (e.g. physics-informed ML), or rigorous validation against physics-based simulations to ensure reliable, explainable performance in real-world grid operations.
Overcoming barriers	A phased and human-in-the-loop deployment strategy is recommended. Rather than full automation, many utilities have started by using AI as a decision-support assistant in control rooms. This approach addresses trust issues, for instance, RTE (France's TSO) introduced AI assistants under the project "ORIGAMI" to help operators manage growing grid complexity, while explicitly prioritising ethics and acceptability in design ([29]). Operators remain in charge, but benefit from AI suggestions for actions like reconfiguring network topology or optimal load shedding, thus building confidence in the algorithms. In parallel, developing real-time validation tools and industry standards is crucial. Governments and industry groups (e.g. IEEE, CIGRÉ) are working on benchmark tests and certification processes for grid AI systems ([30]). Such standards will reassure operators that an

([25]) U.S. DOE, *AI for Energy: Opportunities for a Modern Grid* Report, p.11 (Department of Energy, 2024).

([26]) Centric Consulting – Common Barriers to AI Implementation (2023).

([27]) Common European Energy Data Space - European Commission

([28]) Home - Int:net

([29]) https://www.entsoe.eu/technopedia/techsheets/artificial-intelligence-ai/

([30]) Energy and AI https://iea.blob.core.windows.net/assets/601eaec9-ba91-4623-819b-4ded331ec9e8/EnergyandAI.pdf

	AI will always comply with system constraints. Economic barriers can be tackled by aligning financial incentives with digital optimisation: regulators can reward utilities for improved network capacity use or reliability achieved through AI, rather than only traditional capital expansion. This shifts the business case in favour of AI investments.

Use Case 03	Predictive Maintenance for Energy Assets and Infrastructure
Domains of Application	**Generation assets:** Predictive maintenance for power plants (gas/coal turbines, solar farms, wind turbines); using sensor data and AI to predict equipment failures (e.g. detecting vibration anomalies in turbines or performance degradation in PV panels). **Grid infrastructure:** Transmission and distribution equipment monitoring, transformers, circuit breakers, power lines (using IoT sensors, drones, thermal imaging analysed by AI to predict faults or needed repairs). **Distributed assets:** Monitoring of batteries, EV chargers, smart meters, and other DER for proactive maintenance and improved uptime.
Traditional vs. AI Methods	**Traditional:** Time-based scheduled maintenance or reactive “run-to-failure” repairs. Some use of basic condition monitoring (SCADA alarms, manual inspections). Traditional statistical models or threshold rules (if vibration > X, alert) provide limited prediction, and often maintenance relies on human inspections and manufacturer recommendations (e.g. replace after N years regardless of condition). **AI:** Machine-learning-driven **predictive analytics** on sensor streams (temperature, pressure, vibration, oil quality, etc.) to detect subtle patterns before failure. AI models (e.g. anomaly detection, random forest or neural networks) learn normal vs. faulty behaviour, estimating remaining useful life for components. Advanced approaches include computer vision for equipment inspection (AI image analysis of power lines or plant equipment) and deep learning that fuses multiple sensor inputs to predict complex failure modes.
Impact and Benefits	**Reduced unplanned downtime and cost:** AI-driven predictive maintenance cuts unexpected outages and costly emergency repairs. Studies show **maintenance cost reductions of ~25–30%**, breakdown incidents cut by ~70%, and downtime reduced ~40% on average when moving from reactive to predictive maintenance (using AI/IoT) ([31]). This improves service continuity and saves operational expenses. For example, McKinsey analysis finds advanced predictive maintenance can lower maintenance costs by up to 30% and increase equipment uptime by 5–15%, saving hundreds of billions of Euros across industries by 2025 ([32]). Avoiding just one major transformer failure or power plant forced outage can save millions in damage and lost revenue. **Extended asset life and efficiency:** By fixing issues at early stages and optimising maintenance schedules, assets experience less wear-and-tear. AI maintenance strategies can extend component lifetimes (e.g. avoid catastrophic failures that ruin equipment) and sustain higher efficiency (a well-maintained turbine operates closer to peak efficiency, saving fuel). This deferment of capital expenditures (replacing equipment less often) has significant economic value. **Safety and reliability:** Predictive maintenance reduces the risk of accidents (identifying unsafe device conditions before they lead to fires or explosions) and helps maintain grid reliability (fewer sudden equipment outages mean lower risk of cascading failures). Customers benefit from more reliable energy delivery.

([31]) U.S. Department of Energy. (2021). Predictive Maintenance Applications for Energy Infrastructure. Technical Report

([32]) McKinsey & Co., The Economic Value of Advanced Predictive Maintenance, p.2 (2021).

Prioritisation	This use case is **highly valuable** but often seen as a medium-term priority. It directly impacts O&M costs and reliability, which are important, though perhaps less immediately visible to the public than integration of renewables. Still, for meeting energy transition and resilience targets, **predictive maintenance is a key enabler;** it keeps critical infrastructure running smoothly as assets harder (e.g. running older transformers at higher loads). Many utilities prioritise it after tackling forecasting and operational optimisations. Notably, the U.S. DOE highlights shifting from time-based to AI-driven maintenance as a significant opportunity for cost savings and avoiding outages ([33]). In ageing grids, this is increasingly prioritised to prevent infrastructure failures that could derail transition efforts (e.g. transformer failures delaying renewable projects).
Barriers to Adoption	**Data and sensor availability:** Implementing predictive maintenance requires extensive sensor deployments and data integration. **Older infrastructure may lack IoT sensors** and retrofitting them is costly. Data may exist in silos (different formats from different OEMs) and be of variable quality. Ensuring enough historical failure data to train models is also a challenge; many critical failures are rare, so the AI may initially struggle with limited examples. **Organisational and economic barriers:** Utilities often need to justify the business case for new AI-driven maintenance programs; the benefits (avoided failures) are somewhat invisible until a failure occurs. Regulated utilities might not readily recover investments in AI systems without regulatory support. Additionally, maintenance crews need training to trust and use AI outputs; there can be resistance if the AI suggests actions that conflict with traditional practices or if it's seen as a threat to jobs. Moreover, companies may be reluctant to share data due to concerns about disclosing sensitive business information or trade secrets, highlighting the need for robust privacy-enhancing techniques. **Data ownership and privacy:** Equipment manufacturers (OEMs) sometimes hold the detailed equipment data or proprietary algorithms. Sharing data with AI vendors or across departments raises intellectual property and cybersecurity concerns ([34]). Utilities must also ensure that predictive analytics do not compromise customer data privacy (for customer-owned assets like smart meters or EV chargers). **Accuracy and risk:** False positives (AI predicting failure that doesn't happen) could lead to unnecessary maintenance and costs, while false negatives (missed predictions) could cause a false sense of security. Ensuring the AI model is robust and continuously improved is an ongoing challenge, especially as equipment ages or new failure modes emerge. Conservative culture in engineering means AI recommendations will be double-checked until proven, which can slow adoption. In addition, these systems critically rely on domain-specific data, which is often proprietary and not easily accessible. This lack of accessible data limits the ability to build generalised models and restricts knowledge sharing across organisations, reducing the overall scaleability and effectiveness of AI in industrial environments.
Overcoming barriers	To tackle data challenges, companies are investing in sensor networks and data platforms. Retrofitting assets with sensors (e.g., vibration, temperature or oil quality sensors for transformers) significantly increases the data available for AI models. Utilities are also creating central data lakes to consolidate operational data. Ensuring data quality may involve automated cleansing and adopting common standards, like the EU's common information model for asset data. Combining domain expertise with AI is essential; equipment engineers and data scientists work together to create hybrid models, incorporating physics-based insights into machine learning for more reliable predictions. Knowledge sharing within industry forums accelerates adoption, allowing utilities to replicate successful AI applications. AI integration into existing workflows involves updating processes, such as linking AI predictions to maintenance schedules and inventory management. With leadership support, these strategies overcome barriers and make predictive maintenance a norm.

([33]) U.S. DOE, *AI for Energy* Report, Section 2.3.2, p.17 (2024).

([34]) U.S. DOE, *AI for Energy* Report, p.18 (2024).

Use Case 04	Demand Side Management and Intelligent Flexibility Activation
Domains of Application	**Demand-response programs:** AI-driven systems that adjust or shift consumer loads (industrial, commercial, residential) during peak periods or when supply is low, e.g. smart thermostats lowering HVAC use, EV chargers pausing or timing charging, industrial processes modulating. **Distributed energy resource (DER) coordination:** Using AI to orchestrate thousands of distributed assets (home batteries, smart appliances, EVs, heat pumps) as a "virtual power plant" that provides peak shaving, frequency support, and reserves by intelligently modulating demand and supply at the edge. **Energy efficiency and dynamic pricing:** AI analysing usage patterns to recommend energy-saving actions or optimal rate plans, and automating consumption to respond to real-time electricity prices or carbon intensity signals (e.g. running flexible loads when renewable generation is abundant).
Traditional vs. AI Methods	**Traditional:** Demand side management (DSM) historically relies on static techniques; time-of-use tariffs, manually triggered demand-response events (with advance notification), or direct load control contracts for big industrial users (often with simple on/off controls). These are often rule-based and use coarse heuristics (e.g. "reduce X MW when frequency < threshold"). Customer participation is sometimes manual (responding to price signals or phone/text alerts). **AI: Intelligent DSM** platforms using reinforcement learning and predictive control to continuously adjust loads. AI predicts both grid conditions and customer behaviour, then optimises control of devices in real-time. Methods include multi-agent systems (devices acting autonomously within an AI framework), and optimisation algorithms that solve for thousands of endpoints (e.g. neural network-based control of many smart thermostats to maintain grid frequency). AI can also personalise incentives for consumers (using algorithms to predict which customers are likely to respond, and how, improving targeting of demand-response).
Impact and Benefits	**Peak load reduction and cost savings:** By activating flexibility, AI-optimised DSM can significantly cut peak demand. Studies in Europe show that by 2030, **up to 60 GW of peak supply needs could be met by flexible demand** instead of new power plants, saving roughly €2.7 billion per year in generation capacity investments ([35]). For grid operators, shaving peaks means less need for activating expensive power plants or emergency imports, leading to direct operational cost savings. Consumers participating can also save on bills or earn incentives, while overall wholesale prices are lowered by peak shaving (benefiting all ratepayers). **Integration of renewables and emissions reduction:** DSM improves the alignment of demand with renewable supply (e.g. shifting EV charging to midday solar peak or windy nights). This **reduces curtailment of renewables and reliance on fossil backup**, cutting emissions. A full deployment of demand flexibility in the EU could save an estimated 37.5 Mt of CO_2 in 2030 (an ~8% reduction of power sector emissions) by avoiding fossil generation during peak times ([36]). Flexibility also enhances grid resilience to renewable variability, supporting higher RES targets (by absorbing excess generation or reducing load during lulls). **Grid resilience and reliability:** Active demand management can alleviate grid stress during contingencies, e.g. AI can reduce loads in a region to prevent an overload or help balance the grid during a supply shortfall, acting as an automatic safety net. This fast adaptive load control adds to grid stability (in some cases providing ancillary services like frequency regulation from loads). It also enables consumers to contribute to reliability, creating a more decentralised and robust energy system.

([35]) DNV & smartEn, Demand-Side Flexibility in the EU: Quantification of Benefits in 2030, p.5-6 (2022).

([36]) DNV & smartEn, Demand-Side Flexibility in the EU (2022), p.8.

	Consumer benefits: Beyond incentives and bill savings, AI-driven DSM can provide better comfort management (smart control pre-cools or pre-heats buildings to ride through demand response events with minimal impact) and enable new energy services. It effectively monetises flexibility as a new value stream for end users (turning distributed assets into revenue-generating resources).
Prioritisation	**High priority**, particularly as renewable penetration grows. **Flexibility on the demand side is often cited as the "missing piece"** for a cost-efficient energy transition. By actively managing consumption, we avoid overbuilding generation or storage for peak conditions. Many policymakers and system operators view intelligent DSM as critical for reaching net-zero goals and ensuring reliability with high renewables. For example, studies show a large fraction of grid investment needs can be avoided through flexibility ([37]). In practical terms, enabling demand-side flexibility is a priority for grid operators facing capacity constraints (it's faster and often cheaper to implement demand-response programs than to build new infrastructure). It is also aligned with empowering consumers in the energy transition, a political priority in many regions. As a result, this use case ranks high in strategic roadmaps for modern smart grids and virtual power plant development.
Barriers to Adoption	**Technical implementation:** Widespread DSM requires advanced metering infrastructure (smart meters) and control devices in homes and businesses. Many regions still lack full deployment of these enabling technologies. Additionally, coordinating thousands or millions of devices in real time is data- and compute-intensive; latency or communication failures can hinder responsiveness. Ensuring interoperability across many device types and manufacturers (so they can all respond to AI signals) is an ongoing technical challenge. **Regulatory and market barriers:** Inflexible market structures and regulations often impede demand-side participation. In some areas, **dynamic pricing or demand-response programs are not yet fully enabled by regulators**, or aggregators face hurdles to enter markets. Policy may not remunerate flexibility adequately, making it less attractive. There are also instances of conflicting incentives, e.g. consumer protection rules that limit control over customer devices, or peak tariffs that are too shallow to motivate change. Policy reforms are needed to value and reward flexibility for AI-driven DSM to scale. **Consumer acceptance and social factors:** Reliable demand-response depends on getting consumers or automated systems to adjust energy use based on grid needs. Privacy concerns arise when AI monitors detailed home energy data, and some users fear losing control or comfort. Without clear safeguards and communication, adoption may be slow as people opt out over perceived risks or inconvenience. Building trust is essential. Programs must respect user preferences and operate within agreed limits. Since social behaviour is hard to predict, user-focused design is key. Changing habits in energy use is a major challenge. Targeted outreach can build awareness and trust, while a gradual move to dynamic pricing helps align behaviour with grid conditions. Privacy-enhancing methods like data anonymisation and on-device processing can ease concerns and support responsible AI use. **Economic considerations:** There can be upfront costs for consumers to enable flexibility (smart appliances, controllers) and for utilities to implement AI platforms. If not subsidised or incentivised, this slows uptake. Furthermore, the benefits of DSM (avoided costs) are dispersed across the system, whereas the entity investing in enabling it (e.g. an aggregator or the consumer) might not capture enough value; this misalignment can be an economic barrier to adoption unless market designs or policies address it.
Overcoming barriers	Regulatory and market reforms are crucial for enabling solutions in demand-side flexibility. Governments must implement frameworks that empower aggregators and reward flexibility. For example, simplifying market entry for demand-response providers and ensuring they can compete with traditional generators fosters

([37]) DNV & smartEn, Demand-Side Flexibility in the EU: Quantification of Benefits in 2030, p.5-6 (2022).

	innovation. The EU's upcoming Energy Data Space initiative aims to create a common architecture for secure energy data sharing, enabling participation in flexibility programs ([38]). By standardising APIs and consent mechanisms, data fragmentation can be reduced. Open standards for devices (e.g., OpenADR for demand response signalling) ensure interoperability, allowing a single AI platform to control many device types. Edge computing and IoT solutions also help manage local loads with low latency, improving reliability and reducing communication bottlenecks ([39]). Incentive programs and dynamic pricing, such as time-of-use tariffs or real-time pricing, give consumers financial reasons to shift consumption. AI-enabled demand-response can access revenue from capacity markets or flexibility tenders, improving the business case. Governments can support technology rollout by subsidising smart thermostats or water heaters to enable aggregate flexibility. Consumer engagement is critical; transparent communication and the ability for users to override AI preferences maintain consumer control. Education campaigns showing that comfort can be maintained while supporting grid stability are key. Aligning innovation with supportive policy and user-centric design will reduce barriers to intelligent demand-side management.

Use Case 05	Tacit and Explicit Knowledge Management from Operational Data and Expert Insights
Domains of Application	**Operational data analytics:** By capturing explicit knowledge from sources like operational logs, incident reports, and maintenance records, AI can analyse unstructured data to uncover patterns. such as common causes of outages or optimal troubleshooting steps. Beyond organising this information for future use, AI can also support and facilitate human decision-making by presenting relevant insights in context. **Expert knowledge capture:** Using AI (including large language models (LLMs) and knowledge graphs) to codify **tacit knowledge** of experienced grid operators, engineers and field technicians. This includes insights from decades of experience (e.g. "tribal knowledge" about peculiar grid behaviour or maintenance tricks) that were traditionally passed on informally. **Decision support and training:** Intelligent assistants that can answer questions like "what steps to take for a transformer relay trip" by drawing on both manuals (explicit knowledge) and past expert resolutions. AI-driven training tools, such as scenario simulators and Q&A chatbots for operational procedures, draw on historical knowledge to mentor new staff. By personalising and contextualising this information for each individual, AI makes onboarding and skill transfer more interactive, accessible and effective.
Traditional vs. AI Methods	**Traditional:** Rely on human memory, mentorship and documentation such as manuals and procedure guides. Knowledge management systems (if any) are often basic document repositories or wikis that require manual updates. Tacit knowledge transfer happens via on-the-job training, shadowing experienced workers, or not at all, leading to loss of critical know-how when individuals retire or leave. Searches in traditional document systems can be time-consuming, and many lessons learned remain buried in dense reports or verbal history. **AI: Knowledge mining** with traditional NLP (Using rules or statistical methods instead of machine learning, they include approaches like topic modelling, sentiment analysis and named entity recognition. These methods are reliable, interpretable and efficient, and they can improve AI systems by providing structured input or serving as validation tools.), e.g. AI algorithms (including NLP-based transformers and domain-specific LLMs) that can ingest maintenance logs, emails, engineering reports and automatically extract key lessons or decision rules. LLMs with **retrieval augmentation** are used to create conversational assistants:

([38]) European Commission. (n.d.). *Digital strategy*. Retrieved from https://digital-strategy.ec.europa.eu

([39]) Centre for European Policy Studies (CEPS). (n.d.). *CEPS*. Retrieved from https://cdn.ceps.eu

	an operator can query in natural language ("Have we seen a similar fault before?") and the AI, having been fed decades of reports and sensor data, can respond with synthesised relevant knowledge ([40]). AI can also form knowledge graphs linking concepts (equipment→ failure modes→ solutions) from both data and human expert input, providing a navigable web of insights. Importantly, generative AI can capture tacit knowledge by analysing conversations, forums or conducting structured interviews; turning veteran insights into potentially personalised explicit guidelines.
Impact and Benefits	**Preservation of expert knowledge:** Prevents loss of critical know-how as an ageing workforce retires. The energy sector faces a "Silver Tsunami" of retirements, e.g. ~50% of senior utility workers could retire by 2027 ([41]). Capturing their tacit knowledge via AI systems can save millions: industry studies estimate Fortune 500 firms lose ~$31.5 billion annually by failing to share knowledge, a figure expected to double by 2030 in expertise-intensive sectors ([42]). By systematising expert insights, AI knowledge systems ensure continuity of operations and safety standards even as personnel change. **Improved operational efficiency and decision-making:** Engineers and control room operators can retrieve answers or historical precedents in seconds using an AI assistant, rather than flipping through binders or calling a retired colleague. This speeds up troubleshooting, reducing downtime of equipment and outages. Field case: companies report **38% faster problem resolution and 27% fewer critical errors** after implementing AI knowledge hubs for operational support ([43]). Consistent, data-backed guidance also leads to more optimised decisions (e.g. choosing the fix that has worked best historically), saving costs by avoiding trial-and-error or missteps. **Enhanced training and innovation:** New hires climb the learning curve faster; some utilities saw ~46% reduction in technical onboarding time with AI-assisted knowledge systems ([44]). Interactive simulators or Q&A bots can teach staff using real incidents from the past, preserving not just the "what" but the "why" of best practices. In addition, by analysing a large corpus of operational data, AI might identify patterns unnoticed by individuals; spurring process improvements or innovation (for example, discovering that a particular sequence of grid operations avoids a known issue, and formalising that as a standard practice). Overall, organisational learning becomes continuous and data driven.
Prioritisation	This use case is **strategically important** in the long run. The reliability of the energy system depends on human expertise; thus, managing knowledge is crucial for resilience. It is a **medium priority** currently; many organisations recognise the risk of knowledge loss and the value of AI here, but efforts are in early stages compared to forecasting or DER integration. However, as workforce transitions accelerate, this is rising in priority. Ensuring that decades of hard-earned operational wisdom inform future AI-driven grids (and not lost in filing cabinets) aligns with digitalisation goals. Also, the complexity of smart grids means operators will increasingly need AI decision support; making effective knowledge management foundational for **safe AI adoption** (the AI tools are only as good as the institutional knowledge they're built on) ([45]). Thus, forward-looking utilities are starting to invest in this area as a backbone for digital transformation and operational excellence.

([40]) U.S. DOE, *AI for Energy* Report, Section 3.3.3, p.29-30 (2024) – discussing use of LLMs with retrieval to augment knowledge bases.

([41]) Siegel, Evan – *Capturing Tacit Knowledge from the Great Retirement*, eGain blog (2025), citing BLS and Deloitte figures.

([42]) Siegel, Evan – *Capturing Tacit Knowledge from the Great Retirement*, eGain blog (2025), citing BLS and Deloitte figures.

([43]) eGain case study data, as reported in Siegel (2025).

([44]) eGain case study data, as reported in Siegel (2025).

([45]) U.S. DOE, *AI for Energy* Report, Section 3.3.3, p.29-30 (2024) – discussing use of LLMs with retrieval to augment knowledge bases.

Barriers to Adoption	**Cultural and change management:** Capturing tacit knowledge can be challenging; experts may be reluctant or find it hard to articulate their know-how. There can be trust issues: employees might worry that sharing knowledge makes them replaceable or they might distrust an AI "assistant"'s advice. Overcoming this requires a strong knowledge-sharing culture and change management so staff see AI as a tool, not a threat. Additionally, training staff to use and contribute to these tools can be a barrier. **Data preparation and quality:** A lot of valuable information resides in unstructured form (handwritten notes, disparate databases, old PDFs). Considerable effort is needed to digitise and curate this into a usable form for AI. There's also the challenge of context; tacit knowledge often relies on context that AI might miss. If not carefully curated, an AI tool may misinterpret information or give generic answers. Ensuring the **knowledge base is comprehensive, up-to-date and context-rich** is an ongoing barrier (it requires continuous human oversight and data maintenance). **AI accuracy and trustworthiness:** Present AI (especially LLMs) can hallucinate or give incorrect answers if the knowledge base is incomplete or queries are ambiguous. In a safety-critical industry, any incorrect recommendation can erode trust in the system. Therefore, implementing robust **validation and feedback loops** is essential, which adds complexity. Additionally, training on domain-specific or potentially proprietary data helps reduce risk. Moreover, sensitive information (grid security details, etc.) must be handled carefully; deploying such AI often raises **cybersecurity and privacy** concerns in sharing internal documents with AI platforms. **Lack of formal processes:** Traditionally, many energy companies haven't invested in formal knowledge management and data governance. This organisational gap itself is a barrier: successful AI adoption here requires interdisciplinary effort (IT, HR, operations) to build processes for capturing and using knowledge. It may be unclear who "owns" this initiative. Without management focus and resources, even the best AI tool will falter due to lack of content or use.
Overcoming barriers	A multi-pronged approach is essential for effective knowledge management (KM) using digital tools and AI. Companies should first establish a clear KM strategy addressing specific challenges, such as an ageing workforce or data overload. This plan should have executive support and incentives for participation. To capture tacit knowledge, utilities are implementing mentorship programmes and after-action reviews. AI tools, like natural language processing, help manage explicit knowledge by converting unstructured text into searchable data. Centralised knowledge bases, organised with consistent metadata, improve accessibility. AI-driven software can also auto-classify documents, and AI assistants can query knowledge bases in natural language. For example, Austria's TSO APG deployed an AI chatbot to allow staff to query internal documentation and regulations ([46]). Cultural shifts, such as rewarding knowledge sharing and designating knowledge champions, further support KM. Metrics like reduced troubleshooting time can demonstrate ROI. Collaboration through international bodies, like GO15, enables the sharing of best practices and anonymised data for AI training ([47]). Finally, investing in IT infrastructure and clear data governance policies ensures accessibility, privacy and security in KM systems.

([46]) *Artificial Intelligence Technopedia Entry*, ENTSO-E website (accessed July 2025) – outlines AI applications and challenges for TSOs, including data complexity, integration, and skills gaps (ENTSO-E, 2023).

([47]) *Energy and AI – World Energy Outlook Special Report*, p. 124–131, International Energy Agency, 2025 – provides analysis of AI applications in energy (e.g. forecasting, grid operations, maintenance) and discusses adoption barriers and opportunities.

1.3. Policy Recommendations for AI and Digitalisation in Europe's Energy Sector

Pillar	Action Area	Key Recommendations
Governance and Regulatory Frameworks	Establish AI governance and regulatory clarity	Create an EU-wide AI task force, clarify AI Act/GDPR rules, expand sandboxes.
	Future-proof energy regulations	Update regulations for AI control/trading, mandate oversight, and introduce certification ([48]).
Data Access and Quality	Create a unified European energy data space	Standardise formats/APIs, manage access rights, ensure GDPR-compliant protocols ([49]).
	Enhance data quality and transparency	Use common quality metrics, encourage synthetic datasets, ensure accurate AI inputs ([50]).
Infrastructure and Interoperability	Accelerate grid digitalisation and IT infrastructure	Deploy IoT, sensors, cloud/edge computing to modernise grid and support AI.
	Ensure interoperability via open standards	Adopt common standards/interfaces for grid systems, MIMs for AI tool compatibility.
	Coordinate on digital twin development	Transmission system operator / distribution system operator collaboration on digital twins with shared models and interoperability ([51]).
Cybersecurity and Resilience	Embed cybersecurity in energy digitalisation	Apply NIS2 ([52]) to AI in energy, implement zero-trust and AI-specific security standards.
	Enhance system resilience to cyber threats	Conduct risk assessments, drills, and implement real-time AI monitoring ([53]).

([48]) Lakshminarayana, S., Bessa, R., Vale da Cunha, L., & Steenmeijer, D. "Unlocking the Potential of AI and Generative AI in European Smart Grids: A Strategic Position Paper and Guide for Action." European Technology & Innovation Platform Smart Networks for Energy Transition, 2025.

([49]) International Energy Agency (2025) – "Energy and AI – World Energy Outlook Special Report." IEA, Paris, April 2025.

([50]) Lakshminarayana, S., Bessa, R., Vale da Cunha, L., & Steenmeijer, D. "Unlocking the Potential of AI and Generative AI in European Smart Grids: A Strategic Position Paper and Guide for Action." European Technology & Innovation Platform Smart Networks for Energy Transition, 2025.

([51]) TSO-DSO Challenges & Opportunities for the Digital EU Electricity System." Joint report by the European Network of Transmission System Operators and the EU Distribution System Operators Entity, March 2025.

([52]) Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) (Text with EEA relevance)Text with EEA relevance. https://eur-lex.europa.eu/eli/dir/2022/2555

([53]) Lakshminarayana, S., Bessa, R., Vale da Cunha, L., & Steenmeijer, D. "Unlocking the Potential of AI and Generative AI in European Smart Grids: A Strategic Position Paper and Guide for Action." European Technology & Innovation Platform Smart Networks for Energy Transition, 2025.

Pillar	Action Area	Key Recommendations
Ethics, Trust and Consumer Engagement	Mandate transparency and fairness in AI applications	Enforce explainable AI, fairness audits, and redress mechanisms.
	Build consumer trust and participation	Engage public via demos, education campaigns and participatory events.
	Empower consumers with smart tools and control	Promote apps and devices for real-time usage, integrate with demand response.
	Ensure inclusion and digital literacy	Monitor algorithmic bias, ensure equitable tech access, and enhance literacy ([54]).
Innovation, Skills and Investment	Boost R&D and open innovation	Increase funding via Horizon Europe and pilots, support open-source AI models ([55])
	Develop AI and energy sector skills	Reform education, offer re-skilling, and embed AI in engineering training.
	Support startups and digital SME ecosystems	Grant/test access for startups, streamline procurement, co-fund scaled deployment.
	Invest in digital infrastructure upgrades	Modernise legacy systems, secure data platforms, and invest in lagging regions.
Alignment with EU Goals and Long-Term Vision	Align AI with climate and sustainability goals	Evaluate AI's emissions impact, incentivise green AI and report CO_2 metrics.
	Promote ethical, human-centric AI in energy	Implement EU AI Act with focus on ethics, create sector-specific guidelines ([56]).
	Lead international digital energy standards	Shape global standards, ensure cross-border AI energy system integration.

([54]) Lakshminarayana, S., Bessa, R., Vale da Cunha, L., & Steenmeijer, D. "Unlocking the Potential of AI and Generative AI in European Smart Grids: A Strategic Position Paper and Guide for Action." European Technology & Innovation Platform Smart Networks for Energy Transition, 2025.

([55]) International Energy Agency (2025) – "Energy and AI – World Energy Outlook Special Report." IEA, Paris, April 2025.

([56]) Lakshminarayana, S., Bessa, R., Vale da Cunha, L., & Steenmeijer, D. "Unlocking the Potential of AI and Generative AI in European Smart Grids: A Strategic Position Paper and Guide for Action." European Technology & Innovation Platform Smart Networks for Energy Transition, 2025.

1.4. AI Use Case Summary in the Energy Sector

Use Case	Domains of Application	Key Benefits	Priority Level	Key Barriers	Overcoming Barriers
1. Forecasting Energy Demand, Renewable Generation and Grid Conditions	Forecasting electricity demand, renewable output (wind/solar) and grid conditions	Higher forecast accuracy, cost and emissions reduction, improved reliability	High	Data limitations, model trust, regulatory and skill gaps	Harmonised data formats, hybrid AI-physics models, regulatory sandboxes, AI-skilled teams in utilities
2. Grid Planning, Operations and Real Time Control Optimisation	Grid planning, operations (OPF, voltage control) and real-time control	Efficiency gains, enhanced reliability, higher DER/RES integration	Highest	Trust in AI for control, legacy systems, regulatory uncertainty, cybersecurity	Human-in-the-loop strategies, phased AI deployment, interoperability standards, benchmark testing for AI safety
3. Predictive Maintenance for Energy Assets	Monitoring and maintenance of generation, transmission and DER assets	Reduced downtime and costs, extended asset life, improved safety	Medium-High	Sensor/data gaps, organisational resistance, accuracy/trust, data ownership	IoT retrofits, data lakes, hybrid physics-ML models, cultural change and domain-data science collaboration
4. Demand-Side Management and Intelligent Flexibility Activation	Smart demand-response, DER coordination, dynamic pricing	Peak load reduction, cost and emissions savings, improved grid resilience	High	Tech readiness, regulatory/market hurdles, consumer acceptance, cost-benefit misalignment	Dynamic pricing, aggregator incentives, edge computing, open standards (e.g. OpenADR), public engagement campaigns
5. Tacit and Explicit Knowledge Management	Operational data analysis, expert knowledge capture, AI decision support	Knowledge preservation, improved decision-making, faster training	Medium	Cultural resistance, unstructured data, AI accuracy, lack of formal processes	AI chatbots, NLP for document mining, structured mentorship capture, central knowledge bases, clear KM strategy

2. Task 2: Market analysis of AI in energy

2.1. Use Case 1: AI-Assisted Grid Operation

The increasing complexity of Europe's power grids, driven by renewables, decentralisation, digitalisation and evolving regulatory frameworks, requires new tools to support real-time operation. Artificial intelligence offers promising capabilities to enhance situational awareness, optimise decisions, and strengthen system resilience at both transmission and distribution levels. Fragmented regulations, new tariff structures, benchmarking obligations and compliance requirements add more layers of decision-making complexity.

2.1.1. Status Quo & Market Potential

From a methodological perspective, artificial intelligence offers a growing toolbox for power system operation. Techniques such as reinforcement learning and graph-based neural networks are increasingly applied to learn control policies for topology switching (changing the configuration of the grid by opening or closing switches to reroute power flow), corrective redispatch (adjusting generator outputs to relieve congestion), and voltage regulation (maintaining voltage levels within acceptable limits across the grid). These methods enable AI agents to make context-sensitive, real-time decisions based on grid topology, load flow data, and system constraints.

A prominent platform for benchmarking and validating these techniques is the Learning to Run a Power Network (L2RPN) challenge series, developed in collaboration with LF Energy ([1]), a Linux Foundation initiative dedicated to accelerating the energy transition through open-source collaboration. L2RPN is built on Grid2Op, an open-source simulation environment designed for safe AI experimentation in grid operation. Together, these platforms provide high-fidelity testbeds where academic and industrial stakeholders can evaluate AI agents in simulated but realistic environments.

These platforms are open-source, offering transparency and auditability that strengthen cybersecurity and accelerate innovation. For TSOs and DSOs, this openness supports interoperability and vendor independence. However, adopting open source also requires robust supply chain security and governance. LF Energy addresses these challenges by providing concrete guidance on secure development practices ([2]): all LF Energy projects must meet the OpenSSF Best

([1]) https://lfenergy.org/projects/grid2op/

([2]) The Linux Foundation. (2023). *Cybersecurity in Energy Infrastructure: The Value of Open Source Software*. LF Energy & Open Source Security Foundation

Practices Badge requirements, use Software Bills of Materials (SBOMs) to ensure transparency in third-party components, adopt the Secure Supply Chain Consumption Framework (S2C2F) to vet dependencies, and implement continuous integration/continuous deployment (CI/CD) pipelines to streamline secure updates. Developer training and active community collaboration further reinforce security and resilience. Additionally, the Apache 2.0 licence allows broad rights to modify, redistribute and commercialise; it also ensures attribution and licence continuity, enabling utilities and vendors to build trusted, proprietary solutions on top of open, secure foundations.

LF Energy's broader portfolio includes projects such as PowSyBl (Power System Blocks), an open framework for electricity grid modelling and analysis, and OperatorFabric, a modular platform for control room operations. Another initiative is GridFM, an LF Energy project that provides an open-source framework to enable foundation models for the power grid. By promoting open interfaces, transparency and community-driven development, LF Energy addresses key barriers to AI deployment, such as interoperability, explainability and modular integration, while fostering a global ecosystem that includes TSOs, DSOs, researchers and vendors. AI methods are shown to be 100–400 times faster than traditional AC power flow solvers, enabling the screening of millions of N−k contingency scenarios in minutes instead of days. This massive speedup is crucial for real-time grid security, where timely decisions can prevent cascading failures. However, their accuracy in identifying critical congestion varies with system size and contingency severity. While data-driven models achieve high recall in small or uncongested systems, their performance drops notably in detecting congested scenarios in larger grids. This highlights their value for rapid pre-screening, but also the need for validation of critical cases using traditional AC solvers to ensure grid security. Further research is needed to enhance the reliability and generalisability of these models under realistic and high-stress operating conditions.

In parallel with research advances, industry adoption is progressing steadily. Several European transmission and distribution system operators are piloting AI applications in operational environments. TSOs such as RTE (France) and TenneT (Germany/Netherlands) or DSOs like Alliander are testing AI-based tools for alarm filtering, overload forecasting and remedial action optimisation in control rooms. These systems typically run in shadow mode alongside conventional tools and are supervised by human operators. AI applications are being trialed in initiatives like GNN4GC ([3]), AI4REALNET ([4]) and AI-EFFECT ([5]). These pilots are usually supported by national R&D funds or Horizon Europe programmes.

([3]) Hosseini, I. and Dietz, T., 2024. GNN4GC- Graph Neural Networks for Grid Control. e+ i Elektrotechnik und Informationstechnik, 141(7), pp.464-467.

([4]) https://ai4realnet.eu/

([5]) https://ai-effect.eu/

2.1.2. Barriers to Adoption

Despite its promise, several challenges hinder the widespread deployment of AI in grid operation and planning. Models that perform well in simulations often struggle in real-world environments where data may be noisy, incomplete or inconsistent. This challenge is particularly pronounced in low-voltage networks with limited observability ([6]). High-quality real-world datasets remain scarce, and access is often restricted due to privacy, regulatory or commercial concerns. Datasets for grid operation must provide executable grid models with realistic boundary conditions. In research and benchmarking, simplified network models and synthetic datasets are widely used to ensure comparability across methods, since access to detailed real-world grid data is often restricted ([7, 8, 9]). These synthetic datasets typically include realistic values for loads, generation capacities and network constraints, and some variants are designed to provide particularly challenging conditions for testing algorithms. While highly useful for innovation and reproducibility, such benchmarks remain abstractions and often omit details that are essential for industry-grade operations.

Explainability remains another major barrier. Operators require confidence in AI decisions, particularly for safety-critical tasks. Explainable AI approaches are actively researched, but standardised frameworks for explainability and human trust calibration are still lacking ([10,11]).These gaps limit user acceptance, particularly among operational staff responsible for real-time decisions under regulatory oversight.

Organisational readiness varies widely across Europe. Many DSOs, especially smaller ones, lack the IT infrastructure, data integration pipelines and trained personnel necessary for robust AI deployment. The lack of uniform digital maturity across EU member states presents a persistent obstacle to the scalable adoption of AI in grid operations. While large utilities are advancing rapidly, smaller operators often remain in early digitalisation stages, and the sector as a whole lacks consistent benchmarking for AI readiness.

A further structural barrier is the fragmented and uncoordinated development of AI solutions across the sector. To address this, a structured roadmap for AI innovation in power systems has been proposed, aligned with technology

([6]) Cihangir Martin, N., Lūsis, P., van den Akker, W., & Liverani, S. (2024). *Active distribution network management in low voltage grids* [Technology paper]. European Distribution System Operators (E.DSO)

([7]) Babaeinejadsarookolaee, S., Birchfield, A., Christie, R.D., Coffrin, C., DeMarco, C., Diao, R., Ferris, M., Fliscounakis, S., Greene, S., Huang, R. and Josz, C., 2019. The power grid library for benchmarking ac optimal power flow algorithms. *arXiv preprint arXiv:1908.02788*

([8]) https://grid2op.readthedocs.io/en/latest/available_envs.html#description-of-some-environments

([9]) Barsali, S. (2014). Benchmark systems for network integration of renewable and distributed energy resources.

([10]) Tabas, D., & Zhang, B. (2022). Computationally efficient safe reinforcement learning for power systems. In 2022 American Control Conference (ACC) (pp. 3303-3310). IEEE

([11]) Marot, A., Donnot, B., Chaouache, K., Kelly, A., Huang, Q., Hossain, R. R., & Cremer, J. L. (2022). Learning to run a power network with trust. Electric Power Systems Research, 212, 108487.

readiness levels (TRLs) ([12]). It begins with foundational research and prototyping (TRL 1–2), moves to demonstrators in public-private consortia (TRL 3–5), and culminates in operational validation in testing facilities and operator labs (TRL 6–9). Core enablers include standardised data interfaces, model versioning, retraining pipelines and mechanisms for integrating human feedback. The roadmap also notes that AI development in critical infrastructure is rarely linear. "Switchbacks" between TRLs are common when models fail under operational constraints or must adapt to new regulatory demands. Platforms like Grid2Op are therefore essential for low-risk, high-fidelity sandboxing during early-stage validation.

Finally, regulatory uncertainties persist. While the EU AI Act lays out a horizontal framework for high-risk AI systems, energy-sector-specific certification or validation processes are not yet established. This legal ambiguity combined with cybersecurity and data governance concerns contributes to hesitance in procuring and operationalising AI solutions.

2.2. Use Case 2: Demand-Side Response and Flexibility Activation

The shift to a decentralised, renewable-based energy system necessitates a more flexible approach to balancing supply and demand. Demand-side flexibility is crucial to unlocking this potential, but its full-scale deployment requires automation, scaleability and responsiveness to local grid conditions. AI technologies provide the analytical intelligence to optimise decentralised loads in real time, enabling active consumer participation and system-level efficiency. This use case explores how AI facilitates the activation and orchestration of distributed flexibility assets.

2.2.1. Status Quo & Market Potential

The technical and economic potential of demand-side flexibility has been extensively quantified in EU policy and industry studies. SmartEN estimates that full-scale DSR deployment across Europe could yield between **€11.1 and €29.1 billion in annual savings** on investment needs between 2023 and 2030 ([13]). National-focused studies reflect similar patterns of cost savings and grid

([12]) Cremer, J.L., Kelly, A., Bessa, R.J., Subasic, M., Papadopoulos, P.N., Young, S., Sagar, A. and Marot, A., 2024, October. A Pioneering Roadmap for ML-Driven Algorithmic Advancements in Electrical Networks. In *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)* (pp. 1-5). IEEE.

([13]) SmartEN, *The Benefits of Demand-Side Flexibility by 2030* (2022) – EU-wide

investment deferral at the member state level ([14] [15]). AI plays a key role in enabling this by allowing faster, more granular, and more reliable activation of distributed resources.

Relevant flexibility assets include residential and commercial heat pumps, electric vehicles (EVs) and their chargers, stationary batteries, HVAC systems, industrial demand management units and smart appliances. EVs in particular offer significant latent capacity. The EY–Eurelectric report forecasts that by 2030, up to 114 TWh (approximately 4% of Europe's overall electricity demand) of EV battery capacity could be harnessed for grid flexibility through smart charging and vehicle-to-grid (V2G) systems ([16]).

AI is recognised in international and European publications as having high potential for optimising energy systems, and also for electric vehicle transport and intelligent building management. It can enhance operational efficiency, reduce emissions and manage complexity across increasingly decentralised and dynamic infrastructure ([17, 18, 19]). A key factor for unlocking the business case of decentralised flexibility is the introduction of dynamic grid tariffs. These tariffs designed to reflect real-time or locational grid conditions create economic incentives that AI systems can respond to automatically, optimising consumer loads for both cost and system efficiency. However, research also shows that uncoordinated demand response can trigger so-called avalanche effects (sudden, synchronised load) ([20]) which can be mitigated through dynamically updated control signals, enabling more effective integration of renewables and system cost savings.While AI-driven flexibility services can already be viable under existing market conditions, the introduction of dynamic tariffs would significantly enhance the business case by enabling more precise value stacking and better alignment with local grid needs. Dynamic grid tariffs require DSOs to detect and communicate local grid conditions making DSR linked to the

([14]) Agora Energiewende und Forschungsstelle für Energiewirtschaft e. V. (2023): Haushaltsnahe Flexibilitäten nutzen. Wie Elektrofahrzeuge, Wärmepumpen und Co. die Stromkosten für alle senken können.

([15]) Neon Neue Energieökonomik im Auftrag des ZVEI e. V. (2024): Mehrwert dezentraler Flexibilität. Oder: Was kostet die verschleppte Flexibilisierung von Wärmepumpen, Elektroautos und Heimspeichern? Kurzstudie vom 14. März 2024. Autoren: Anselm Eicke, Lion Hirth, Jonathan Mühlenpfordt

([16]) EY und Eurelectric (2025): Plugging into potential: unleashing the untapped flexibility of EVs. Why smart charging is essential for a more sustainable and resilient energy system.

([17]) BDEW. (2020). Künstliche Intelligenz in der Energiewirtschaft – Status quo, Potenziale und Herausforderungen. Online available: https://www.bdew.de/media/documents/BDEW_KI_LAUNCH_2406_1ADiAzP.pdf

([18]) Daniel, Claus, Gehin, Jess C., Laurin-Kovitz, Kirsten, Morreale, Bryan, Stevens, Rick, Tumas, William, Anitescu, Mihai, Poczatek, Alec, Siegel, Andrew, Som, Sibendu, Vilim, Richard, Grout, Ray, Kroposki, Benjamin, Yue, Meng, Rose, Kelly, Al Rashdan, Ahmad, Ritter, Christopher, Balaprakash, Prasanna, ... Dietrich, Emily M. (2024). Advanced Research Directions on AI for Energy. https://doi.org/10.2172/2340139

([19]) IEA (2025), Energy and AI, IEA, Paris https://www.iea.org/reports/energy-and-ai, Licence: CC BY 4.0

([20]) Kühnbach, M., Stute, J., & Klingler, A. L. (2021). Impacts of avalanche effects of price-optimized electric vehicle charging-Does demand response make it worse?. Energy Strategy Reviews, 34, 100608.

digitalisation of grid operations, as further explored in Use Case 1 (**AI-Assisted Grid Operation)**.

Scientific literature on AI-supported DSR is highly advanced. It includes reinforcement-learning agents for real-time EV charging, multi-agent coordination of home energy systems, and hybrid optimisation of heat pumps and storage systems ([21]). Furthermore, AI helps solve a fundamental challenge in battery storage systems where the state of charge cannot be measured directly because it is an internal property. Estimation must rely on indirect signals such as current, voltage and temperature, which are often noisy and affected by operating conditions. Traditional methods like Kalman filters typically achieve a normalised root mean square error of around 2.59 %. In comparison, deep learning approaches have reduced this error to just 0.10 % ([22]). In industry, leading aggregators and energy service providers are already deploying AI in real-world flexibility applications, particularly in the domains of EV charging, building energy management and decentralised storage. For example, Tibber, active in the northern European countries and Germany, leverages AI for real-time control of heat pumps and EVs based on spot market prices and grid signals. GridBeyond provides an AI-powered platform for dynamic optimisation of batteries and industrial loads, participating in balancing and capacity markets in Ireland and mainland Europe. Similarly, Octopus Energy applies AI via its Kraken platform to manage distributed residential loads and EVs in several European countries. These use cases demonstrate that AI-driven flexibility orchestration has already progressed into the early commercial phase in digitally advanced regions, even as other markets remain at a pilot or preparatory stage.

2.2.2. Barriers to Adoption

Despite the potential, deploying AI-based DSR faces significant hurdles, some general to flexibility markets, others specific to AI applications. A key issue is the lack of harmonised frameworks for baseline calculation, performance measurement and compensation in aggregated flexibility services[5]. Legacy tariff systems remain prevalent across the EU and often fail to provide the short-term, locational price signals needed for AI systems to optimise flexibility in real time. As Europex underlines, this results in a **systemic lack of incentives** for demand-side resources to participate in flexibility markets, particularly at the distribution level ([23]). The slow rollout of dynamic, cost-reflective grid tariffs therefore represents not only a regulatory barrier, but also a **missed economic enabler** for unlocking the full value of flexibility.

([21]) Shahid, A., Ahmadiahangar, R., Rosin, A., Blinov, A., Korõtko, T., & Vinnikov, D. (2025). Leveraging the machine learning techniques for demand-side flexibility–A comprehensive review. Electric Power Systems Research, 238, 111185.

([22]) Zafar, M. H., Khan, N. M., Abou Houran, M., Mansoor, M., Akhtar, N., & Sanfilippo, F. (2024). A novel hybrid deep learning model for accurate state of charge estimation of Li-Ion batteries for electric vehicles under high and low temperature. *Energy*, *292*, 130584.

([23]) Europex – Association of European Energy Exchanges (2024): Response to ACER public consultation on prioritising the removal of barriers to electricity demand response. Brussels, 2 February 2024.

The implementation of such tariff systems requires that DSOs gain visibility into low-voltage grid conditions and acquire the capability to manage real-time communication with decentralised resources. This, in turn, demands substantial investment in metering infrastructure, ICT systems and advanced analytics. As explored further in Use Case 1, the maturity of these capabilities varies significantly across Member States.

EU and national authorities, together with system operators, should strengthen energy data management by establishing secure, privacy-compliant registries that map available flexibility resources such as smart appliances, storage systems ([24]). For demand-side response, the authors are not aware of dedicated benchmarking initiatives comparable to those established for grid operation. Instead, research and development activities rely on isolated open datasets, such as electric vehicle charging behavior or operational data from heat pumps, which are mostly collected by OEMs or private consumers. Broader **market data**, such as order book information from wholesale markets, does exist but is typically **not openly available** (e.g. from EPEX). As a result, while fragmented datasets are accessible, no coordinated, open-access benchmarking initiative currently exists that would allow reproducible evaluation and systematic comparison of flexibility activation methods. These registries, combined with the regular publication of detailed national and EU-level statistics, will provide the high-quality, standardised data needed for AI systems to effectively optimise. Eurelectric also underlines that enabling AI-based control of EV flexibility requires standardised data protocols and open-access APIs, which are not yet widely deployed ([25]).

Trust and explainability are critical barriers. Operators and regulators are reluctant to adopt “black-box” AI solutions without transparency or clear audit trails. Several regulatory frameworks, including the General Data Protection Regulation (GDPR), the EU Artificial Intelligence Act and the EU Cyber Resilience Act (CRA), among others, define the legal boundaries for AI deployment. For example, the EU AI Act classifies AI systems deployed in critical infrastructure like electricity sector as high risk, mandating requirements for transparency and human oversight. Similarly, the General Data Protection Regulation (GDPR) provides the legal basis for processing consumer energy data and imposes several key obligations. Data processing must follow the principle of data minimisation and, where appropriate, may require anonymization to limit the use of personally identifiable information. Article 22 of the GDPR restricts fully automated decision-making that significantly affects individuals. Although research on explainable AI and human-in-the-loop systems is growing, operational use remains largely limited to advisory roles, with full AI autonomy a distant goal.

([24]) Gorenstein Dedecca J., Ansarin, M., Bene, C., van Delzen, T., van Nuffel, L., Jagtenberg, H., 2025, Increasing Flexibility in the EU Energy System - Technologies and policies to enable the integration of renewable electricity sources, Publication for the Committee on Industry, Research and Energy, Policy Department for Transformation, Innovation and Health, European Parliament, Luxembourg.

([25]) Eurelectric (2025). Plugging into Potential: Unleashing the Untapped Flexibility of EVs – Policy Recommendations. Union of the Electricity Industry – Eurelectric, Brussels.

Cybersecurity concerns add complexity. AI systems controlling energy assets create new vectors for adversarial attacks or data manipulation, potentially threatening system stability, which is an issue that is currently under-addressed.

2.3. Use Case 3: Generation and Load Forecasting

As renewable penetration and load variability increase, accurate short-term forecasting is becoming critical for maintaining grid stability, optimising market participation and activating flexibility. AI-based forecasting methods enhance the precision and responsiveness of generation and demand forecasts, enabling system operators, aggregators and traders to better manage uncertainty and reduce balancing costs.

2.3.1. Status Quo & Market Potential

The growing complexity of energy systems has created a strong demand for improved forecasting capabilities. **Renewable generation forecasting**, particularly for wind and solar, has matured significantly and is widely deployed across Europe. These forecasts, strongly influenced by weather conditions, have benefited from major advances in AI-based approaches such as neural networks, ensemble models and hybrid physical-statistical methods ([26]). Established companies like Quadra Energy (Germany, now part of TotalEnergies, France) and Meteomatics (Switzerland, with offices within the EU), as well as innovative startups such as ENFOR (Denmark), provide operational forecasting solutions for renewable energy generation, supporting grid operators and energy suppliers with high-resolution, data-driven predictions. Their services are used by TSOs, utilities and energy traders to inform market bids and system operation. These platforms often combine meteorological data with machine learning algorithms to offer high-resolution, site-specific forecasts.

Increasing attention is being directed towards **probabilistic forecasting**, where models provide not only point estimates but also full distributions over possible future outcomes. This approach is especially valuable in high-renewable energy systems, where uncertainty plays a critical role in balancing, reserve allocation and risk-aware decision-making. Recent advances in machine learning, particularly **generative models** based on **diffusion architectures**, now enable the creation of realistic ensembles that capture the full range of plausible weather scenarios ([27]) coming hours. A key example is **GenCast**, developed by **Google**

([26]) Lafuente-Cacho, M., Izquierdo-Monge, O., Peña-Carro, P., Hernández-Jiménez, Á., Callejo, L. H., Losada, A. M. P., & Lamadrid, Á. L. Z. (2025). State of the Art for Solar and Wind Energy-Forecasting Methods for Sustainable Grid Integration. Current Sustainable/Renewable Energy Reports, 12(1), 1-12.

([27]) Price, I., Sanchez-Gonzalez, A., Alet, F. *et al.* Probabilistic weather forecasting with machine learning. *Nature* **637**, 84–90 (2025). https://doi.org/10.1038/s41586-024-08252-9

DeepMind, which demonstrates that machine learning can **outperform traditional ensemble numerical weather prediction systems** (such as the ECMWF ENS) in both accuracy ([28]) around 20% up to lead times of 2 days and computational efficiency. GenCast delivers well-calibrated probabilistic forecasts for essential variables like wind speed, temperature and extreme weather events, marking a significant shift towards the operational deployment of ML-based forecasting in applications such as renewable integration, disaster and climate resilience.

In the area of **load forecasting**, the situation is more complex. While some TSOs and large DSOs have deployed forecasting tools, particularly at the aggregated level, forecasting demand at higher spatial and temporal resolution especially for households and local grids remains a research and deployment challenge ([29]). The rise of electric vehicles, heat pumps and time-variable tariffs introduces irregular patterns in consumption. Traditional statistical models often perform well for temperature-driven loads like space heating and cooling but struggle with behavioural loads such as hot water usage, which exhibit less predictable, user-specific variability. AI-based models show growing promise in capturing these complex patterns. AI-based forecasting tools are increasingly being applied by companies such as Dexter Energy (Denmark), Sensewaves (France) or **Nectaware (Italy)**.

The economic value of improved forecasting is well documented. It supports more efficient market participation, reduces balancing costs, improves reserve sizing, and enables congestion avoidance. In addition, accurate forecasts are a key enabler for both **AI-based grid operation(Use Case 1)** and **flexibility activation (Use Case 2),** providing the predictive basis for optimal scheduling and the coordinated dispatch of distributed energy resources in line with expected system conditions.

Finally, privacy-preserving technologies such as **federated learning** are increasingly explored to enable fine-grained demand forecasting without requiring centralised access to individual ([30]). In federated learning, machine learning models are trained locally on users' devices or edge systems and only the model updates, but no the raw data is shared. These approaches are particularly relevant for DSO-level forecasting and flexibility use cases, where GDPR compliance is a central concern.

Data-driven forecasting relies on comprehensive historical datasets. For renewable generation, several benchmarking frameworks have been established

([28]) Soliman, A. (2024). *DeepMind AI weather forecaster beats world-class system*. *Nature*. https://www.nature.com/articles/d41586-024-03957-3

([29]) E.DSO. (2025). *Technology Radar v4* (Last updated version). European Distribution System Operators. https://www.edsoforsmartgrids.eu/technology-radar/

([30]) Fekri, M. N., Grolinger, K., & Mir, S. (2022). Distributed load forecasting using smart meter data: Federated learning with Recurrent Neural Networks. *International Journal of Electrical Power & Energy Systems*, *137*, 107669.

such as the IEA Wind Task 51 ([31]), which provides structured benchmarks for wind forecasting, and international exercises such as WIRE ([32]), which compared solar and wind forecasts on common datasets. In the meteorological domain, open initiatives such as WeatherBench ([33,34]) make standardised datasets and metrics available to evaluate forecasting models. For load forecasting no initiative is known that coordinates benchmarks, but individual datasets ([26,27,35]) are available online.

2.3.2. Barriers to Adoption

Despite growing maturity in many forecasting domains, several challenges continue to limit widespread adoption, particularly in the case of short-term and decentralised load forecasting. One of the most significant barriers is the issue of data privacy and sharing. Smart meter data, which is essential for high-resolution demand prediction, is classified as personal data under the GDPR. As a result, access to and use of this data is tightly controlled, and concerns about surveillance, profiling and data misuse have led to strong hesitance around data sharing, especially when it comes to individual-level forecasting for applications like residential flexibility activation.

To address these constraints, federated learning and edge-based training approaches have been proposed and tested in research settings. These techniques allow machine learning models to be trained on distributed local devices without transferring sensitive raw data to a central serverInstead, each client performs local training on its private dataset and transmits only aggregated model weight updates to a coordinating server, which maintains a shared global model. The model updates generally contain much less information than the raw data and therefore aligns with the principle of data minimisation. Moreover, they can be made ephemeral and transmitted without identifying metadata, further enhancing privacy protection ([36]). While they show great promise, their deployment in operational systems remains limited due to the need for reliable IT

([31]) https://iea-wind.org/task51/

([32]) Sperati, S., Alessandrini, S., Pinson, P., & Kariniotakis, G. (2015). The “weather intelligence for renewable energies” benchmarking exercise on short-term forecasting of wind and solar power generation. *Energies*, *8*(9), 9594-9619.

([33]) Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: a benchmark dataset for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2020MS002203.

([34]) Rasp, S., Hoyer, S., Merose, A., Langmore, I., Battaglia, P., Russell, T., ... & Sha, F. (2024). WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems*, *16*(6), e2023MS004019.

([35]) Emami, P., & Graf, P. (2018). BuildingsBench: A Large-Scale Dataset of 900K Buildings and Benchmark for Short-Term Load Forecasting. [Data set]. Open Energy Data Initiative (OEDI). National Renewable Energy Laboratory. https://doi.org/10.25984/1986147

([36]) McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282). PMLR.

infrastructure, robust communication protocols and organisational readiness to integrate such solutions into existing utility processes.

2.4. Use Case 4: Anomaly Detection & Predictive Maintenance

Ensuring the reliability and resilience of energy infrastructure requires timely detection of anomalies and condition-based maintenance strategies. Conventional maintenance approaches often rely on fixed schedules or reactive interventions, leading to inefficiencies and increased risk of failure. AI enables a shift towards proactive asset management by identifying abnormal patterns, predicting failures before they occur and optimising maintenance planning. This use case explores AI's role in strengthening operational continuity and cost efficiency in both grid and renewable assets.

2.4.1. Status Quo & Market Potential

In the grid context, AI models are increasingly deployed for the detection of operational anomalies such as voltage sags, frequency deviations, transformer degradation and cable insulation faults using real-time data from SCADA, PMU, and sensor networks. As part of a broader shift from time-based to risk-informed asset management, TSOs and DSOs are beginning to integrate these tools into their maintenance workflows to extend asset lifetimes, reduce unplanned outages and optimise maintenance planning.

Several DSOs are deploying AI-based systems for predictive maintenance, focusing on asset condition forecasting, anomaly detection and vegetation management[29]. Tools like *Analytics4Assets* and *CartoLine* enable early fault prediction on transformers, lines and voltage conditions. AI is also used to anticipate storm-related outages and neutral loss situations, improving preventive maintenance and grid resilience.

Predictive maintenance is one of the most impactful AI applications in the energy sector, enabling early failure detection, optimised maintenance cycles and reduced unplanned outages. According to McKinsey, AI-driven predictive maintenance can increase asset productivity by up to 20% and reduce overall maintenance costs by up to 10% ([37]). In the renewable energy domain, predictive maintenance using machine learning is particularly well developed for wind and solar power. For example, vibration signal analysis, thermographic imaging and

([37]) McKinsey & Company (2017). *Smartening up with Artificial Intelligence (AI) – What's in it for Germany and its Industrial Sector?*

supervisory control data are analysed to detect wear and faults in turbine blades or solar inverters.

Several companies are actively driving innovation in this space. EnBW is using AI to improve operational efficiency in wind power through advanced condition monitoring and fault prediction. Turbit provides machine learning-based solutions that analyse turbine data in real time to identify deviations and performance issues. Deutsche Windtechnik is engaged in multiple research projects on predictive maintenance for wind turbines, aiming to reduce downtime and optimise service intervals. NeuronSW develops AI-based diagnostic and predictive systems across energy infrastructure, including wind and solar assets. Meanwhile, PSI Software applies AI (PSIdetect) to identify potential threats to power grids early (e.g. anomalies that can be attributed to Cyber attacks), supporting both grid reliability and asset health management.

2.4.2. Barriers to Adoption

Despite increasing research and promising pilot projects, several challenges continue to hinder the widespread deployment of AI-based anomaly detection and predictive maintenance in grid and renewable energy environments. A primary barrier is data-related ([38,39]): high-quality, high-resolution and labelled operational data are not consistently available across asset types, grid levels or national contexts. Legacy infrastructure, data silos between departments or systems, and the absence of harmonised data standards complicate data aggregation and limit interoperability especially in heterogeneous, multi-vendor environments.

To address these issues, the development of energy data spaces offers a systemic solution ([40]). These shared, governed platforms enable secure, standardised and sovereign data exchange across actors and systems, breaking down silos while preserving control over sensitive operational information. By improving data accessibility, harmonisation, and interoperability even across legacy and modern systems data spaces create the foundation for training more generalisable AI models and deploying predictive maintenance applications at scale.

([38]) Alam, M. M., Hossain, M. J., Habib, M. A., Arafat, M. Y., & Hannan, M. A. (2025). Artificial intelligence integrated grid systems: Technologies, potential frameworks, challenges, and research directions. *Renewable and Sustainable Energy Reviews*, *211*, 115251.

([39]) Mortensen, L. K., Sundsgaard, K., Shaker, H. R., Hansen, J. Z., & Yang, G. (2024). *Designing digitally enabled proactive maintenance systems in power distribution grids: A scoping literature review*. Energy Reports, 12, 1–21. https://doi.org/10.1016/j.egyr.2024.08.044

([40]) Farrell, E.; Minghini, M.;Kotsev, A.; Soler-Garrido, J.; Tapsall, B.; Micheli, M.; Posada, M.; Signorelli, S.; Tartaro, A.; Bernal, J.; Vespe, M.; Di Leo, M.; Carballa-Smichowski, B.; Smith, R.; Schade, S.; Pogorzelska K.; Gabrielli, L.; De Marchi, D., European Data Spaces: Scientific insights into data sharing and utilisation at scale, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/400188, JRC129900.

2.5. Use Case 5: AI-Supported Permitting of Energy Infrastructure

Slow and fragmented permitting processes represent a major barrier to the timely expansion of renewable energy infrastructure. Complex legal frameworks, unstructured data and high administrative workloads lead to project delays and investor uncertainty. AI technologies can streamline permitting procedures by automating document analysis, extracting regulatory insights and predicting approval timelines. This use case examines how AI can help overcome bureaucratic bottlenecks while maintaining legal compliance and stakeholder transparency.

2.5.1. Status Quo & Market Potential

Permitting remains one of the most entrenched bottlenecks in the expansion of renewable energy infrastructure across the European Union ([41]). Despite legislative advancements, approval processes for wind farms, solar installations and grid expansions are still characterised by fragmentation, legal complexity and administrative inertia. These inefficiencies result in significant delays and rising development costs, undermining both investor confidence and the EU's ability to meet its climate and energy targets.

At the same time, the emerging capabilities of AI, particularly in the form of generative models and natural language processing, are opening new possibilities for modernising and accelerating permitting workflows ([42]). AI can assist in reviewing and interpreting complex legal and technical documents, automate the drafting of permit applications, and extract and classify regulatory requirements from heterogeneous sources. It also enables the bulk extraction and structuring of data from past permits and environmental reviews, creating valuable datasets on project characteristics and development timelines. This structured information can be used to identify correlations between project types and permitting delays, improve mitigation strategies and enhance the overall efficiency of environmental reviews. In addition, AI-powered geospatial screening can help identify suitable locations for new infrastructure by using e.g. satellite data ([17]).

While concrete implementations in Europe remain limited, there are early signals of progress. A pioneering example can be found in Denmark where the Danish Environmental Protection Agency partnered with the technology firm cBrain to

([41]) Banet, C., & Donati, F. (2024, October). *Speeding up renewable energy permitting in Europe: Overcoming implementation challenges* (Report). CERRE – Centre on Regulation in Europe.

([42]) Benes, K. J., Porterfield, J. E., & Yang, C. (2024). AI for energy: Opportunities for a modern grid and clean energy economy. *U.S. Department of Energy*.

develop and implement a fully digital, AI-supported permitting system ([43]). The agency succeeded in converting its permitting procedures into a standardised and structured digital format using a government-specific platform, eliminating the need for custom software development. This digital transformation laid the groundwork for the secure integration of AI, including large language models trained on historical permitting data within isolated, on-premise environments. The solution demonstrates how rule-based automation and context-specific AI can significantly enhance the speed, transparency and legal robustness of complex environmental permitting processes while ensuring full compliance with data privacy and regulatory standards. Similarly, EU-level initiatives are beginning to encourage broader experimentation. One such effort is the European Regions Research and Innovation Network (ERRIN) call to apply AI and generative AI in public administrations ([44]).

By contrast, the United States has shown stronger momentum in this area. Recent work by the NREL, including the SolarAPP+ platform and research on using large language models to automate the extraction of local siting ordinances, demonstrates how AI can enhance both permitting efficiency and the structured management of policy-relevant data for renewable energy deployment ([45,46]). Another example is PermitAI, developed by the Pacific Northwest National Laboratory for the U.S. Department of Energy, which leverages a generative AI-driven database of thousands of environmental impact statements to help regulators interpret NEPA requirements and streamline project reviews by making previously inaccessible information searchable and actionable ([47]).

2.5.2. Barriers to Adoption

One of the most pressing challenges is the legal fragmentation and complexity across jurisdictions. Permitting frameworks differ widely between EU member states, regions and municipalities, which makes it difficult to design AI systems that can generalise across diverse legal codes, administrative procedures and environmental standards. To operate effectively and compliantly, AI models, especially those based on natural language processing, must be continuously trained and updated using jurisdiction-specific legal and regulatory data. This

([43]) Knudsen, C. T. (2023). *AI for environmental permitting: Accelerating government environmental permitting with use of AI* (Whitepaper). cBrain. https://www.cbrain.com/post/ai-for-environmental-permitting

([44]) https://errin.eu/calls/apply-ai-genai-public-administrations

([45]) Williams, Juliana, Jeffrey J. Cook, Jesse R. Cruce, Kaifeng Xu, Seth Crew, Minahil Qasim, and Matt Miccioli. 2022. SolarAPP+ Pilot Analysis: Performance and Impact of Instant, Online Solar Permitting. Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-81603 https://www.nrel.gov/docs/fy22osti/81603.pdf

([46]) Buster, G., Pinchuk, P., Barrons, J., McKeever, R., Levine, A., & Lopez, A. (2024). Supporting energy policy research with large language models: A case study in wind energy siting ordinances. *Energy and AI*, *18*, 100431.

([47]) Dunkin, A., & Sharma, R. (2025, June 24). *Enhancing U.S. power grid by using AI to accelerate permitting*. Federation of American Scientists. https://www.fas.org/publication/enhancing-us-power-grid-by-using-ai-to-accelerate-permitting/

underscores the need for sandbox environments, where AI systems can be securely developed, tested and refined using internal, context-specific data without compromising privacy or breaching regulatory constraints ([48]). Such sandboxed AI approaches are particularly relevant under the EU AI Act, which imposes stricter obligations on high-impact public-sector applications, including those used in land-use and environmental governance.

A second major barrier is the lack of adequate digital infrastructure within permitting authorities. Many agencies still rely on outdated IT systems, fragmented databases and non-standardised documentation formats, which are incompatible with modern AI tools. The absence of machine-readable planning documents and harmonised datasets severely limits the portability and effectiveness of AI models. Addressing this gap will require significant investment in open data infrastructure, standardised regulatory ontologies and interoperable digital workflows that can support AI integration across institutional boundaries.

Institutional trust in AI-assisted decision-making also remains low, particularly in domains involving environmental protection, land rights and public participation. Concerns about transparency, legal accountability and fairness are especially acute in processes where communities must be consulted or where ecological impacts are assessed. Without robust explainability and audit mechanisms, AI-generated outputs are unlikely to be accepted as legitimate by legal authorities or stakeholders. The EU AI Act reflects this concern by mandating human oversight, transparency obligations and documentation requirements for high-risk AI systems. Therefore, explainable AI methods must be built into permitting tools from the outset, ensuring that AI recommendations are understandable, traceable and aligned with procedural safeguards.

2.6. Policy recommendations

After outlining relevant use cases and describing their functionality, market potential land adoption barriers, this section now turns to policy recommendations aimed at establishing and scaling up AI in the EU energy system. These recommendations are designed to both generally enable AI deployment across the energy sector and address specific needs of individual use cases. Given that AI is a cross-sectoral technology, a holistic policy approach is essential. Each recommendation is presented with an explanation of its relevance and a discussion of its potential impact on the previously described use cases.

([48]) Knudsen, C. T. (2023). *How to unlock the potential of context-specific AI by leveraging a sandbox approach* (Whitepaper). cBrain. https://www.cbrain.com/post/how-to-unlock-the-potential-of-context-specific-ai-by-leveraging-a-sandbox-approach

2.6.1. Establish a SME-Friendly AI Governance Framework for the Energy Sector

The growing integration of AI into Europe's energy systems, across forecasting, maintenance and demand-side services, makes it increasingly important to simplify and harmonise the regulatory framework, particularly to support small and medium-sized enterprises (SMEs). These innovators often face disproportionate burdens navigating complex and overlapping rules, including the AI Act, GDPR, NIS2 and national energy regulations. This fragmented "compliance matrix" discourages SME participation and slows innovation. Studies show that SMEs struggle with ambiguous risk classifications, overlapping legal obligations and high documentation demands ([49],[50]). Compliance costs for AI systems classified as "high-risk" under the EU AI Act can reach hundreds of thousands of euros, making them inaccessible without support ([51]). This high-risk classification, if not clearly scoped and aligned with sector-specific regulation, could lead to a regulatory lock-in effect that discourages innovation and excludes early-stage solutions from the market. Without tools like regulatory sandboxes or simplified conformity paths, current frameworks risk reinforcing incumbent dominance and limiting market access ([52]). To address these barriers, the EU should establish a harmonised, SME-oriented AI governance framework tailored to the energy sector. This could include modular "trust" toolkits, standardised audit templates and clear certification routes. Such toolkits could comprise customisable components that help SMEs systematically address legal and technical requirements when developing or deploying AI systems in the energy sector. These might include risk assessment checklists to identify potential harm or regulatory concerns early in the development process, and data protection impact assessment (DPIA) templates to ensure that personal data (for example, smart meter readings or user profiles) is handled in line with GDPR rules. They could also offer model documentation guidelines to help developers explain how their AI models work, what data they rely on and how decisions are made. In addition, toolkits could include preapproved plug-ins or software libraries for specific functions such as explainability (for example, explaining why the AI made a certain decision) or privacy-preserving computation (such as federated learning or data anonymisation). By providing these tools in an easy to adapt, ready-made format, the toolkits would reduce the burden on smaller firms, which often lack the legal, technical, or compliance staff needed to meet complex regulations like the EU AI Act or GDPR. Instead of building everything from scratch, SMEs could choose from prevalidated modules and templates that match their risk level and

([49]) Joswig, T., & Kurz, W. (2025). Regulatory and compliance requirements for SMEs operating AI systems through data centres in the EU, with a focus on data protection challenges in Germany. *Journal of Next-Generation Research 5.0*, *1*(2). https://doi.org/10.70792/jngr5.0.v1i2.89

([50]) Walters, J., Dey, D., Bhaumik, D., & Horsman, S. (2023). Complying with the EU AI Act. *arXiv preprint*. https://doi.org/10.48550/arXiv.2307.10458

([51]) Haataja, M., & Bryson, J. (2021). What costs should we expect from the EU's AI Act? *OSF Preprints*. https://doi.org/10.31235/osf.io/8nzb4

([52]) Apráez, B. E., & Noorman, M. (2024). Regulating AI in the 'twin transitions': Significance and shortcomings of the AI Act in the digitalised electricity sector. *Review of European, Comparative & International Environmental Law*. https://doi.org/10.1111/reel.12574

use case, helping them to innovate responsibly without being overwhelmed by red tape. Practical building blocks already exist, for example Microsoft's open source Responsible AI Toolbox ([53]); IBM's AI Explainability 360 ([54]) and AI Fairness 360 libraries; the DARPA ([55]) funded XAITK explainability toolkit; and the OECD/G7 Toolkit for AI in the Public Sector, which bundles procurement and risk management checklists that can be adapted by energy regulators and municipal utilities ([56]).

International frameworks such as the OECD AI Principles and the NIST AI Risk Management Framework offer valuable foundations, but they need to be adapted to the specific needs of the energy sector, ideally through collaboration with European SME stakeholders, and operationalised via shared infrastructure and coordinated governance mechanisms ([57]). This would strengthen AI deployment across a wide range of energy applications, with particular importance in areas where technical innovation intersects with regulatory complexity. Harmonised governance would reduce entry barriers for SMEs working on grid optimisation, decentralised flexibility and advanced forecasting, which are domains where legal uncertainty and compliance overhead currently slow progress. It would also support the scaling of AI tools in predictive maintenance and permitting, where standardised evaluation criteria and clearer legal pathways are critical. Overall, a more coherent and SME-friendly regulatory environment is essential for unlocking innovation and ensuring that AI contributes meaningfully to Europe's energy transition.

2.6.2. Strengthening the Societal Acceptance of AI in Energy through Transparency and Inclusion

As AI systems play an increasingly central role in both day-to-day operations and long-term energy planning, their societal legitimacy will hinge on public trust, regulatory alignment and institutional readiness. Societal acceptance should not be viewed as an afterthought but as a core condition for the effective and sustainable deployment of AI ([58]). This acceptance depends on whether AI systems are transparent, explainable and aligned with principles of fairness and accountability. Without these attributes, even technically sound solutions may be

([53]) Microsoft. *Responsible AI Toolbox* (GitHub repository). Available at: https://github.com/microsoft/responsible-ai-toolbox (accessed 25 July 2025).

([54]) IBM. *AI Explainability 360* and *AI Fairness 360* (GitHub repositories). Available at: https://github.com/Trusted-AI/AIX360 and https://github.com/Trusted-AI/AIF360 (accessed 25 July 2025).

([55]) XAITK. *Explainable AI Toolkit* (project website). Available at: https://xaitk.org/ (accessed 25 July 2025).

([56]) OECD. *G7 Toolkit for Artificial Intelligence in the Public Sector* (2024). Available at: https://www.oecd.org/en/publications/g7-toolkit-for-artificial-intelligence-in-the-public-sector_421c1244-en.html (accessed 25 July 2025).

([57]) Buocz, T., Pfotenhauer, S., & Eisenberger, I. (2023). Regulatory sandboxes in the AI Act: Reconciling innovation and safety? *Law, Innovation and Technology*, *15*(2), 357–389. https://doi.org/10.1080/17579961.2023.2245678

([58]) Ahmad, M., Anshari, M., Mardani, A., & Alkhalaf, S. (2024). Ethics and public trust in AI governance: A literature review. *AI*, *5*(2), 212–225. https://doi.org/10.3390/ai5020013

rejected by operators, consumers or regulators. Poorly governed AI can lead to discriminatory or opaque outcomes, eroding public trust, particularly in domains like dynamic pricing or permitting where social equity and procedural legitimacy are essential. Research highlights the importance of embedding user-friendly, interpretable interfaces and system-level transparency mechanisms such as dashboards, bias indicators and fairness metrics to support trust and informed participation ([59,60,61]). Ethical guidelines under discussion in the industry are a step forward, but they must be complemented by enforceable regulatory standards and incentives for trustworthy design.

The importance of societal acceptance is evident across all major AI applications in the energy sector. AI systems must be transparent, trustworthy, and aligned with user needs and expectations, whether they support grid operation, enable demand-side flexibility (e.g., through dynamic grid tariffs), improve forecasting, optimise maintenance, or accelerate permitting. Operators, consumers and regulators are more likely to adopt and rely on AI when its outputs are explainable and its decisions auditable. As AI becomes integral to critical grid functions like real-time dispatch and flexibility activation, it also becomes a target for cyberattacks and system failures—making it essential to embed cybersecurity safeguards, fallback mechanisms, and human oversight into every AI-driven process. Recent EU regulations such as the Network Code on Cybersecurity and the AI Act reflect this imperative by mandating continuous risk assessments, tamper-proof authentication, model transparency and human-in-the-loop controls, ensuring that resilience and trust are foundational to digital transformation in the energy sector. Without such trust, even technically advanced systems risk rejection or underutilisation. Embedding transparency, fairness and inclusivity across these use cases is therefore not optional; it is a foundational requirement for the responsible and effective integration of AI into Europe's energy transition.

2.6.3. Launch an EU-wide Benchmark Suite for Evaluating Energy-AI Systems

The absence of consistent evaluation tools for AI systems in the energy sector remains a key barrier to trustworthy and effective deployment. While AI applications are advancing rapidly, policymakers, regulators and system operators still lack shared criteria to assess performance, fairness and sustainability. Recent studies highlight the urgent need for standardised,

([59]) Kim, S. S. Y. (2024). Establishing appropriate trust in AI through transparency and explainability. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613905.3638184

([60]) Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., & Holzinger, A. (2022). Fairness and explanation in AI-informed decision making. *Machine Learning and Knowledge Extraction*, *4*(2), 556–579. https://doi.org/10.3390/make4020026

([61]) Knowles, B., Richards, J. T., & Kroeger, F. (2022). The many facets of trust in AI: Formalizing the relation between trust and fairness, accountability, and transparency. *arXiv preprint*. https://doi.org/10.48550/arXiv.2208.00681

transparent and transferable evaluation metrics for energy AI systems, as current approaches lack agreed-upon measures for explainability, forecasting accuracy, and environmental impact ([62,63,64]). To address this gap, the EU should develop a central Energy-AI Benchmark Suite aligned with the Common European Energy Data Space. This suite should include metrics for load forecast error, demand-side pricing fairness, cybersecurity resilience and carbon intensity of model training and inference. Benchmarking tools should be made openly accessible and co-developed with industry, operators and civil society to ensure broad usability and trust. Initiatives such as LF Energy, with projects like Grid2Op and the L2RPN challenge series, provide valuable blueprints for open, high-fidelity benchmarking environments and could serve as foundational components or partners in this effort. These benchmarks would improve transparency in public procurement, reduce uncertainty for innovators and support regulatory oversight. The impact would be most significant in domains like forecasting, flexibility orchestration and grid planning, where confidence in model performance and fairness is critical to societal acceptance and market integration. For use cases such as AI-assisted grid operation, reliable benchmarks are essential to assess safety and support operator trust. In demand-side response, standard metrics can guide aggregators and distribution system operators in evaluating fairness and accuracy in pricing or load-shifting decisions.

2.6.4. Ensuring Technological Sovereignty in the Adoption of AI in Europe's Energy Sector

The increasing adoption of AI in the energy sector increases the urgent need for policymaking that secures national and European technological sovereignty, as the sector forms part of Europe's critical infrastructure. Yet Europe's ability to shape these systems in line with its democratic values is coming under growing pressure. Control over key components of the AI stack, ranging from foundational models and compute infrastructure to cloud services and the entire chip manufacturing value chain, is concentrated in a small number of non-European firms. With limited exceptions such as ASML and Siltronic, this concentration creates strategic dependencies. Without coordinated policy action, public and municipal energy actors risk becoming locked into opaque, proprietary systems that undermine transparency, flexibility and democratic oversight ([65]). Europe's structural weaknesses in AI development amplify these risks. The EU lags substantially behind the United States in foundation model development and

([62]) Mouakher, A., Inoubli, W., Ounoughi, C., & Kő, A. (2022). Expect: EXplainable Prediction Model for Energy ConsumpTion. *Mathematics*, *10*(2), 248. https://doi.org/10.3390/math10020248

([63]) Necula, S.-C. (2023). Assessing the potential of artificial intelligence in advancing clean energy technologies in Europe: A systematic review. *Energies*, *16*(22), 7633. https://doi.org/10.3390/en16227633

([64]) Delanoë, P., Tchuente, D., & Colin, G. (2023). Method and evaluations of the effective gain of artificial intelligence models for reducing CO_2 emissions. *Journal of Environmental Management*, *331*, 117261. https://doi.org/10.1016/j.jenvman.2023.117261

([65]) Bellanova, R., Carrapico, H., & Duez, D. (2022). Digital sovereignty and European security: Shaping the next generation EU. *European Security*, *31*(1), 1–18. https://doi.org/10.1080/09662839.2022.2043031

private generative-AI investment, capturing only a limited share of global activity across the AI value chain ([66]). See figure below on the number of notable AI models by geographical areas. Analysts warn that Europe's fragmented markets, chronic underfunding and reliance on non-EU infrastructure continue to constrain its innovation capacity, casting doubt on its ability to translate regulatory leadership into technological sovereignty ([67]). This imbalance extends into the energy domain, where most large data centres are not owned by European companies, and persistent grid and capacity constraints are already "hampering the development of AI" within the sector ([68]). Together, these conditions risk entrenching Europe's dependency on external providers for the digital foundations of its energy transition, undermining both autonomy and democratic control.

Figure on foundation models ([69]):

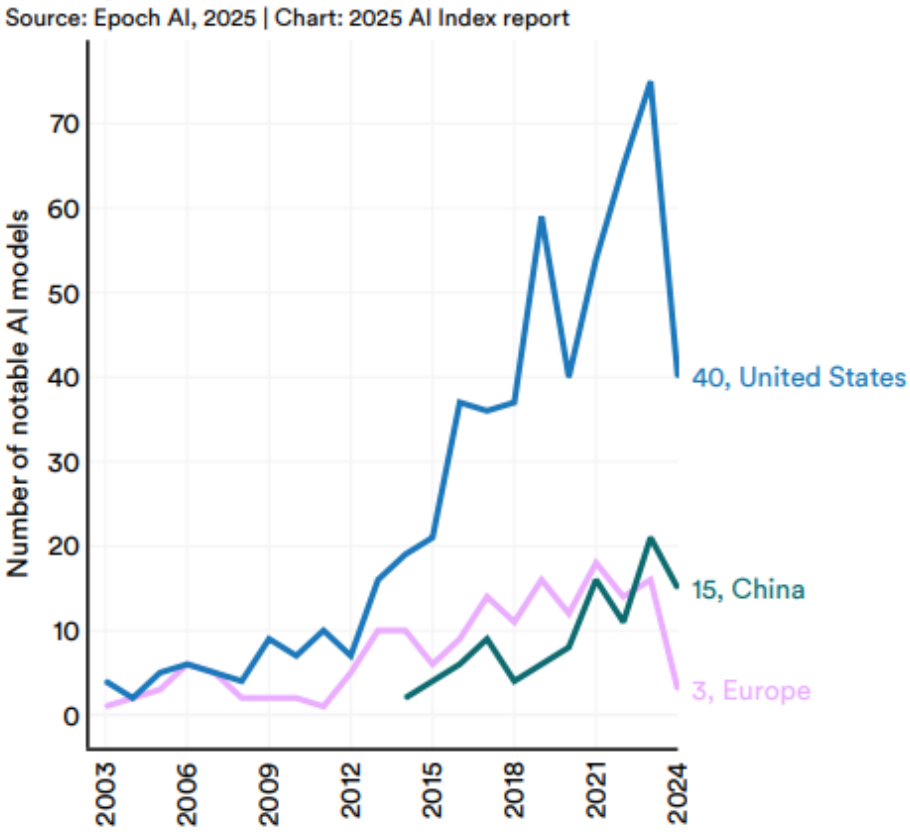


Figure 1: Number of notable AI models by select geographic areas, 2003-24

([66]) European Economic and Social Committee, & Centre for European Policy Studies. (2025). *Generative AI and foundation models in the EU: Uptake, opportunities, challenges, and a way forward*. Publications Office of the European Union. https://doi.org/10.2864/8377116

([67]) Csernatoni, R., Broeders, D., Andersen, L. H., Hoijtink, M., Bode, I., Lindsay, J. R., & Schwarz, E. (2025). Myth, power, and agency: Rethinking artificial intelligence, geopolitics and war. *Minds and Machines*, *35*(3), 37.

([68]) Widuto, AI and the energy sector. Briefing, European Parliamentary Research Service (EPRS), PE 775.859, Jul. 2025

([69]) Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., ... & Oak, S. (2025). Artificial intelligence index report 2025. arXiv preprint arXiv:2504.07139.

To safeguard its technological sovereignty, the EU should establish a Sovereignty AI Energy Hub Initiative. Addressing this complex challenge requires coordinated experimentation, collaboration across the full AI value chain and alignment with societal priorities, none of which can be achieved through fragmented efforts alone. The proposed hub would build on and interconnect existing initiatives, such as Digital Innovation Hubs (e.g., HPC4AI in Italy, ATTRACT DIH in Portugal, de:hub in Germany) and national AI testbeds (e.g., AI Sweden Federated-Learning Testbed, TEF-Health, CitCom.ai for smart cities, and the Test Area Autonomous Driving Baden-Württemberg), within a unified framework anchored in the EuroHPC supercomputing infrastructure and the Common European Energy Data Space. By extending its scope to adjacent sectors like +mobility, buildings and industry, where energy-related AI applications are rapidly converging, the initiative would promote scale, open standards and interoperability. This would reduce vendor lock-in and preserving Europe's ability to shape the digital backbone of its clean energy transition.

2.6.5. Towards a Trusted, Interoperable and Privacy-Preserving Data Foundation for AI in Energy

A secure, interoperable data foundation is essential for scaling AI in Europe's energy systems. While initiatives like the Common European Energy Data Space and Gaia-X promote secure data exchange, siloed infrastructure and inconsistent standards still limit AI's effectiveness. Today, only a small fraction of smart grid data is used due to privacy concerns, lack of interoperability and technical fragmentation. This constrains performance in areas such as forecasting, flexibility and predictive maintenance. Open standards (e.g., IEEE 2030.5), federated architecture, and privacy-enhancing technologies like anonymization and federated learning are needed to address these gaps. In addition, synthetic data provides a crucial cornerstone: it is artificially generated information that mirrors the statistical properties of real measurements, allowing AI developers to train, test and benchmark models without exposing any individual household, plant or commercial secret. By removing re-identification risk and safeguarding proprietary information that underpins a firm's competitive advantage, synthetic datasets let multiple utilities, vendors and researchers collaborate on common algorithms for forecasting, flexibility orchestration or anomaly detection while still respecting privacy and competition law. Because commercially sensitive patterns, customer lists or operational set-points are never revealed, companies are far less reluctant to participate in cross-industry data-sharing initiatives, accelerating collective learning and sector-wide innovation ([70]). Together with federated learning and anonymisation, these technologies would allow utilities to train AI models locally, without sharing raw customer data, thereby preserving

([70]) Hao, S., Han, W., Jiang, T., Li, Y., Wu, H., Zhong, C., ... & Tang, H. (2024). Synthetic data in AI: Challenges, applications, and ethical implications. *arXiv preprint arXiv:2401.01629*

privacy while still enabling system-wide learning and optimisation ([71]). By protecting sensitive information, they also reduce fears of losing competitive advantage, giving data owners stronger incentives to participate in data-sharing frameworks ([72]). Without them, AI tools risk limited deployment or social pushback.

Promoting trusted privacy-preserving techniques would unlock all the use cases discussed. By enabling secure collaboration, these technologies support shared diagnostics and structured asset data, which are crucial for predictive maintenance. Additionally, access to machine-readable planning data would streamline permitting workflows ([73]). Together, these capabilities form the backbone for deploying AI across the energy sector.

([71]) Shrestha, R., Mohammadi, M., Sinaei, S., Salcines, A., Pampliega, D., Clemente, R., Sanz, A. L., Nowroozi, E., & Lindgren, A. (2024). Anomaly detection based on LSTM and autoencoders using federated learning in smart electric grid. *Journal of Parallel and Distributed Computing*, *193*, 104951. https://doi.org/10.1016/j.jpdc.2024.104951

([72]) Syu, J.-H., Lin, J. C.-W., & Srivastava, G. (2024). Distributed learning mechanisms for anomaly detection in privacy-aware energy grid management systems. *ACM Transactions on Sensor Networks*. https://doi.org/10.1145/3640341

([73]) Farahani, B., & Monsefi, A. K. (2023). Smart and collaborative industrial IoT: A federated learning and data space approach. *Digital Communications and Networks*, *9*, 436–447. https://doi.org/10.1016/j.dcan.2023.01.022

3. Task 3: Foundation Models for the Energy Sector

3.1. Introduction

Foundation models (FMs) are a type of artificial intelligence (AI) technology that is trained on vast amounts of data with different characteristics and types from diverse datasets, allowing them to be adapted to a wide range of tasks and operations ([1,2]). Unlike traditional machine learning (ML) models, which are trained for specific purposes, FMs learn general-purpose representations that can be transferred to various domains or fine-tuned for various applications, ranging from natural language processing to computer vision.

The energy sector is undergoing a transformation, driven by the need for increased sustainability, resilience and efficiency ([3]), which present growing challenges in grid renewal, integration of renewable energy and distributed energy resources, demand-side management, grid resilience and cybersecurity. These challenges create opportunities for applying advanced AI techniques, including FMs, to enhance forecasting accuracy, optimise grid planning and operations, automate complex tasks, and to reduce operational costs, among other benefits([4,5]). While the potential of FMs in various industries is increasingly gaining attention, their specific application within the energy sector, particularly in grid operations and demand-oriented production, requires more focused investigation and committed investment.

This report presents a study on the potential of FMs to address the needs in the energy sector. The scope of this study encompasses technical requirements for FM development and deployment, cybersecurity risks associated with their use in critical infrastructure and the collaborative ecosystems necessary for successful implementation. The primary objectives of this report are:

([1]) Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E. and Brynjolfsson, E., 2021. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

([2]) Competition & Markets Authority CMA (2023) AI Foundation Models: Initial Report, https://www.gov.uk/government/publications/ai-foundation-models-initial-report.

([3]) IEA (2023), World Energy Outlook 2023, IEA, Paris https://www.iea.org/reports/world-energy-outlook-2023, Licence: CC BY 4.0 (report); CC BY NC SA 4.0 (Annex A)

([4]) Hines, P., Dowling, A., & Blumsack, S. (2020). Machine learning for electric power systems: A critical review. Electric Power Systems Research, 189, 106599

([5]) Tong, L., Wang, X. and Zhao, Q., 2025. Grid monitoring with synchro-waveform and AI foundation model technologies. Energy Systems, pp.1-28.

1. to assess the potential applications of FMs in grid operations and demand-oriented production (i.e. forecasting of energy demand, grid optimisation, predictive maintenance and knowledge management),
2. to identify technical requirements for deploying FMs, including data infrastructure, computational resources and expertise,
3. to evaluate the cybersecurity vulnerabilities introduced by FMs in energy infrastructure and recommend mitigation strategies,
4. to analyse the collaboration aspects necessary for successful FM implementation, including data sharing, standardisation and regulatory support frameworks, and
5. to provide recommendations for proactive preparation to leverage FMs effectively in the energy sector.

This report aims to provide a balanced assessment of the potential and challenges associated with deploying FMs in the energy sector, contributing to informed decision-making and strategic planning for a more resilient and sustainable energy future.

3.2. Foundation Models and the Energy Sector Use Cases

There is a wide range of applications where deploying FMs can be beneficial for grid operation and demand-oriented production applications. This study has been conducted on five prioritised use cases presented in Table 1, based on the expected impact of enhancement and the current state of development and technology integration in the use cases.

Table 1 – Selected Use Cases and Selection Criteria

Use Case 01: Forecasting of Energy Demand, Renewable Generation, and Grid Conditions	
Selection Criteria	• High impact on grid stability and renewable energy integration • Improves forecasting accuracy and consequently short-term planning reducing reliance on backup generation • Enables real-time adjustments and better operational efficiency
Data Types	Historical weather data (temperature, wind speed, solar radiation), smart meter data (demand patterns), IoT sensor data
Models/Methods	Time-series forecasting, machine learning (e.g., LSTM, XGBoost), regression models
Use Case 02: Optimisation for Grid Planning, Operations, and Real-Time Control	
Selection Criteria	• Essential for managing complex, decentralised grids • Improves grid efficiency by optimising power flow in real-time • Enhances resilience by predicting and addressing grid stress points

Data Types	Real-time grid status (voltage, active and reactive power), weather data energy demand, renewable generation data (solar, wind), grid infrastructure status
Models/Methods	Reinforcement learning (RL), optimisation algorithms (e.g., mixed-integer programming, genetic algorithms)
Use Case 03: Predictive Maintenance for Energy Assets and Infrastructure	
Selection Criteria	• Reduces unplanned downtime and increases equipment longevity • Saves maintenance costs and avoids major failures • Improves operational reliability and safety • Supports asset management and operational optimisation efforts
Data Types	Sensor data (vibration, temperature, oil quality), maintenance logs, machine instructions, images, videos, and equipment failure history
Models/Methods	Machine learning (e.g., Random Forest, SVM), anomaly detection, regression models)
Use Case 04: Demand Side Management and Intelligent Flexibility Activation	
Selection Criteria	• Enables dynamic grid balancing by adjusting demand in real-time • Reduces reliance on peaking power plants • Helps integrate more renewable energy into the grid • Provides economic benefits to consumers and utilities alike
Data Types	Smart meter data, appliance usage patterns, pricing signals, weather data, and consumer preferences
Models/Methods	Reinforcement learning, optimisation algorithms, predictive models
Use Case 05: Tacit and Explicit Knowledge Management from Operational Data and Expert Insights	
Selection Criteria	• Facilitates the retention of institutional knowledge • Improves decision-making through AI-assisted insights • Enhances workforce training and reduces knowledge gaps • Supports continuous improvement and innovation in energy management
Data Types	Expert decision logs, narratives from senior engineers, incident resolutions, training materials, and operational reports
Models/Methods	Large Language Models (LLMs) such as GPT-4 for natural language understanding and generation, and decision-making patterns to generate real-time decision support and knowledge

Foundation models offer transformative potential for enhancing efficiency, reliability and resilience in both grid operation and demand-oriented production. These benefits come from FMs' capabilities to learn complex patterns from diverse, structured and unstructured, and high-dimensional datasets; to generalise for unseen situations and to adapt to evolving conditions to address

the multifaceted challenges of modern energy systems ([4]). In the current state of research and development, there is significant effort from academic institutions, national labs, relevant industries and technology providers that is strongly related to creating FM for the energy sector. The efforts are twofold: create FMs specific for different use cases in the energy sector, and develop FMs to address general categories of challenges shared among various use cases, such as data characteristics and performance of the deployed advanced models. Table 2 presents an overview of the existing FMs and their potential to address the challenges in various use cases.

Table 2 – Foundation Models and Potential Challenges to Address in Selected Use Cases

FM Model	Category	Potential Use Case and Approach	TRL
GPT (General Family -4/4v)	General LLMs	• Knowledge retrieval, content generation, situation awareness (multimodal), optimisation • General large language/multimodal capabilities; used as backbone or via prompting for tasks like summarising documents, generating control instructions, multimodal scene description, and solving simplified OPF ([6,7])	5-7
BERT	General LLMs	• Underlying time series backbones, text tasks • Early NLP FM was used as a backbone for time series models or for text-based energy tasks like document analysis ([8,9])	3-5
LLaMA	General LLMs	• Underlying time series backbones. • Open-source LLM used as a backbone for time series models ([8])	3-5

([6]) Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., ... & Tacchetti, A. (2018). Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261.

([7]) Liu, S., Dong, L., Tian, C. and Xie, L., 2025. Unlocking Multi-Task Electric Energy System Intelligence: Data Scaling Laws and Performance with Limited Fine-Tuning. arXiv preprint arXiv:2503.20040.

([8]) Puech, A., Weiss, J., Brunschwiler, T., & Hamann, H. F. Optimal Power Grid Operations with Foundation Models. Available at: https://arxiv.org/abs/2409.02148.

([9]) Tu, S., Zhang, Y., Zhang, J., Fu, Z., Zhang, Y. and Yang, Y., 2024. Powerpm: Foundation model for power systems. Advances in Neural Information Processing Systems, 37, pp.115233-115260.

FM Model	Category	Potential Use Case and Approach	TRL
MetaIndux	Industrial/Multimodal	• Multiple industrial challenges. • Prototype Industrial FM system with multimodal pretraining, mechanism embedding, industrial agent collaboration; addresses Q&A, scene cognition, decision-making, terminal control, content generation, scientific discovery in industrial/energy context.	4-6
GridFM-v0 ([16])	Industrial/Grid	• Power grid operation, planning and control • Pre-trained models on grid data to speed up computation, and enhance flexibility in a scaleable and economical way ([6,10,16])	6-7
PowerPM[9]	Time series specific	• Forecasting, imputation, anomaly detection, classification • Pre-trained on large, hierarchical electricity time series data; uses temporal & hierarchical encoders; novel self-supervised pre-training ([2,9,10])	5-7
IRM-based FM	Time series specific	• Real-time monitoring, fault detection, data compression • Physics-based architecture learns innovation representation from high-resolution synchro-waveform data for rapid anomaly detection and compression; uses sequential statistical tests ([6,11])	4-6
TimesFM, GPT4TS, Time-LLM, UniTime	Time series specific (categories of LLM enhanced timeseries FM)	• Forecasting, imputation • Enhance time series models using pre-trained LLMs as backbones or via prompt-based methods; show improved performance on time series tasks ([10,12,13])	3-5

([10]) Madani, S., Tavasoli, A., Khoshtarash Astaneh, Z., & Pineau, P.-O. Large Language Models integration in Smart Grids. Available at: https://arxiv.org/abs/2504.09059

([11]) Liu, Q., Liu, X., Liu, C., Wen, Q., & Liang, Y. TIME-FFM: Towards LM-Empowered Federated Foundation Model for Time Series Forecasting. Available at: https://arxiv.org/abs/2405.14252

([12]) Ren, L., Wang, H., Wang, Y., Huang, K., Wang, L. and Li, B., 2025. Foundation Models for the Process Industry: Challenges and Opportunities. Engineering.

([13]) Das, A., Kong, W., Sen, R. and Zhou, Y., 2024, July. A decoder-only foundation model for time-series forecasting. In Forty-first International Conference on Machine Learning.

FM Model	Category	Potential Use Case and Approach	TRL
GNN-based FMs	Industrial/Multimodal (General category)	• Power flow, state estimation, grid stability analysis • Graph neural networks used as underlying architecture for FMs; represents grid topology as graphs; learns power flow mappings and analysis ([2,7])	4-6
Physical informed ML	Industrial (General category)	• Understanding mechanisms, grid dynamics. • Integrates physical laws/equations into ML models (e.g., via loss functions) to improve accuracy and generalisation ability for grid dynamics, R&D tasks ([2,6,12,22])	3-5

3.3. Foundation Models and Development for the Energy Sector

Foundation models represent a new generation of machine learning models that are trained on a substantial and diverse dataset, making them capable of addressing a wide range of tasks and domains. At their core, FMs are usually built using large-scale, deep neural networks. The most popular and successful architectures for these models are transformers, which utilise attention mechanisms to effectively capture relationships in complex data, such as sequences of text or time series. However, recent studies show effective utilisation of other architectures, such as graph neural networks (GNNs) ([11,14]).

3.3.1. Technical Characteristics of FMs

One distinct feature of FMs is a two-phase training approach. The first phase, pre-training, aims to create a general-purpose base model with a broad understanding of underlying data structures and patterns. To achieve this, pre-training requires vast and diverse datasets. Depending on the type of FM, the data corpus may come from public sources (e.g. web text for training GPT-4), from years of internal operational data (e.g. four years of data from an electricity grid operator used for training PowerPM) ([9]), or from synthetic generation (e.g. simulated power flow solutions for GridFM) ([16]). In the case of more advanced FMs, pre-training is often multimodal – incorporating multiple data types such as text, images, time series or other structured formats to enable cross-modal understanding.

([14]) Miller, J. A., Aldosari, M., Saeed, F., Barna, N. H., Rana, S., Arpinar, I. B., & Liu, N. (2024). A Survey of Deep Learning and Foundation Models for Time Series Forecasting, http://arxiv.org/abs/2401.13912.

The second phase, fine-tuning, adapts the pre-trained FM to specific domains or tasks. It involves training the model further on smaller, specialised datasets relevant to the desired application. In contrast to the pre-training stage, fine-tuning often requires substantially less data and computational power. In the energy domain, fine-tuning can be used, for instance, to adapt a foundation model to a specific city using only regional data, or to tailor a model like GridFM to a utility's own network topology.

This combination of rapid adaptation and reduced resource requirements during fine-tuning makes FMs a particularly attractive solution for organisations with limited resources, such as small and medium-sized enterprises (SMEs). This two-step approach allows FMs to be highly versatile.

3.3.2. Prerequisites and Resource Considerations

FMs and traditional ML models have several prerequisites in common, especially regarding data and infrastructure. From a data perspective, both scenarios necessitate comprehensive digitalisation of processes, widespread sensor deployment, robust data management systems and effective data governance to ensure data quality, availability and reliability. Regarding development and training, both FMs and traditional ML require expertise in ML methodologies and sufficient computational hardware resources. However, due to the extensive scale and broader generalisation capabilities of FMs, additional and more specialised prerequisites become important. The nature and extent of these additional requirements depend significantly on the specific development approach chosen for the FM application.

3.3.3. Approaches to Developing FM Applications

There are various approaches to integrating FMs into operational practices. While this overview outlines the three main types of development strategy for FM applications, it should be noted that many more variants are possible in practice ([5]).

3.3.3.1. Approach I: Using Existing Foundation Models

The novel generalisation capabilities of FMs enable the development of customised AI applications with minimal or no additional training. For many organisations, this approach offers a highly effective means of leveraging advanced AI without the need for extensive resources. This is particularly advantageous in use cases involving generative foundation models for text or image generation, where the pre-trained models already demonstrate strong performance across a wide range of tasks. In such scenarios, training these models from scratch is not a viable option due to the substantial computational resources and large-scale datasets required ([15]).

([15]) Myers, D., Mohawesh, R., Chellaboina, V.I., Sathvik, A.L., Venkatesh, P., Ho, Y.H., Henshaw, H., Alhawawreh, M., Berdik, 16 D. and Jararweh, Y., 2024. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. Cluster Computing, 27(1), pp.1-26.

A common approach to utilising existing FMs involves accessing them via APIs offered by model creators such as OpenAI, MistralAI, or third-party service providers. In this scenario, the primary development effort centres on the effective integration and application of the pre-trained model within the target use case. As a result, expertise in software development and prompt engineering becomes significantly more important than traditional ML knowledge ([10]).

3.3.3.2. Approach II: Training Foundation Models from Scratch

Training an FM from scratch or conducting the pre-training phase for the base model involves developing a completely customised solution, starting from model architecture design to the full-scale training process. This approach allows for complete control over the choice of data, modelling decisions and the resulting performance characteristics. It is particularly suited to highly specialised domains where existing models do not adequately reflect the complexity of the application context. However, this strategy is resource-intensive and typically requires access to large-scale datasets and high-performance computing infrastructure (such as GPU clusters) ([8,16]).

Training a competitive FM for unstructured data types, such as text or images, is generally beyond the reach of many companies, especially SMEs, due to the extensive resources required. Projects of this scale typically require investments in the order of several million euros, e.g., the OpenEuroLLM project has a budget of approximately €37.4 million ([17]). Many AI use cases in the energy sector are based on structured data, such as tables or graphs. FMs tailored to these data types can achieve meaningful performance with significantly less training data and computational effort, and lower costs. For example, TabPFN, a leading FM for tabular data, was trained with less than one-thousandth of the computing power required for training LLMS ([18]). This difference illustrates the difficulty of making generalised statements about FMs, particularly regarding costs and sustainability.

Additionally, training an FM from scratch in the energy domain still demands modern high-performance hardware, but the requirements are typically far below those for general-purpose LLM models. This is due to the characteristics of industrial data, such as data sources and types, and lower verbal and lingual diversity, as well as the deployed types of models fitting for various industrial use cases. For structured data–oriented models of a size relevant to typical energy applications, prior work suggests that 0.1–1 petaFLOP-days can be sufficient, depending on model architecture and dataset size ([9,19]). Consequently, this approach can be particularly relevant for tech companies that already have expertise in ML, existing hardware and sufficient datasets.

([16]) Hamann, H.F., Gjorgiev, B., Brunschwiler, T., Martins, L.S., Puech, A., Varbella, A., Weiss, J., Bernabe-Moreno, J., Massé, A.B., Choi, S.L. and Foster, I., 2024. Foundation models for the electric power grid. Joule, 8(12), pp.3245-3258.

([17]) European Commission. Funding & Tenders Portal – News, https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/news/30300

([18]) Huang, C., Li, S., Liu, R., Wang, H., & Chen, Y. (2023). Large Foundation Models for Power Systems, http://arxiv.org/abs/2312.07044.

([19]) N. Hollmann et al., “Accurate predictions on small data with a tabular foundation model,” Nature 2025 637:8045, vol. 637, no. 8045, pp. 319–326, Jan. 2025, doi: 10.1038/s41586-024-08328-6.

3.3.3.3. Approach III: Hybrid Approach – Combining Existing and Self-Trained FMs

A promising trend in applying FMs is the hybrid approach, in which FMs are combined with task-specific models, rules and data pipelines in order to create flexible, intelligent systems. One common implementation of this is in agentic systems. Rather than solving a problem end-to-end with a single model, the FMs act as reasoning engines that interpret user input or environmental data and trigger a sequence of tailored operations. For example, in an energy planning tool, the FM might translate a text prompt into sub-tasks such as demand forecasting, policy lookup and scenario comparison, each of which is handled by specialised software components or ML models.

These systems often rely on pre-trained generative and LLM FMs to facilitate user interaction, commonly via chat interfaces and transparently present each step of the underlying decision-making or workflow process back to the user. An example use case can comprise: Step 1: Load demand data, Step 2: Run short-term forecast, Step 3: Generate risk score. Therefore, LLMs in agentic systems prevent the feeling of interacting with a black box, which could facilitate adaptation. The core tasks, such as performing a data analysis, are then carried out by machine learning models or other software tools.

Hybrid applications can also include multimodal data support, combining FMs specialised in each data modality. For instance, a single FM might process language queries, while another FM handles time series sensor data, and a third one processes geographic information. To create advanced agentic systems, the requirements from Approaches I and II are necessary. Additionally, the development of such complex systems demands new software architecture and pipelining skills.

Table 3 – A summary comparison between the three integration approaches.

Approach I – Using Existing Foundation Models	
Advantage	**Disadvantage**
• Minimal computational requirements • Rapid prototyping and deployment • Leverages ongoing model improvements • No need for deep ML expertise if fine-tuning is not required	• Higher risk of domain misalignment, hallucinations, and bias • Dependency on third-party APIs and licensing constraints • High degree of domain specialisation is limited even when fine-tuning is applied
Example for AI Energy Use Cases	
Deployment of large language FMs for AI-assisted knowledge management in energy operations. Pre-trained models like GPT-4 analyse expert logs, training material and incident reports to extract decision patterns.The system provides context-aware recommendations, supports real-time operator decisions, and enables new engineers to query a knowledge base for guidance on problem-solving.	
Approach II – Training Foundation Models from Scratch	
Advantage	**Disadvantage**
• Full control over architecture, training data, and evaluation	• High computational and financial cost • Demands significant ML and infrastructure expertise

• Enables strong data/domain alignment and better interpretability	• Large data corpora needed
Example for AI Energy Use Cases	
Training of a grid foundation model (GridFM) using graph-based architectures to solve power flow equations under contingency scenarios. Once trained, the models enable the millisecond-level simulation of thousands of scenarios, allowing for the real-time assessment of grid stability under a large number of failure configurations.	
Approach III – Hybrid Approach	
Advantage	**Disadvantage**
• Enables additional complexity and flexibility through Agentic design • Only use FM when actually needed and avoid unnecessary black box tasks • Especially suitable for multimodal tasks	• Require both skills and resources from Approaches I & II • High complexity in system integration and evaluation
Example for AI Energy Use Cases	
Agentic system setup using GPT-4 for natural language reasoning, a self-trained GridFM for fast scenario evaluation, and plotting tools for dynamic visualisation. The system assists grid operators by translating queries into analysis actions, generating simulation results and delivering insights through interpretable visuals and explanations.	

3.3.4. Evaluating FM-Based Energy Applications

The evaluation of FM-based applications depends substantially on their input modalities (e.g., structured vs. unstructured data), output types and application contexts. For FMs that use structured data as both input and output—such as time-series, tabular datasets, or graph-based representations—the same evaluation metrics commonly applied in traditional ML or statistical modelling can be used. FMs with tasks such as forecasting, anomaly detection and predictive maintenance can be evaluated using standard quantitative metrics like root mean squared error (RMSE), mean absolute error (MAE), precision, recall and F1-score. These metrics provide clear, interpretable benchmarks for assessing model performance in well-defined, data-driven energy applications ([5,13]).

This contrasts with tasks involving unstructured data, such as knowledge queries and chat-based operational decision support in grids and compliance systems. FM applications for these tasks are subject to qualitative evaluation, which can introduce new challenges for companies. These qualitative assessments typically involve user studies and evaluations centred on explainability, fairness, ethical alignment and task alignment in general. In these cases, human-in-the-loop evaluation can be an effective strategy to assess the usability and added value of FM-based systems. For instance, user-centric metrics—such as user satisfaction (measured via surveys), task completion rate and task completion time—can provide insights into how well the FM integrates with existing workflows and supports decision-making. Complementing these, economic indicators such as cost effectiveness, return on investment (ROI) and operational efficiency gains (e.g., reduced outage durations or improved energy efficiency) are essential to quantify the tangible benefits of deploying FM applications in practice.

This shift means that companies need new competencies that go beyond traditional ML and quantitative evaluation. Teams developing AI use cases with Generative FM must also have the skills to design and interpret user studies and implement feedback loops. Close collaboration between domain experts and developers or UX designers is required to ensure that FM applications deliver clear benefits and are trusted by their users ([8,11]).

FMs offer significant flexibility due to their two-phase training process, enabling broad generalisation capabilities and efficient adaptability to carry out specialised tasks. However, the required resources and evaluation methodologies differ significantly according to the specific data modalities and application contexts of the FM. Companies developing FM applications, therefore, need to carefully align their resource planning and skill sets with the specific demands of their data modalities and target use cases.

3.4. Cybersecurity Risks in Foundation Models

Foundation models have potential broad applicability in various domains, which also introduces unique safety and security risks and potential for misuse, in addition to the inherent risk of AI in systems. Table 4 summarises some of the currently identified security threats of FMs.

Table 4 – FMs Security Threats in the Energy Sector

Data Security and Privacy
• Protecting sensitive grid operational data and consumer energy consumption data from breaches • Anonymisation and differential privacy techniques for consumer data
Model Security ([2,6,7,9,11,20])
• Adversarial attacks by crafting malicious inputs to manipulate FM outputs, e.g., causing incorrect forecasts, hiding system anomalies, triggering inappropriate control actions • Model poisoning by injecting malicious data into training or fine-tuning datasets to compromise model integrity or introduce backdoors • Model theft through unauthorised access and extraction of the trained model, resulting in intellectual property loss and potential for reverse engineering attacks
Integration Risks ([6,12])
• New attack vectors can be introduced by integrating AI systems into critical OT infrastructure and industrial IT systems • API security vulnerabilities
Supply Chain Risks ([7,19,21])
• Reliance on third-party FM providers and potential vulnerabilities in their development or deployment processes • Lack of transparency into FM architectures or training data sources

([20]) Liu, S., Dong, L., Tian, C. and Xie, L., 2025. Unlocking Multi-Task Electric Energy System Intelligence: Data Scaling Laws and Performance with Limited Fine-Tuning. arXiv preprint arXiv:2503.20040.

([21]) Stanford Center for Research on Foundation Models, Stanford RegLab & HAI (2024). 'Request for Comment on the U.S. Artificial Intelligence Safety Institute's Draft Document: Managing Misuse Risk for Dual-Use Foundation Models'. Submission to Department of Commerce, National Institute of Standards and Technology, U.S. Artificial Intelligence Safety Institute (Docket No. 240802-0209, XRIN 0693-XC137).

Resilience and Availability ([6,7,9,11,12,19,20])
• Availability and continuous operation of FM-dependent systems • Protection against denial-of-service attacks targeting the AI infrastructure • Need for robust backup and failover mechanisms if the AI system is critical
Regulatory and Compliance Security Aspects ([2,9,11,12])
• Meeting cybersecurity standards mandated by energy sector regulators.

FMs are applicable across a wide range of applications and, due to their self-training process, can expose serious vulnerability for the systems they are integrated into, and become the single point of failure, as well as introducing additional ambiguity when applying generated outcomes outside their intended use cases ([20,23]).These major concerns with FM security, safety and trustworthiness are generally recommended to be addressed through reproducibility, third-party evaluation support and model monitoring usage post-development ([7,19]).

The FMs in grid and demand-oriented production can be exposed to cyberattacks from various aspects and different scales, including but not limited to local entities, control devices in the grid, grid operators and national distributed entities and energy markets. Exposed vulnerabilities of legacy systems due to integration of FMs create opportunities for cyberattacks and manipulation of technical equipment and automated control actions, as well as the local and regional forecasts that can have severe consequences and impact on disturbances in the local grid. Any manipulations of data within an FM-based grid operator extend the scale of the consequences to the regional and national levels. The time series FMs for energy pricing, tariffs, distributed resource scheduling and charging coordination cybersecurity vulnerabilities would result in market costs manipulation and energy trading functions, with potential risk on the local, regional and national level, including but not limited to the manipulation of the grid sensor feeds, grid topology data, and outage indicators.

For FMs in the energy sector, cybersecurity is not merely a technical add-on but a crucial requirement. Secure data management, ownership of model development and control over computing infrastructure form a robust defence by ensuring data integrity and privacy, safeguarding model integrity and protecting the operational environment.

A secure data repository and sharing infrastructure for FM, as for all digital technology, can reduce and mitigate cybersecurity risks by directly addressing data integrity, privacy and adversarial attacks. Such infrastructure ensures reliable model training and accurate learning while safeguarding sensitive operational and personally identifiable data.

Having ownership and control over AI model development provides security throughout the entire lifecycle of FMs, from concept design to its application, including architecture and training process, and the enforcement and implementation of security practices. This direct control ensures model integrity and safeguards against adversarial attacks, maintaining the model's intended safe and reliable behaviour, specifically crucial for dual-use FMs. It can also enhance supply chain security by reducing reliance on potentially less secure third-party components. The initial ownership and control also facilitate the

embedding of ethical principles to actively prevent the generation of harmful outputs from the start.

Physical and operational security of the AI infrastructure can be ensured through control over the computing capacities to develop and deploy AI models, including FMs. The direct oversight enables physical security of hardware, preventing unauthorised access and malicious code injection, as well as resource attacks such as Denial-of-Service by allocating dedicated resources and implementing load management, which is critical for maintaining uninterrupted availability in real-time energy applications. Having control over the computing infrastructure allows for the creation of secure environments that protect sensitive development processes and deployed models from external cyber threats. It supports controlled deployment strategies, staged releases and gradual access expansion to monitor and mitigate risks before wider integration.

These elements are interdependent and vital for building and deploying trustworthy and resilient FMs as a part of larger AI systems in critical applications like grid and demand-side production. However, an in-depth investigation is a crucial prerequisite for identifying the requirements of integrating FM in an intended application and its connectivity and accessibility, as well as potential vulnerabilities of the future system.

3.5. Foundation Models Adoption in The Energy Sector

FMs are increasingly recognised as a promising and rapidly evolving technology that is poised to transform the energy sector. Different types of FMs are demonstrating capabilities suited for carrying out specific tasks within this sector. Language models, e.g., the GPT family ([22]), showing proficiency at natural language processing tasks, including extracting insights from energy reports, generating summaries of complex grid events, and knowledge management. Vision models, e.g., CLIP ([23]) and YOLO ([24]), are sophisticated in image recognition and can be applied to identifying equipment failures through visual inspections or analysing satellite imagery for solar panel monitoring tasks. Time-series models, often specialised transformer architectures ([25]), are crucial for energy forecasting, predicting load demand, renewable energy generation, and

([22]) Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

([23]) Radford, A., Kim, J. W., Xu, C., Perez, G., Hessel, C., Goh, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. arXiv preprint arXiv:2103.00020.

([24]) Terven, J., Córdova-Esparza, D.M. and Romero-González, J.A., 2023. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. Machine learning and knowledge extraction, 5(4), pp.1680-1716.

([25]) Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764.

grid conditions. Multimodal FMs can process and analyse historical data, weather patterns, and other relevant factors to generate more reliable forecasts than traditional methods. Anomaly detection capabilities enable FMs to identify unusual patterns in grid data, model complex energy networks, optimise grid topology, and potentially indicate cyberattacks, equipment failures, or other disruptions ([6,26]).

FMs are transitioning from general language and vision tasks to specific, complex and often multimodal problems in the energy sector. There is significant effort in research and development from academic institutions, national labs, relevant industries and technology providers that is highly related to creating FM for the energy sector. However, there is no publicly released and operationally ready FM for the energy sector. Most current models are still at the research and prototype stages, where significant challenges for their successful deployment remain to be addressed ([9,10,12,20]), ranging from technology maturity and infrastructure requirements to ambiguity in frameworks for collaboration governance and regulatory compliance.

3.5.1. Key Areas to Address for FM Integration in the Energy Sector

While FMs show promising potential to carry out complex tasks, it is crucial to address their limitation in the energy sector to address the challenges and avoid over-generalisation. Some of the identified key areas to consider for FM integration are presented in Table 5.

Table 5 - Key Areas for FM Integration in the Energy Sector

Data Strategy and Infrastructure	
FMs are data-hungry; a lack of high volume, diverse, clean, accessible and integrated data (due to various reasons such as the unwillingness to share the data by data owners, or lack of comprehensive data governance and regulatory restrictions, e.g. GDPR) is a major barrier. Infrastructure must support massive training, fine-tuning, and real-time inference	
Focus	Assessing the current data landscape, identifying data gaps, establishing data pipelines, governance, quality standards and required compute and storage infrastructures for development and implementation
Use Case Identification & Prioritisation	
Identifying high-value problems FMs can solve across grid operations, planning and demand-side management; evaluating technical feasibility, potential ROI, and risk	

([26]) Gao, Y., Wang, Z., & Zhao, J. (2022). A review on anomaly detection in power systems based on machine learning. Energy, 254, 124277.

Focus	Investing in areas with the greatest potential impact and feasibility, starting with less critical applications if necessary. Support for startups and continuity of the promising projects in the prototyping phase (some mentioned in this report)
Model Development, Fine-tuning, and Management	
Generic FMs will not perform optimally without domain-specific fine-tuning. A lifecycle management plan is needed for continuous improvement and adaptation to changing systems in operation.	
Focus	Strategies for acquiring or building base FMs, fine-tuning them for specific energy tasks, ongoing model monitoring, retraining, and version control. Deciding on buy vs. build vs. partner approach – outlined in the previous section.
Integration with Existing Systems	
FMs provide insights; these insights must trigger actions within existing energy management systems to provide value. This is often complex due to legacy systems	
Focus	Planning how FM outputs (forecasts, recommendations, anomaly alerts) will be integrated into operational technology (SCADA, EMS, DMS) and information technology (asset management, billing, customer platforms) systems
Talent Development & Organisational Structure	
Successful FM adoption requires a blend of AI expertise and in-depth knowledge of specific areas in the energy domain. Organisational silos can hinder progress in cross-disciplinary team formation and collaboration.	
Focus	Identifying required skill sets (AI/ML engineers, data scientists, ML/AIOps specialists, domain experts with AI understanding), training needs and how teams should be structured to support FM development and initiatives
Security & Risk Management	
FMs introduce new attack vectors. Compromising AI systems could have severe consequences for grid stability and data privacy. Security must be proactive and built in.	
Focus	Developing robust cybersecurity frameworks specifically for AI systems in critical infrastructure, addressing data security, model security (adversarial attacks, poisoning), integration risks, and compliance
Regulatory & Policy Engagement	
The energy sector is highly regulated. Regulatory uncertainty or unsupportive policies can hinder innovation. AI use in critical systems may require new regulatory approaches.	

Focus	Proactively engaging with regulators to inform them about FM capabilities and risks, discuss potential policy implications (e.g., data sharing, market rules, security standards) and advocate for supportive frameworks, providing clear practical guidelines for development
Explainability, Trust and Accountability	
Stakeholders need to trust the system's recommendations. In the event of errors or incidents, understanding why something happened is crucial for diagnostics, liability and regulatory compliance.	
Focus	• Developing methods and policies to ensure AI/FM decisions can be understood, validated and trusted by human operators and stakeholders, especially in critical applications. • Defining accountability for FM design, development, integration, and operation.

3.5.2. Technical FM Integration Roadmap

From a technical perspective, the direct application of FMs to complex industrial tasks, such as grid operations and demand-oriented production, is limited by challenges related to industrial data characteristics, quality, sensitivity and specificity, performance reliability and trustworthiness, computational efficiency, and adaptation to inherently complex, dynamic and multimodal environments ([6,10,11]). A structured roadmap that can address the existing challenges is presented in Table 6.

Table 6 – Technical FM Integration Roadmap in Grid and Demand-Oriented Production Applications

Phase 1 – Data Foundation for Grid and Demand	
Purpose	• To address multimodality, heterogeneity, scale, real-time availability and quality of industrial data ([6,10]). • To implement strong data governance, privacy-preserving techniques (federated learning, secure enclaves) and access controls for sensitive/proprietary data ([6,11]).
Activity	Establish robust, secure and scaleable data pipelines for collecting and curating diverse, high-resolution data from various sources relevant to grid and demand-oriented production ([6,8]). This includes: - Real-time and historical grid state data (SCADA, potentially high-resolution synchro-waveform from WMUs) ([7,8]) - Demand and generation time series data (load, smart meters, renewable forecasts, dispatch signals) at various granularities and locations (energy data are highly regionally dependent) ([10]) - Multimodal sensor data (images, audio, vibration) from grid infrastructure and generation assets ([6,10]) - Exogenous variables (weather, market prices) ([10])

	Industrial documents and knowledge bases related to grid operations and equipment ([6])
Outcome	Large, pre-processed and machine-ready, integrated, and governed multimodal datasets tailored for grid and demand-oriented applications
Key stakeholders	Data providers, utilities and grid operators, academics, industry R&D
Phase 2 – Base Model Training for Spatio-Temporal Grid Dynamics	
Purpose	To address the limitations of general FM architectures for causal, spatio-temporal, multimodal industrial data ([6,9,12]) To manage the high computational demands of training with efficient architectures and distributed systems[6,11])
Activity	- Develop and train novel multimodal FM architectures specifically designed for spatio-temporal grid data, leveraging insights from innovation representation model (IRM)-based and GNN approaches ([10,12]) - Use self-supervised pre-training objectives (e.g., masked modelling, dual-view contrastive learning) that capture spatio-temporal dynamics, hierarchical dependencies and underlying physics ([10,12]) - Develop physics-informed models (integration of physics principles into AI models) for robustness and embed physics-based constraints (e.g., power flow equations, grid dynamics) directly into the training process (e.g., via loss functions) to ensure consistency ([6,12])
Outcome	An Industrial Base Model capable of understanding general patterns in grid state, demand, and generation data, with inherent spatio-temporal and physical inductive biases
Key stakeholders	FM developers, academics, cloud service providers (CSPs)
Phase 3 – Adapting for Grid & Demand Applications	
Purpose	To ensure accurate and reliable adaptation to diverse grid scenarios and specific utility requirements ([7,12]) To manage domain-specific knowledge, avoid model overfitting and enhance interpretability ([10,12])
Activity	- Adapt the Industrial Base Model using task-oriented and domain-specific fine-tuning methods ([6,20]) - Fine-tune for core grid and demand management tasks like OPF, load/renewable forecasting, anomaly detection, and over-current protection using paired task descriptions and input/output data ([6,10,12,20,27]) - Embed domain-specific knowledge (grid topology, operational constraints, equipment specifications) and integrate relevant knowledge graphs for domain adapter fine-tuning ([21]) - Explore lightweight adaptation techniques (few-shot, prompt-based) suitable for real-time adjustments and resource-constrained environments ([6,11])
Outcome	A set of Task-Oriented and Domain-Specific FMs specialised for critical grid and demand-oriented production applications, validated on representative datasets
Key stakeholders	Downstream developers/deployers (utilities), FM developers, academics
Phase 4 – Deployment & Real-time Operation	

([27]) NIST (U.S. AI Safety Institute) (2024). Managing Misuse Risk for Dual-Use Foundation Models. U.S. Department of Commerce, National Institute of Standards and Technology, NIST AI 800-1 Initial Public Draft. Available from: https://doi.org/10.6028/NIST.AI.800-1.ipd.

Purpose	To ensure high real-time inference performance, address data overload from sensors, integrate into heterogeneous computing systems, maintain trustworthiness and interpretability in operational settings ([6,10,12])
Activity	- Deploy the adapted FMs into real-time grid operation environments ([21]) - Integrate FMs into existing SCADA, energy management zsystems (EMS), and distribution management systems (DMS) ([6]) - Develop industrial agent systems combining FMs for perception, planning, memory and action, enabling human-agent and machine-agent interaction ([20]). - Implement robust monitoring systems to ensure model performance, reliability, and security during deployment ([2,6,12]).
Outcome	FM-powered systems actively supporting real-time grid control, demand management, generation forecasting and automated monitoring and safety, enhancing operational efficiency and resilience
Key stakeholders	Utilities/grid operators, downstream developers/deployers, FM developers, academics, regulators, third party auditors and evaluators

The success of the interlinked phases described in the roadmap is a step towards enabling the long-term vision and the final phase of widespread, integrated deployment of an FM across the entire industrial product lifecycle in the energy sector. This goes beyond individual applications to encompass R&D, manufacturing, testing, management and maintenance processes ([11]).

3.6. Foundation Model Stakeholders

The deployment of FMs in the grid and demand-oriented production is a complex task that goes beyond addressing the technical and infrastructure requirements. The current state of FM development also indicates an increased tendency for partnerships to overcome the hurdles of FM data hunger, high computational demands, large capital investment, and cross-domain and -disciplinary AI literacy and expertise ([28]). The development and deployment of FMs in the energy sector require the active involvement of various stakeholders who have distinct roles in both current and future scenarios, as presented in Table 7.

Table 7 – FM Development and Deployment Stakeholders.

Core Developer and Research Entities	
FM Developers	Companies and startups (e.g., OpenAI, Google, Anthropic, Siemens) that develop and distribute FMs, often invest significantly in their training and deployment ([1,2])
Academia and Research Labs	Crucial for theoretical advancements, developing new architectures, and exploring new applications for Industrial FM or GridFMs ([7,10,22]), and often contributing to open-source initiatives ([11])
Energy Sector Operators and Partners	
Electricity Utilities and Grid Operators	They are the primary end-users, managing their own private and siloed data and grid-specific requirements. They directly benefit from fine-tuned FMs without needing to develop AI models ([10])

[28] *Competition & Markets Authority CMA (2024) AI Foundation Models: technical update report.*

Aggregators, Vendors, and Demand-Side resource	Involved in the control and management of DERs and inverter-based resources (IBRs), facing challenges with data visibility and real-time analysis ([10])
Engineers and Policymakers	Utilise FMs for power system modelling, operations, energy management and strategic planning ([10,16])
Infrastructure and Support Providers	
Cloud Service Providers	Crucial for providing the vast computational resources needed for FM training and deployment, potentially leading the model deployment ([2,7])
Energy Software Vendors	Critical to enable integration of FM into legacy IT and OT systems (SCADA, EMS, DMS, ADMS providers)
Hardware Manufacturers	Essential for developing AI-specific hardware and edge computing devices to support lightweight FM deployment in industrial settings ([6]).
Regulatory and Governance	
Government Agencies and Regulators	Responsible for setting standards, ensuring safety, security, privacy, and promoting fair competition in the AI market. Developing robust regulatory frameworks and addressing potential issues like market concentration and misuse ([2,10,24]).
Standardisation Organisations	Develop standards for AI system evaluation and performance (e.g. ISO, IEEE). Develop frameworks and guidelines for independent auditing
Investment Community	
Provides the significant funding required for FM development, including for smaller startups	
Civil Society Organisations and Third-party Evaluators	
Play important roles in assessing misuse risks and promoting responsible AI development and bringing the needs and inputs for continuous improvements after deployment	

3.6.1. Collaboration of Foundation Model Stakeholders

To accelerate the development and realise the deployment of FMs in the energy sector, regardless of specific application, a joint effort and cross-collaboration among a wide range of stakeholders across the AI supply chain and energy ecosystem is essential. Table 8 presents the stakeholders and their contributions aligned with the phases outlined in the technical roadmap.

Table 8 – Stakeholders' Contributions with Respect to Position in the Technical Roadmap.

Industry, Developers, and Deployers	
Action	Systemic collection of diverse operational data and generating high-fidelity synthetic data to overcome the scarcity of labelled data for specific industrial tasks. Invest in developing and training industrial base models and task-specific IFMs ([11,12]). Implement robust security measures and transparent practices throughout the FM lifecycle ([2,7,20,22]). Deploy FMs responsibly, starting with low-risk applications and scaling gradually based on rigorous testing ([2,11]).
Contribution to roadmap	Lead Phase 2 (Model Training), Phase 3 (Adaptation), and Phase 4 (Testing) by developing and validating specific FMs. Execute the long-term deployment vision in real-world energy systems.

Collaboration	Partner with industry to access real-world data (while ensuring privacy/security) and validate research findings on operational systems ([6,9,11]). Engage with civil society and regulators to keep them informed on research priorities and findings ([2,11,19]).
Academia, Researchers, and Open-Source Initiatives	
Action	Conduct fundamental research into FM theory, architecture and training objectives tailored for multimodal industrial data and mechanism embedding ([20,21,22,24]). Develop rigorous evaluation methodologies and open benchmarks for industrial FM capabilities, including interpretability and robustness ([2,11,12,19]). Lower the barriers to entry for smaller developers ([2]).
Contribution to roadmap	Drive Phase 1 (Data) and Phase 2 (Model Training) by creating open datasets, researching novel architectures and developing training techniques. Lead Phase 4 (Testing) by developing evaluation methods and conducting independent audits.
Collaboration	Partner with industry to access real-world data (while ensuring privacy/security) and validate research findings on operational systems.[6,9,11,20] Engage with civil society and regulators to keep them informed on research priorities and findings ([2,11,19,20]).
Government and Regulators	
Action	Establish clear, adaptive regulatory frameworks focusing on safety, security and trustworthiness for dual-use FMs in critical infrastructure like energy ([2,11])
Contribution to roadmap	Primarily influence Phase 4 (Testing) and the Long-term Vision (Deployment) by defining standards, testing protocols, and certification pathways. Provide funding for public research infrastructure (compute, data) ([7]).
Collaboration	Coordinate across relevant agencies (NIST, DOE, FERC, CMA) ([7]), engage with industry and academia to understand technical capabilities and risks ([2,11,19,20])
Civil Society Organisations and Expert organisations	
Action	Provide feedback on the usability, performance, and societal impacts of deployed FM-powered energy applications ([2,11,19]). Advocate for transparency, accountability and ethical development practices ([2,11,19,20]).
Contribution to roadmap	Inform requirements in Phase 3 (Adaptation) and Phase 4 (Testing) by providing user feedback. Their concerns and needs should guide the entire long-term deployment vision.
Collaboration	Participate in user studies and provide domain expertise for data annotation and model evaluation. Engage in dialogue with developers and regulators on concerns and priorities ([2,7,23,25]).

Stakeholder collaboration can be enhanced and facilitated through the establishment of mechanisms for sharing information about the models' capabilities, risks and mitigations, as well as standardised reporting that covers computational, energy and deployment strategies. The joint development can be promoted through joint academic and industry projects to bridge the gap between theoretical breakthroughs and application needs. Formation and support for independent and third-party evaluation and auditing provide the capabilities for verification of FM capabilities, establishing independent third-party auditing and evaluation can enable FM capabilities and limitations to be verified. This oversight provides an accountability framework for dependable collaboration for responsible FM development.

3.7. Core Recommendations

In the current state and advancement of FM development, proactive preparation and strategic and impact-oriented investment are crucial to effectively leveraging the potential of this new technology while mitigating associated risks and ensuring a responsible and secure deployment ([4]). The findings of this study do not suggest the immediate deployment of FM at scale; instead they focus on building the essential prerequisites, such as data, initial understanding of the use cases, a security foundation, and collaborative and support frameworks, as outlined in Table 9.

Table 9 – Core Recommendations for FM Adoption in the Energy Sector.

Prioritise Domain-Specific FM Development
To build experience and demonstrate value, identify high-potential and lower-risk use cases (e.g., improved load forecasting for non-critical applications, or specific asset anomaly detection) where FMs can be piloted using existing or soon-to-be-available data. Use-case selection requires assessment and prioritisation that looks beyond technology readiness level and market analysis and extends to cover factors in the design and development of the concept of operation phases. It should include feasibility and technical requirements as well as interdependencies for integration into the existing systems, and potential security concerns, and societal consequences.
Access and Enhance Data Readiness
FMs cannot deliver value without high-volume, diverse, high-quality, accessible data. Conduct a comprehensive audit of existing data sources (SCADA, AMI, DER data, weather, market, asset history), identify data quality issues and gaps, and develop a phased plan for data cleaning, integration, and establishing secure, scaleable data pipelines that can feed potential FM applications.
Advance Energy Sector Ecosystems
Work to standardise, collect, process and make available diverse and high-quality energy data, especially at the distribution level. Utilise secure and privacy-preserving technologies to facilitate data sharing, provide a trusted environment for validating and certifying the FM-based systems' integration into different applications (e.g., energy test and evaluation facilities and hubs) and collaborative training across entities.
Develop Security-by-Design Framework for AI/FM
Security in critical infrastructure cannot be addressed after the development. Implement a robust risk management framework and integrate cybersecurity planning from the very beginning. Avoid FM as a single point of failure by Including threat modelling for inherent AI-specific risks (adversarial attacks, data poisoning) as well as FMs dual use risks. Plan for secure data handling, model deployment, and integration points. Ensure compliance with relevant energy sector cybersecurity regulations.
Reinforce Collaboration and Governance Reinforcement
Establish strong partnerships among stakeholders. Promote open-source initiatives to democratise access and reduce development costs. Develop clear governance policies and accountability frameworks for FM development and deployment to ensure an ethical and safe operation.

4. Task 4: Flexibility and Heat Reuse Potential of Data Centre

4.1. Executive summary

This document provides a comprehensive analysis of the waste heat recovery and flexibility potential of data centres, focusing on technical, economic, and policy aspects. It covers data centre energy use, cooling technologies, heat pump technologies, district heating integration, flexibility and sector coupling, techno-economic assessments, and real-world business cases, with the aim to support sustainable energy transitions in the EU.

4.1.1. Data centre sector overview

Data centres are critical to the digital economy but consume substantial energy, with cooling systems currently accounting for up to 40% of their total energy use.

Energy demand is rapidly increasing, driven by AI, cloud computing, and hyperscale deployments, with data centres projected to represent 3-4% of global electricity consumption by 2030. In the EU, the current electricity demand for data centre is of the order of 55 TWh/y, corresponding to about 2.2 % of overall electricity consumptions. This share is expected to grow to 15-30 % by 2035: assuming a linear growth and the most conservative case, this would correspond to a 7.5 % share and hence to 187.5 TWh/y by 2030.

Data centres are categorised into enterprise, co-location, hyperscale high-performance computing (HPC), cryptocurrency, and edge data centres, each with different characteristics and heat reuse potentials. Enterprise and crypto data centres are often most suitable for heat reuse due to full control over IT equipment and cooling systems.

The share of high-power density racks (> 20 kW per rack) is currently about 10 %, but it might grow to 20-70 % by 2030. Since high-power density is related to liquid cooling, these shares can be considered roughly valid for the latter as well.

4.1.2. Cooling technologies

Cooling is essential to maintain optimal IT equipment temperatures. Air cooling remains dominant but is limited in handling high heat fluxes, leading to increased adoption of liquid cooling technologies such as direct-to-chip and immersion cooling. Liquid cooling enables higher waste heat temperatures (up to 65°C),

improving heat recovery potential and reducing power usage effectiveness (PUE) to below 1.1, compared to around 1.6 for conventional air cooling.

Waste heat temperatures vary by cooling technology and chip operating temperatures, influencing the need for heat pumps. Higher chip temperatures reduce heat pump requirements but may increase IT power consumption slightly, requiring trade-offs in system design.

4.1.3. Heat pumps

Heat pumps (HPs) are key for upgrading data centre waste heat to temperatures suitable for district heating networks, typically around 90°C. Electrically driven compression heat pumps operate on refrigerant cycles, with performance measured by coefficient of performance (COP). COP depends strongly on heat source and sink temperatures, with typical values close to 3 for heat recovery from air cooling, but able to arrive to 7 or more with liquid cooling.

Various refrigerants are used in heat pumps, constrained by environmental regulations such as the F-gas phase-out policies (with PFAS as an upcoming concern). Natural refrigerants like ammonia and carbon dioxide partly avoid these issues, but have to be properly evaluated in terms of temperature limits and performance. Heat pump configurations impact efficiency and flexibility, with turbo-compressors enabling fast modulation beneficial for grid services.

4.1.4. Integration with district heating

District heating (DH) is the most common and effective method for distributing recovered heat from data centres, even though local heat recovery examples exist. DH networks vary by generation, with newer generation systems operating at lower temperatures (< 70 °C) than older ones (> 90 °C), facilitating easier integration of low-grade waste heat. The temperature and distance between data centres and DH networks influence the need for heat pumps and connection costs.

Heat recovery projects typically involve a heat recovery substation (including heat pumps, heat exchangers and controls) and a connection pipeline to the DH network. Investment and operational costs vary widely depending on system size, technology, electricity prices and local conditions. Moreover, contractual agreements are needed, similarly to other waste heat recovery applications.

4.1.5. Economic aspects

Levelised cost of heat (LCoH) is used to assess economic feasibility, accounting for capital expenditures (CAPEX) and operational expenditures (OPEX).

Operational costs include electricity costs and are affected by system efficiency. Due to the large variety of possible heat recovery combinations, LCoH values range broadly from around 26 €/MWh under favourable conditions to over 135 €/MWh under less favourable ones. Based on this range, a reference LCoH of 68.50 €/MWh was identified as representative. Heat recovery from air-cooled data centres is much more economically challenging than from liquid cooling, but remains viable especially if the temperature to be supplied to the DH network is not too high (order of 70 °C).

4.1.6. Flexibility and sector coupling

Data centres can provide flexibility services to support power grid stability through demand response, load shifting and participation in frequency regulation markets. Techniques include activating uninterruptible power supply (UPS) battery systems and temporarily shutting down or pausing servers. Examples include Google's carbon-aware computing and Microsoft's grid-interactive data centres. The inclusion of large-scale heat pumps for heat recovery introduces an additional degree of freedom for flexibility.

Sector coupling involves integrating data centres with other energy sectors, such as heating, cooling and power to optimise resource use and reduce emissions. Challenges include technical complexity, regulatory alignment, investment needs and data privacy concerns.

4.1.7. Examples of business cases

Several real-world examples demonstrate successful waste heat recovery and flexibility integration. Many identified cases highlight the technical feasibility and environmental benefits of integrating data centre waste heat into local energy systems, though barriers, especially related to economic feasibility in the case of air-cooled data centres and distant networks, have been identified.

4.1.8. Recommendations

The last chapter of this report presents recommendations and suggests possible support measures which could be applied to the sector (e.g., at policy level) to foster the diffusion of waste heat recovery applications and the application of flexibility measures for data centres.

4.2. Introduction

4.2.1. Background

Digitalisation and artificial intelligence (AI) have the potential to make the future energy sector much more operationally efficient, though as a relatively new technology there are still many questions about the costs, benefits and optimised use of AI in the energy sector. The Strategic Roadmap for Digitalisation and AI in the Energy Sector was announced in the Affordable Energy Action Plan as a plan to incorporate AI in the energy sector, but analyses to answer some of the aforementioned questions are first necessary to fully develop this roadmap. The EnTEC2 study, articulated in multiple Tasks, will inform the Strategic Roadmap through a cost-benefit analysis, market analysis and reports on the incorporation of AI in specific areas. In particular, this report addresses Task 4, focused on the flexibility and heat reuse potential of data centres.

Data centres are a rapidly growing sector. Their operation requires large amounts of electricity. Such energy goes dissipated into heat within electronic equipment, which then requires cooling to maintain proper operating temperatures. The overall efficiency of these systems can be increased exploiting two opportunities: waste heat (WH) recovery towards nearby district heating (DH) networks and demand response or similar ancillary services for the electric grid. The natural link between these two opportunities is provided by heat pumps (HPs), which are typically needed to raise the waste heat temperature to useful levels. HPs can also be used for flexible operation to support the power grid in certain cases.

This report evaluates these opportunities in EU data centres. It examines the business case for these possibilities, highlighting the flexibility and waste heat reuse opportunities within EU data centres. It showcases best practices and analyses the conditions that facilitate the adoption of these services by the data centre industry. Ultimately, the report provides specific recommendations for EU actions to enhance flexibility and promote waste heat reuse in the data centre sector.

4.2.2. Objectives

The study is focused on analysing the incorporation of digitalisation and AI – particularly AI – in the energy sector. This includes market analysis, cost-benefit analysis and sector-specific applications. In all analyses, the study aims to identify concrete recommendations for the EU on optimising this promising new technology in the energy sector.

The study has three main objectives:

- **O1**: Assess the waste heat recovery potential of data centres, especially in connection with district heating.
- **O2**: Assess the flexibility potential provided by data centres to the electricity grid.
- **O3**: Identify possible actions the EU could take to promote these efficiency measures.

The report is arranged in 9 chapters:

1. Introduction
2. Data centre sector overview
3. Overview of data centre cooling technologies
4. Relevant heat pump technologies
5. Connecting data centres and district heating
6. Flexibility and sector coupling
7. Overall techno-economic estimates
8. Examples of business cases and best practices
9. Recommendations

4.2.3. Methodology

This section describes the methodology used to collect the technical information reported in chapters 2-8. In general, sources were identified based on starting with the authors' experience combined with AI-powered web search. Each source suggested by AI was later directly checked. All reported information is based either on the authors' original content or on citations from specified sources.

Chapter 2, Data centre sector overview. In this chapter, different data centre categories, data centre trends and an overview of their energy use are presented based on long experience with these matters and external references, like JRC (“Energy Consumption in Data Centres and Broadband Communication Networks in the EU”) and LBNL (“2024 United States Data Center Energy Usage Report”) reports and sector consultants.

Chapter 3, Overview of data centre cooling technologies. The main logic driver for this chapter consisted in identifying the available waste heat temperatures for the different cooling technology. In this respect, investigating the current diffusion of the various technologies and the trends for dominant solutions was also considered necessary. This input was needed for the planned waste heat recovery techno-economic analysis. Key literature used as a reference for this chapter includes reports from ASHRAE (“Emergence and Expansion of Liquid

Cooling in Mainstream Data Centers"), NREL ("Best Practices Guide for Energy-Efficient Data Center Design"), besides the aforementioned JRC and LBNL reports.

Chapter 4, Relevant heat pump technologies. Since heat pumps are usually a key component for heat recovery from data centres, a dedicated analysis was carried out in this chapter. The analysis assessed available heat pump sizes and operating temperature ranges in order to explore their compatibility with the various cooling technologies and the main expected "user" of the recovered heat, i.e., district heating. Moreover, environmental aspects related to refrigerant types were considered. The key references for this chapter are provided by Agora/Fraunhofer ("The roll-out of large-scale heat pumps in Germany") and IEA-HPT (Annex 58 Final Report) reports. A few single heat pump manufacturers were also checked to validate the general information collected. They were not explicitly mentioned to avoid commercial bias.

Chapter 5, Connecting data centres and district heating. In this chapter, reusing waste heat from data centres is described, both in the proximity of the data centre and at greater distances through connection with district heating networks. The content is based on a mix of long experience with these subjects and external references.

Chapter 6, Flexibility and sector coupling. In this chapter, opportunities for flexible operation of the data centres to support the power grid are described together with other possible sector couplings such as reusing excess heat and making use of natural gas flaring from landfills. This chapter is mostly based on external references but also on the authors' experiences in research and development projects.

Chapter 7, Overall techno-economic estimates. This chapter is devoted to the economic analysis (investment and operating costs) for waste heat recovery from data centres to district heating. The operating conditions assessed in the previous chapters constituted the inputs for the technical estimates. Heat pump costs were collected from the Danish Energy, the aforementioned Agora/Fraunhofer report and academic literature. District heating costs were taken from the standard textbook by Frederiksen and Werner, "District heating and cooling". All costs were updated and referred to EU average through Eurostat databases.

Chapter 8, Examples of business cases and best practices. In this chapter, specific existing or planned cases are presented, both related to flexible operation of data centres and to sector coupling such as reusing excess heat and making use of natural gas flaring from landfills. The content is mostly based on external references, but also includes interviews with personnel at the sites. Details concerning specific business cases are considered confidential and could not be shared with the authors of this report.

Finally, recommendations reported in chapter 9 were elaborated evaluating and integrating key opportunities and barriers identified within single topics.

4.3. Data centre sector overview

4.3.1. Introduction to data centres

Data centres are foundational to Europe's digital economy, supporting cloud services, computational power and a digital infrastructure. They require substantial amounts of energy to operate, making it particularly important for them to efficiently utilise any excess heat, to contribute to grid stability, and to carefully plan other sustainability measures.

The data centre sector in western Europe is well developed and has a large selection of colocation data centres and service providers. The highest density of data centres in Europe is found in FLAP-D (Frankfurt, London, Amsterdam, Paris, and Dublin). These cities are characterised by robust infrastructure and connectivity, and strategic geographical advantages. The number and size of data centres have increased continuously all over the globe and are expected to continue to do so in the future as demand for data storage and computational power increases, e.g. for AI. Open maps of existing data centres exist on the web ([1]), though with limited reliability.

Data centres fall into several different categories, depending on the resources they provide and how the business model of the data centre is designed. The sections below describe some typical data centre categories.

4.3.1.1. Hyperscale data centres

These are usually the largest data centres. The operator owns and is responsible for everything both within and around the data centre. It is usually part of a worldwide network of several similar data centres. IT hardware, cooling systems and software are also often developed in house and tailored to the specific services running in the hyperscale data centres.

4.3.1.2. Co-location data centres

Co-location data centres are usually large data centres too, where several customers rent parts of the data centre or servers for their operations. This type of data centre therefore has several different customers that can consist of everything from banks to individual small companies that use different types of computing power. The structure of the different actors can vary. In some cases, the data centre itself handles everything, while in other cases the customer takes care of their own servers. In order to facilitate power grid support and effective

([1]) See e.g. https://www.datacentermap.com/ and https://map.datacente.rs/ (visited August 2025).

heat reuse, co-location data centres should incorporate these considerations into their client agreements.

4.3.1.3. Enterprise data centres

An enterprise data centre belongs to a company or authority and handles its own data. It can, for example, be a data centre that handles all data from a large industrial company such as salaries, documents or production data from the industry.

4.3.1.4. Edge data centres

Edge data centres are located close to the user and are usually small. They are primarily used to handle calculations with a high demand for speed, typically when data transfer time between the user and data centre significantly affects user experience. Such cases include control of industrial processes, autonomous systems and VR/AR/AI ([2]) applications. These data centres are usually not suitable for flexible operation.

4.3.1.5. High performance computing (HPC) data centres

HPC data centres are used for complex, advanced calculations. They can be located at universities and research institutes, but also in companies. These data centres are often used for the training of AI before it is distributed for inference. The availability requirements for an HPC data centre are not as high as for a regular data centre because the calculations are usually not time-dependent to the same extent.

4.3.1.6. Cryptocurrency data centres

Cryptocurrency data centres are equipped with HPC hardware and cooling systems designed to secure and validate digital currencies like Bitcoin. Crypto data centres are similar to HPC data centres in the sense that they both provide high-computing capacity. Both types of data centres perform calculations that can be paused and resumed without major problems and there is no great requirement for response time. These inherent characteristics render these data centres well suited for contributing to grid stability.

([2]) Virtual reality, augmented reality and artificial intelligence.

4.3.2. Data centre energy use

The energy use of data centres is driven primarily by two components:

1. Running the computer equipment: Including servers, storage and networking devices. This accounts for most of the energy use and scales with computational demand, especially in AI and high-performance computing environments.
2. Cooling Systems: Responsible for maintaining optimal operating temperatures. Cooling can consume up to 40% of a data centre's total energy, depending on the technology and climate conditions.

Globally, data centre energy use is rapidly increasing, where the growth in electricity consumption is largely driven by AI, cloud computing and hyperscale deployments. In the USA, data centres are projected to account for 12% of total electricity demand by 2028, up from 4% in 2023 ([3]). Power availability has become a strategic constraint and bottleneck. Data centre owners are increasingly prioritising onsite power generation (e.g., fuel cells, microgrids, diesel generator sets) to reduce dependency on strained grids and to accelerate deployment timelines. Some hyperscale data centre campuses now require over 1 GW of power, equivalent to the electricity needs of 800,000 homes ([4]). On an EU level it is calculated that if all planned data centres are built, they could increase Europe's total power demand by up to 30% over the next decade. Even under conservative assumptions, a 10–15% increase is expected, due to AI-driven infrastructure expansion according to Goldman Sachs ([5]).

The Joint Research Centre reports (6) that data centres in the EU used an estimated 45–65 TWh of electricity in 2022 (1.8–2.6% of total EU electricity use), while telecommunication networks used an estimated 25–30 TWh of electricity (1–1.2% of total EU electricity use).

Even with progress in data processing efficiency, data centre growth will result in increased global electricity and water consumption as well as generation of waste heat. As illustrated in Figure 2, Goldman Sachs projects that global data centre electricity consumption will increase from 400 to 1060 TWh between 2023 and 2030 and will represent 3-4% of global electricity requirements in 2030.

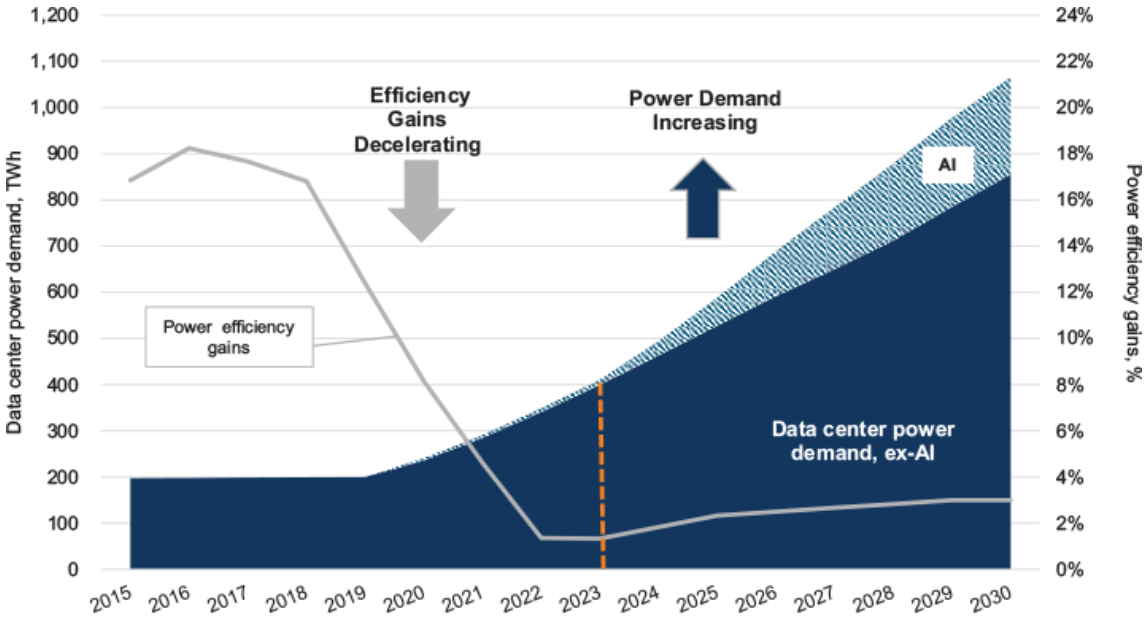
([3]) Shehabi et al., "2024 United States Data Center Energy Usage Report", Lawrence Berkeley National Laboratory, 2024.

([4]) www.datacenterdynamics.com/en/whitepapers/2025-data-center-power-report-mid-year/ (visited August 2025).

([5]) https://www.goldmansachs.com/insights/articles/data-centers-could-boost-european-power-demand-by-30-percent (visited August 2025).

([6]) Kamiya and Bertoldi, "Energy Consumption in Data Centres and Broadband Communication Networks in the EU", Joint Research Centre (JRC), 2024.

Figure 2 – Projected global data centre electricity consumption



Source: Goldman Sachs ([7]), "AI/data centers' global power surge: Five drivers of upside/downside and the Reliability investment tailwind", Fig. 4.

Power efficiency gains have decelerated in recent years (grey curve), while data demand is continuously growing. This is driving a surge in datacentre power use, with AI acceleration expected to continue. (Note: The term *power demand* in the figure refers to demand of electric energy.)

To give a national, concrete example from the EU, data centres in Sweden are estimated to consume between 4.0–4.4 TWh in 2025 and then increase to 4.4–5.0 TWh (3–4% of the national electricity use) by 2030 ([8,9]). This estimate, however, may turn out to be on the low side, as the Swedish Energy Agency report was published before the need for AI hardware had taken off. The development of AI is expected to accelerate and take an increasingly larger share of total data centre energy use. Indeed data centres are entirely driven by user needs and requests, a significantly growing demand with the rapidly increasing use of AI and large language models.

In 2024, the European Commission adopted regulations for rating the operations of data centres. In general, they require DC operators to ([10]):

- Measure their energy efficiency.

([7]) Goldman Sachs, "AI/data centers' global power surge: Five drivers of upside/downside and the Reliability investment tailwind", 2025. Available online: https://www.goldmansachs.com/pdfs/insights/goldman-sachs-research/five-drivers/FinalReport.pdf (visited August 2025).

([8]) Swedish Energy Agency 2023: Energianvändning i digitala system, datacenter och kryptovaluta. ER 2023:04.

([9]) Wallin, M., Werner, R., & Olofsson, S. (2020). *Datacenter i Sverige 2020-2025.* Radar Ecosystem Specialist.

([10]) https://www.districtenergy-digital.org/districtenergy/q1_2025/MobilePagedArticle.action?articleId=2037059#articleId2037059 (visited August 2025)

- Show how much of their energy emanates from renewable energy.
- Describe how the data centre reuses waste heat.
- Measure the effectiveness of cooling.
- Show the effectiveness of carbon use.
- Report their use of fresh water.

Data centre companies would be required to evaluate heat recovery potential as part of their planning and siting process. This consideration would involve both the design of the cooling systems as well as location of the data centre relative to potential heat loads. There can be multiple economic and environmental advantages to siting data centres closer to potential users of surplus heat.

Environmental performance standards could be developed for key performance parameters, e.g.:

- Energy Reuse Factor (ERF) – the amount of reused energy divided by the total amount of electrical energy supplied to the DC servers.
- Power Usage Effectiveness (PUE) – this is a commonly used factor that is equal to the total energy of the data centre divided by the IT energy consumption.
- Percentage of rejected heat that is productively used.
- Water Usage Effectiveness (WUE), typically measured in litres/kWh (water consumption per kWh of electricity used).
- Performance standards could be applied on an individual facility basis or on the company's data centre operations as a whole.

4.3.3. Data centre trends

Today (existing inventory), high-density racks make up a small portion of data centres. The vast majority of data centres have moderate rack power (on average ~5–8 kW/rack), and high power densities are uncommon ([11,12]). An industry survey in 2020 showed that only ~16 % of organisations had racks over 20 kW (and only ~10 % had racks over 30 kW) ([13]). Even in 2024/25, racks >30 kW per

([11]) "Data Center Trends: Rack Density Rises While PUE and Outage Frequency Remain Flat", 2025, https://www.upsite.com/blog/data-center-trends-rack-density-rises-while-pue-and-outage-frequency-remain-flat/ (visited August 2025).

([12]) Navigating Data Center Power Density in the Age of HPC and AI", 2024, https://www.tierpoint.com/blog/data-center-power-density/ (visited August 2025).

([13]) "Data Center Trends: Rack Densities and Outages On the Rise", 2021, https://www.raritan.com/blog/detail/data-center-trends-rack-densities-and-outages-on-the-rise (visited August 2025).

rack are rare, found almost exclusively in hyperscalers/HPC facilities ([14]). In other words <10 % of today's total racks are "high-density" according to the >20–30 kW definition (mainly in AI/HPC environments).

To give a relatable consumer example, a ChatGPT query typically needs nearly 10 times as much electricity to process as a normal Google search, according to research published by Goldman Sachs. Requirements for data processing are growing exponentially, having increased by an average of 120 % annually since 2010 – from 2 to 120 zettabytes, according to the World Economic Forum (1 zettabyte = 1 trillion gigabytes).

In ~10 years (2035): The proportion is expected to increase significantly in line with the AI boom. Already, about a third of operators are planning to build high-density rack capacity, and almost half of data centres anticipate that their average rack will soon exceed 11 kW ([15]). Over the next decade, AI and HPC demands will drive up power density – many new data centres will contain significantly more racks above 20–30 kW. Exact forecasts vary, but it is reasonable to expect a sizeable minority (perhaps 30–50 % of racks) to be considered high-density by the mid-2030s. For context, McKinsey predicts that ~70 % of all computing capacity by 2030 will be used for AI computation ([16]), indicating that high-density solutions will then become commonplace (as AI racks are often in the 20–40 kW range).

In ~30 years (2055): Today's high-density racks are expected to be the norm in most data centres. The trend is that as servers and chips become more power-hungry; almost all major data centres will need to support high power densities by 2050^{+}. The industry assumes that high density will eventually become the standard ([15]) – there is strong evidence that the majority of racks around mid-century will be at 20 kW or well above. In short, the share of high-density racks is expected to increase from today's ~10 % to a significant share within 10 years and further to constitute the majority of data centre infrastructure in ~30 years (driven by the fact that almost all heavy IT workloads will then require high density).

Looking in particular at the trends in HPC and e.g. Nvidia which can be considered in the forefront of GPU development, the trends show that in a couple of generations there will be rack setup with a maximum power usage up to 1 MW, which is approaching the same power density as in a nuclear reactor. For this equipment the possibility for air-cooling is long gone; it will require liquid cooling (see next chapter). Along with this, the power distribution needs a paradigm shift.

([14]) "Uptime's 15th Annual Global Data Center Survey Results Shows Both Commitment and Hesitancy as Industry Plans for Wider AI Usage, Climate Change Reporting, and the NVIDIA Revolution to Come", 2025, https://uptimeinstitute.com/about-ui/press-releases/uptimes-15th-annual-global-data-center-survey-results-shows-both-commitment-and-hesitancy (visited August 2025).

([15]) "Why Density per Rack is Going Up", https://phoenixnap.com/blog/rack-density-increasing (visited August 2025).

([16]) "Data center demands", 2025, https://www.mckinsey.com/featured-insights/week-in-charts/data-center-demands (visited August 2025).

Generally, power density is going up and the design point temperature is going down, which will increase the challenges for cooling and ability to for heat reuse.

A final emerging trend to be noted is the possible switch to optical solutions for data centre networking ([17]). Data centre networking consists of the connection between different racks, the speed of which is becoming increasingly important for specific calculations. Photonic networking, based on optical switches and fibres, can greatly increase such speed, at the same time reducing electricity dissipation and thereby energy consumption. However, this is expected to be mostly used to increase computational density and efficiency, and not to decrease overall energy consumptions as such.

4.3.4. Data centre sector summary data

The current electricity demand for data centres at EU level is of the order of 55 TWh/y, corresponding to about 2.2 % of overall electricity consumption (average of ranges provided by JRC). This share is expected to grow to 15-30 % by 2035 (Goldman and Sachs): assuming a linear growth and the most conservative case, this would correspond to a 7.5 % share and hence to 187.5 TWh/y by 2030. This will be compared to the district heating demand in chapter 5.

The share of high power density racks is currently about 10 %, but it might grow to 70 % by 2030 (McKinsey). Most conservative estimates expect a 40 % share by 2035, corresponding to a 20 % share by 2030 with a linear growth assumption. The prediction uncertainty is shown by these large differences.

4.4. Overview of data centre cooling technologies

Proper cooling is crucial for IT equipment. In its absence, electronic devices can overheat and cause problems of varying severity, like service disruptions, data loss, increased energy costs, emissions and operating costs and other types of damage. According to ASHRAE ([18]), around 2018, after a period “where significant performance increases were delivered, generation over generation, accompanied by modest and predictable power increases”, large power increases in the compute, memory, and storage subsystems of IT equipment started challenging the data centre sector. The corresponding high thermal power fluxes to be removed from chips made air-based cooling insufficient for certain

([17]) “Lightspeed: Photonic networking in data centers”, 2025, https://www.datacenterdynamics.com/en/analysis/lightspeed-photonic-networking-in-data-centers/ (visited August 2025).

([18]) ASHRAE, “Emergence and Expansion of Liquid Cooling in Mainstream Data Centers”, white paper, 2021.

data centre applications, thereby causing the progressive spread of more efficient liquid-based solutions.

This chapter presents a concise review of such cooling technologies. Data centre cooling is a sector where research is continuously in progress, as witnessed by the number of published papers. Searching with the text string “review cooling data center” in the Elsevier journals database revealed that 9 review papers on data centre cooling technologies had been published in the last 2 years only. The 3 most prominent review papers on liquid cooling published in 2025 ([19]) contain in total about 660 references.

4.4.1. Introduction to cooling technologies

Liquid cooling technologies can be broadly classified in two main categories, namely air-based cooling and liquid-based cooling, each coming with some variants. Though hybrid solutions are also possible, this classification mainly refers to cooling solutions adopted at the server level (some air cooling at room level can anyway be possible).

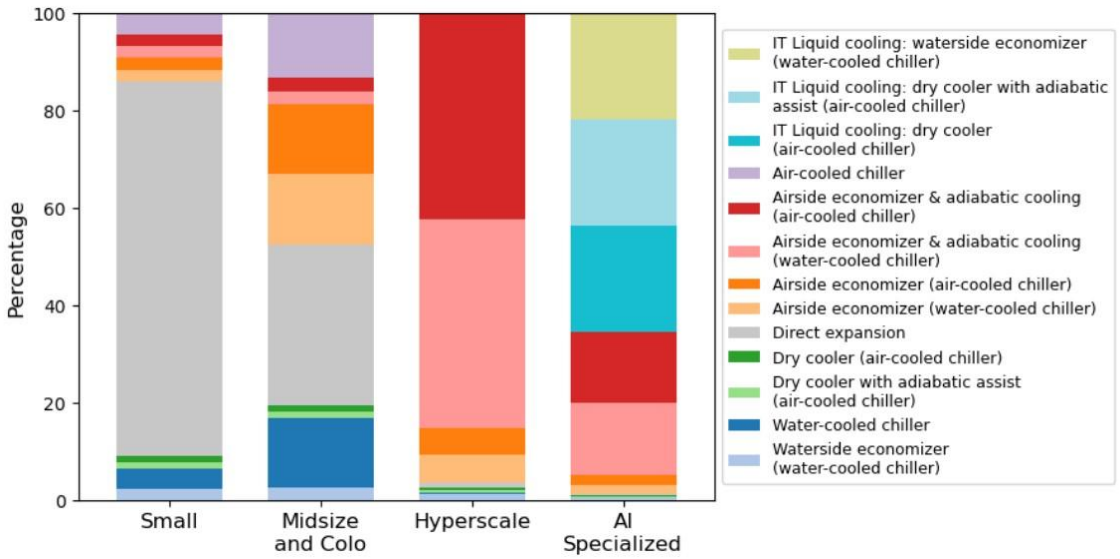
Air-based cooling is the first developed and still dominant method. It is the most suitable solution for limited power fluxes, where it can provide the required cooling at the cheapest costs. However, increasing CPU/GPU densities make it unsuitable for some emerging cases, where air cooling is insufficient when operated at usual temperatures (order of 25°C) or not economically convenient if operated at lower temperatures. Consequently, liquid-based cooling solutions are emerging. The typical power intensity of servers actually depends on the type of data centre (see previous chapter). High power fluxes are mainly present in HPC and AI applications, which in turn are more common in large-scale colocation or hyperscale data centres. With reference to the US market, as reported by LBNL ([20]), more than 60 % of servers were delivered for such data centres in 2024, with a share of 70 % expected by 2026. Figure 3 shows the percentages of cooling technologies for some data centre types. Due to the complexity of the data centre ecosystem this is not enough to offer a specific quantitative share for liquid cooling, but the emerging trend of the latter technology is evident. Within Europe, a representative case is provided by Germany where, according to current projections, the market share of liquid cooling is rising rapidly from 8 % in 2021 to a predicted 33 % in 2028 ([21]).

([19]) Dlzar Al Kez *et al.*, *Sustain. Energy Techn.*, **82**, 104511 (2025); Shuai Zheng *et al.*, *Appl. Therm. Eng.*, **272**, 126385 (2025) ; Zhijun Wu *et al.*, *Chem. Eng. Sci.*, **310**, 121525 (2025).

([20]) A. Shehabi et al., “2024 United States Data Center Energy Usage Report”, Lawrence Berkeley National Laboratory (LBNL), 2024.

([21]) KPMG, https://kpmg.com/de/en/insights/business-performance-and-resilience/three-success-factors-for-the-future-of-german-data-centres.html (visited February 2026).

Figure 3 – Percentage of cooling solutions in different data centre types.



Source: LBNL, "2024 United States Data Center Energy Usage Report", Fig. 4.2 (2024).

Estimated percentage of various cooling solutions for different types of data centres. See next sections for more details about cooling technologies. It can be seen that liquid cooling is largely dominant for AI.

The threshold for the transition from air to liquid cooling was estimated by ASHRAE at a CPU power of about 300-400 W. It is also useful to refer to rack power. Racks typically include multiple servers, where each server can be equipped with more than a CPU/GPU. For usual data centres, the rack power limit is about 5-8 kW. HPC and AI can require rack powers in the range of 20-80 kW, according to "Research and Markets" firm ([22]). The threshold for the usage of liquid cooling is at rack powers of the order of 10 kW. Liquid cooling technology providers are already addressing rack powers of 100-150 kW, with a upward trend.

Besides being able to manage higher power fluxes, liquid cooling can also improve the PUE. The cooling liquid indeed operates at higher temperatures than air (between 40-60°C), possibly allowing for direct heat rejection without the need of compression chillers. Moreover, liquid heat transfer is generally more efficient than air heat transfer (the convective heat transfer coefficient for liquids is much higher than for gaseous media, like air). While conventional data centres have a PUE of the order of 1.6, the claimed PUE for most efficient data centres exploiting liquid cooling is below 1.1. In contrast, the declared PUE is sometimes ambiguous (with liquid cooling a "partial" PUE, pPUE, being often implicitly used), especially if coupled to heat recovery (applications see below).

The main cooling technologies can be classified as follows:

([22]) Research and Markets, report 5311272, https://www.researchandmarkets.com/reports/5311271/u-s-data-center-cooling-market-landscape-2024 (visited August 2025).

- Air cooling
    - Server fans (coupled to CRAC or CRAH, see below)
- Liquid cooling
    - Direct-to-chip
        - Single-phase
        - Two-phase (either heat pipe or thermosiphon)
    - Immersion
        - Single-phase (either enclosed chassis or open bath)
        - Two-phase (either enclosed chassis or open bath)

Further details and hybrid classes can be introduced. For example, rear-door heat exchangers can handle approximately 85–90 kW per rack and are sometimes combined with direct-to-chip solutions to upgrade existing air-cooled halls without extensive modifications ([23]). In practice, hybrid solutions combining air and liquid cooling are often necessary, as some components continue to require air cooling. Consequently, a combination of liquid and air cooling is expected to remain standard in AI-focused data centres for the foreseeable future ([24]).

4.4.2. Air-based cooling

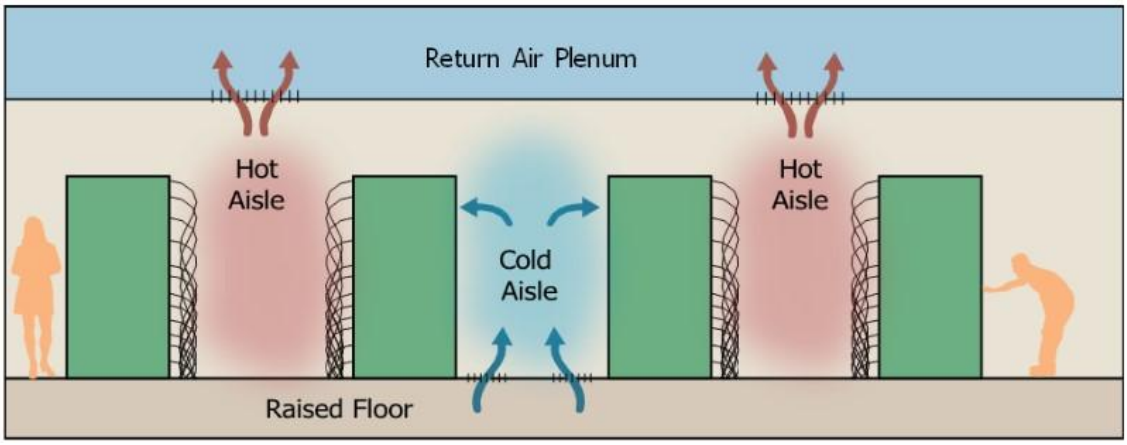
Air-based cooling is still the dominant solution. It is also likely to remain a significant part of the market in the future, as liquid cooling solutions typically do not capture all of the generated heat and need some form of additional air-cooling.

Air-based cooling occurs in multiple stages. First, some airflow (typically driven by fans at server level) cools the electronic equipment within the racks. High-power components like CPUs are typically connected to some finned heatsinks (a metallic block exchanging heat with the air flow). Highly thermally conductive thermal paste is used to ensure the contact between the CPU and the heatsink. Second, a proper air flow is ensured at room level, with fans and ducts arranged in order to favour convection among rack rows (see Figure 4). Third, the air is cooled by rejecting heat outside of the building with an external cooling system.

([23]) https://datacentrereview.com/2025/03/the-four-techniques-you-need-to-know-to-cool-ai-data-centres/ (visited August 2025).

([24]) Optimizing AI Infrastructure: The Critical Role of Liquid Cooling, Victor Avelar and Wendy Torell, 2025.

Figure 4 – Air cooling arrangement in data centre rooms.



Source: NREL, "Best Practices Guide for Energy-Efficient Data Center Design", Fig. 4.1 (2024).

Typical arrangement for air circulation in data centre rooms. Rack rows are alternated by hot and cold aisles. Raised floor and possibly false ceiling allow for proper air distribution. Hot and cold aisles can also be sealed by physical separators, in order to improve flow efficiency.

The main distinction among possible room cooling systems is typically made between CRAC (computer room air-conditioning) and CRAH (computer room air handler) systems ([25]). Both systems rely on a combination of compression chillers, economisers and air-water heat exchangers (e.g., in rooftop or various dry cooler versions). The main difference is, however, that CRAC are smaller modular units, while CRAH are centralised units which typically offer higher efficiencies.

Due to the rather continuous operation of data centres, the design of these systems typically benefits of some oversizing with respect to conventional units for other types of buildings. Variable speed drives and similar regulating systems are therefore important to guarantee efficiency.

Several additional details and efficiency measures can be taken, also depending on local climate (e.g., the possibility to implement free cooling or evaporative/spray cooling at the external air interface depending on external conditions). More recent and hybrid variants of air cooling are given by in-row air cooling and rear-door heat exchanger cooling, which place cooling coils (typically with some liquid refrigerant) directly between racks or on their rear side.

While the room temperature admits a certain degree variability, a reference operating temperature for air-based cooling systems can be considered 25°C, with variations again depending on season and climate. As such, the resulting waste heat is definitely low-temperature waste heat and needs a heat pump in

([25]) Otto Van Geet and David Sickinger, "Best Practices Guide for Energy-Efficient Data Center Design", National Renewable Energy Laboratory (NREL), funded by US Federal Energy Management Program (FEMP), 2024

order to be reused in common district heating systems (with a possible exception being provided by 5th generation district heating and cooling systems, see below).

4.4.3. Liquid-based cooling

The two main classes of liquid cooling are direct-to-chip (DTC) and immersion cooling. Both can come with single-phase or two-phase processes. For immersion cooling, moreover, the use of a dielectric (i.e., electrically insulating) fluid is necessary. Actually, this can be recommendable for DTC applications as well, in order to avoid any issue in case of leakages.

Liquid-based cooling can significantly lower the PUE. It is important to distinguish between energy efficiency and power capacity: systems with low PUE can still require high peak power. For example, immersion cooling offers both high capacity and energy efficiency, while cooled doors deliver high capacity but not the same efficiency.

While not strictly belonging to liquid cooling at chip level, in this context it is also worth mentioning projects of underwater/undersea cooling ([26]). An underwater data centre houses servers in watertight, container-like pods submerged in the sea or ocean. It uses cold seawater for natural cooling, eliminating or reducing traditional air-conditioning system and hence strongly improving PUE. Successful operation has been demonstrated for periods of several months.

Finally, when high temperatures are available at the primary cooling circuit, one can also consider exploiting such heat to drive an absorption chiller ([27]). This concept is, however, not further discussed here, as not related to WH recovery.

4.4.3.1. DTC - Direct-to-chip cooling

In direct-to-chip cooling, the coolant circulates through cold plates in direct contact with components. If non-dielectric fluids are used, the sealing of circuits is crucial. From this point of view the usage of natural convection systems can be beneficial, as they are based on a two-phase process (evaporation of the fluid in the liquid phase at the CPU plate and condensation of the fluid in vapour phase at an external heat exchanger connected to an external circuit) which does not require circulation pumps and can be completely sealed. This the case for heat pipes as well as for thermosiphons: in heat pipes a wick facilitates the fluid circulation independently of the circuit orientation, while in thermosiphons natural circulation occurs by gravity and hence the circuit orientation, with a proper tilt, is crucial.

([26]) https://brightlio.com/underwater-data-centers/ (visited August 2025).

([27]) See, e.g., A. Haywood et al., “Thermodynamic feasibility of harvesting data center waste heat to drive an absorption chiller”, *Energy Convers. Manage.*, **58**, 26 (2012).

Figure 5 – Example of direct-to-chip cooling.



Source: LIFE4HeatRecovery project.

Example of a prototype for an innovative two-phase (thermosiphon) direct-to-chip cooling developed in the LIFE4HeatRecovery project, 2024 ([28]). The copper pipes and plates contain the refrigerant.

DTC liquid cooling can be operated at relatively high temperatures, typically up to 60°C. The higher is the temperature, the easier is to reject heat towards the external environment (typically through an external secondary circuit); correspondingly, the easier is the heat recovery as well. On the other hand, the limiting temperature must consider the overall thermal resistance of the server circuit. Chip suppliers typically provide a “case temperature” limit (related to the actual chip temperature limit through laboratory tests). The cooling system must guarantee that such a temperature is not exceeded.

4.4.3.2. Immersion cooling

Immersion cooling submerges entire servers, achieving even higher waste heat temperatures—up to 70°C has been recorded in laboratory settings. Similarly to DTC cooling, immersion cooling can also exploit single-phase or two-phase conditions. Moreover, for both of them two main casing options exist: either the enclosed chassis option or the open bath option. The open bath option poses higher challenges especially for the case of two-phase cooling, where vapours can enter the room environment. Maintenance can anyway be expensive (even lifting servers from the bath can require a crane or multiple operators). An image of an immersion cooling installation is shown in Figure 6.

([28]) LIFE4HeatRecovery project, funded by LIFE programme, contract number LIFE17 CCM/IT/000085, website https://www.life4heatrecovery.eu (visited August 2025).

Figure 6 – Example of immersion cooling.



Source: Hypertec.

Image posted in the article “Liquid Cooling in the Data Center Is a Good Thing”, by Bill Kleyman, Data Center Frontier, 2025 ([29]).

Immersion cooling has been shown to reduce energy consumption by up to 50% compared to air cooling ([30]). This technology is particularly attractive for AI-driven data centres with high packing densities and low PUE, although it requires significant changes to operations and design.

4.5. Relevant heat pump technologies

Even though data centre cooling can be operated at relatively high temperatures (see previous chapter), it is usually not possible to apply direct heat recovery towards conventional district heating networks (with supply temperatures of about 90°C). Hence, heat pumps (HPs), and in particular electrically driven compression heat pumps, are expected to play a crucial role in this sector. This chapter therefore focuses on these machines, starting with the key point of performances and of how the latter depend on operating temperatures. Then, market readiness, refrigerants, compressors and related evolving regulation are also discussed, in order to provide a broader overview of this technology. Finally, a connection with PUE and the possible allocations of HP electricity consumption are addressed.

Comments about heat recovery towards new-generation thermal grids, where lower operating temperature can allow data centre heat to be recovered without heat pumps, are reported in the next chapter.

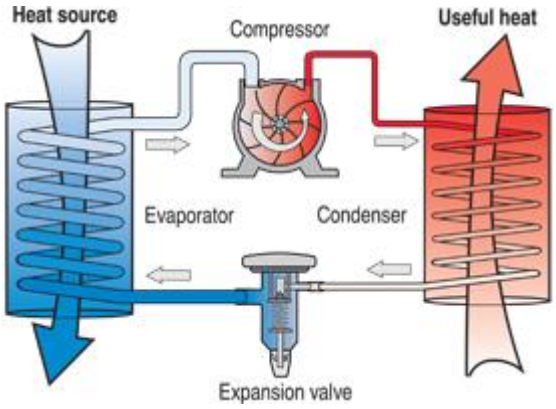
([29]) https://www.datacenterfrontier.com/sponsored/article/55266936/liquid-cooling-in-the-data-center-is-a-good-thing (visited August 2025).

([30]) https://info.pcxcorp.com/blog/top-five-benefits-of-immersion-cooling (visited August 2025).

4.5.1. Heat pump performance metrics and qualitative dependence on temperature

Compression heat pumps are machines able to absorb heat and deliver it at a higher temperature by the same thermodynamic principle of a fridge. The process needs electricity. A refrigerant goes through a cycle (see Figure 7), typically involving phase transitions: the refrigerant in gas phase is compressed to liquid phase, increasing its temperature, then passed through a heat exchanger (condenser) to transfer heat to a heat-sink. After condensation, the liquid goes through a phase change to gas by an expansion valve. The expansion decreases its temperature, and another heat exchanger (evaporator) absorbs a heat source, before starting the cycle again. The heat source could be air or a liquid circuit, and the heat-sink could be air or a liquid circuit. An electric motor is normally used to drive the compressor to compress and circulate the refrigerant. More complex cycles, possibly avoiding phase transitions, exist.

Figure 7 – Heat pump scheme.



Source: IEA-HPT, https://heatpumpingtechnologies.org/articles/heat-pump-work/.

Representation of the heat pump internal refrigerant cycle. The refrigerant goes through a cold heat exchanger (where it absorbs heat from an external fluid acting as the heat source), enters the compressor (where the refrigerant temperature increases), goes through a hot heat exchanger (where it supplies heat to another external fluid acting as heat sink), enters the expansion valve (where the refrigerant temperature decreases) and finally returns to the cold heat exchanger.

The key parameters to evaluate the performance of a HP are the **COP** (for heating uses) and the complementary **EER** (for cooling uses). It is crucial to understand that **these parameters depend on the operating temperatures** on the source and sink side (the higher the temperature lift, the more electricity is consumed by the HP). Consequently, the performance of a HP in a data centre heat recovery application depends on the temperature at which the cooling technology operates.

The exact definition and the dependence of COP and EER on temperature are explained with approximate formulas in the following subsection, for readers interested in efficiency measurement details.Using these approximate COP formulas, one can estimate the COP values for **scenarios of interest**, as

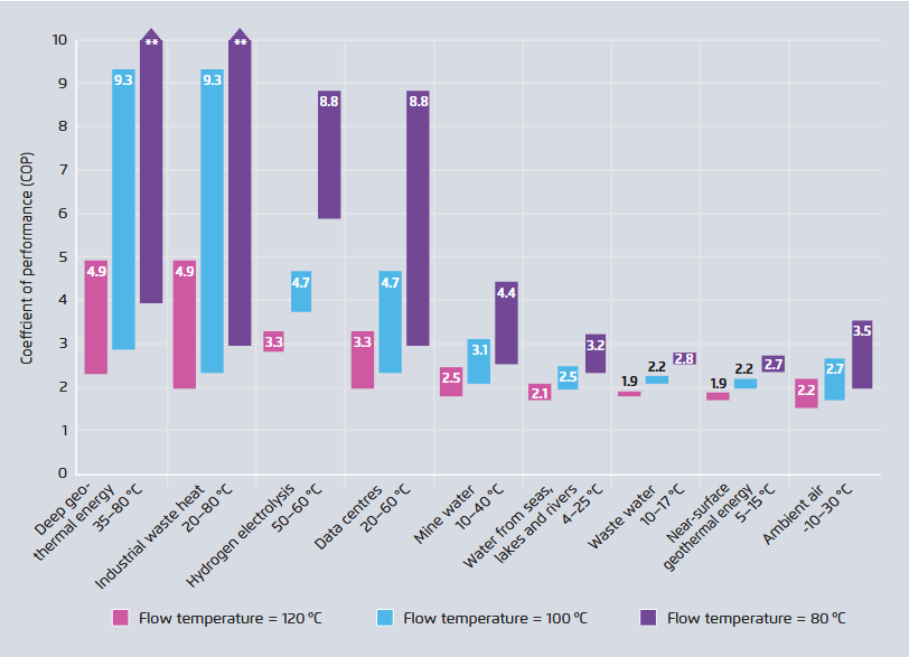
reported in Table 10 below. Scenarios S1-S3 correspond to 3 different source temperatures (60°C, 40°C, 20°C, which can be interpreted as high-temperature liquid cooling, medium-temperature liquid cooling, and air cooling), all for a relatively high supply temperature (90°C, corresponding to conventional district heating). Scenarios S4-S5 correspond to the same source conditions, but to a lower supply temperature (namely 75°C, corresponding to recent and more efficient DH networks). For all scenarios, three values of the efficiency factor (40%, 50%, 60%) are considered (the efficiency factor contributes to the COP calculation and is defined in the next subsection). The performance of real machines is expected to lie in this range (more likely towards the lower end in the experience of authors), with variations depending on compressor and refrigerant type. Hence, the provided Carnot-like estimates can be used as lower and upper bounds for the performances of real HPs. Moreover, a broader overview of heat pump performances for different types of heat sources and a visualisation of the COP dependence on temperature lift (i.e., temperature difference between heat source and sink) are shown in Figure 8 and Figure 9.

Table 10 – Estimated COP values for some operating conditions

Carnot-like estimates for COP based on various efficiency factors η = 40, 50, 60% for different ranges of source temperatures (inlet temperature on cold side, T_C) and sink temperatures (outlet temperature on hot side, T_H).

Variable	Unit	S1	S2	S3	S4	S5	S6
T_C	°C	60	40	20	60	40	20
T_H	°C	90	90	90	75	75	75
COP (η = 40 %)	-	4.8	2.9	2.1	9.3	4.0	2.5
COP (η = 50 %)	-	6.1	3.6	2.6	11.6	5.0	3.2
COP (η = 60 %)	-	7.3	4.4	3.1	13.9	6.0	3.8

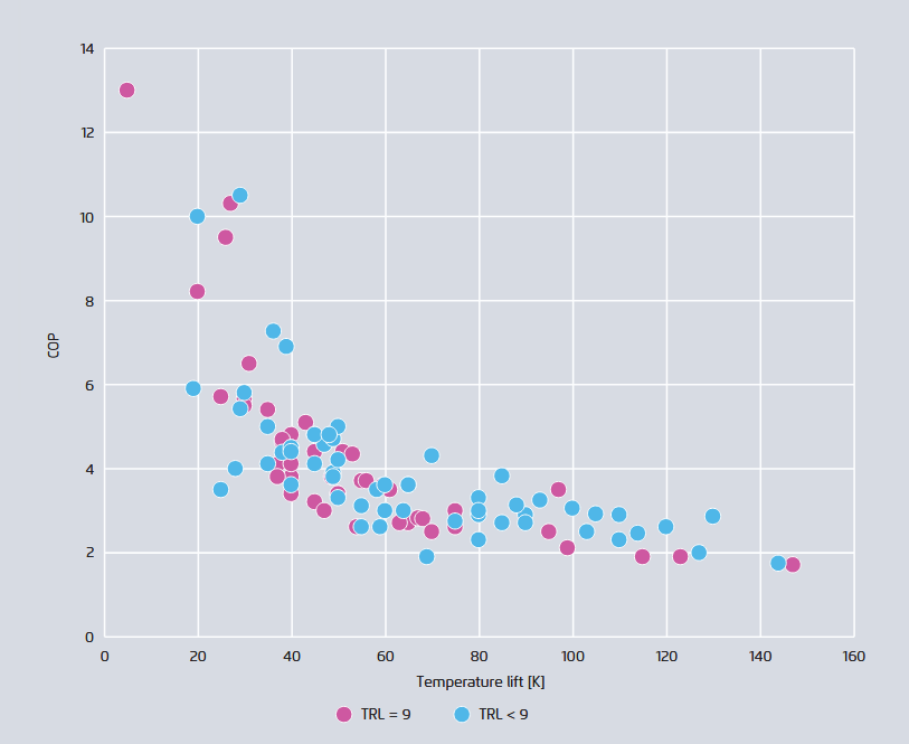
Figure 8 – COP dependence on source.



Source: "The roll-out of large-scale heat pumps in Germany", 2023, Fraunhofer / Agora (2024), Fig. 22.

Dependence of COP on source type and temperatures. Overall HP efficiency assumed to be 50%, COP estimates based on Carnot-like approximation.

Figure 9 – COP dependence on temperature lift.



Source: "The roll-out of large-scale heat pumps in Germany", 2023, Fraunhofer / Agora (2024), Fig. 29.

Dependence of COP on temperature lift for HPs with different level of technology readiness level (TRL).

4.5.1.1. Heat pump COP and EER

As anticipated above, the HP performance is typically measured by two slightly different metrics, depending on whether the focus is on heating or cooling mode. Exactly the same thermodynamic process is used for HP and cooling machines. The performance for heating is expressed in terms of the "coefficient of performance" (COP), given by the ratio between the delivered thermal energy (heat-sink) on the hot side (the condenser side) $E_{th,H}$ and the electricity consumed by the HP $E_{el,HP}$:

$$\mathrm{COP} = \frac{E_{th,H}}{E_{el,HP}}.$$

The performance for cooling is expressed in terms of the "energy efficiency ratio" (EER), given by the ratio between the absorbed thermal energy (heat source) on the cold side (the evaporator side) $E_{th,C}$ and the consumed electricity E_{el}:

$$\mathrm{EER} = \frac{E_{th,C}}{E_{el,HP}}.$$

Neglecting thermal losses (as usually acceptable for these machines), energy conservation yields $E_{th,H} = E_{th,C} + E_{el,HP}$ and hence

$$\mathrm{COP} = \mathrm{EER} + 1\,.$$

The above formulas show that the less electricity consumed for a given useful heating (or cooling) power, the high the HPE COP (or EER).

Moreover, for basic thermodynamic reasons, the HP performance is related to the operating temperatures: the higher the temperature lift between the cold and the hot side, the higher the electricity consumption and hence the lower the performance. While the detailed calculation of the COP as a function of temperatures for a given HP requires complex models, it is possible to obtain acceptable estimates with relatively simple formulas. In order to provide an insight into the impact of data centre cooling and district heating temperatures on HP performances, an approximation related to the ideal Carnot cycle is presented here (see, e.g., the online tool https://grosswaermepumpen-info.de/en-us/products/cop/#calculator, supported by Fraunhofer IEG). The COP is calculated as a function of the cold temperature T_C of the heat source (equal to the data centre cooling outlet temperature) and of the hot temperature T_H of the heat sink (equal to the DH supply temperature):

$$\mathrm{COP}(T_C, T_H) = \eta \frac{T_H}{T_H - T_C}.$$

In this formula, temperatures must be expressed in K (T[K] = T[°C]+273.15). As it can be seen, the function includes an efficiency factor η. For η = 1, this expression corresponds to the COP of the ideal Carnot cycle, $\mathrm{COP}_{id}(T_C, T_H) = T_H/(T_H - T_C)$, which yields the theoretical limit for HP performances. Hence, $\eta = \mathrm{COP}/\mathrm{COP}_{id}$ can be interpreted as the ratio between the real and the ideal performance ([31]). For real applications, one then has $\eta < 1$, with typical values in the range 40-60 % for large scale heat pumps (see the aforementioned website). The efficiency factor takes into account irreversibility effects, related to losses occurring during compression, expansion and heat exchange: its variation depends on the different types of refrigerants used, compressor models and other technical solutions.

Regarding temperature dependence of performances, the simplified Carnot-like formula disregards the impact of inlet-outlet temperature differences in the fluid entering/leaving the heat pump. This can be addressed, still within a simplified framework, by other formulas, as discussed in the technical box below.

The extent of irreversibility (deviation from the ideal Carnot cycle) depends on the chosen HP components. On the other hand, for a given HP, it also depends on operating conditions. Normally, losses tend to be higher at the edges of the operating area. HPs should therefore be sized so that most operating hours fall well within the operating area, for high energy efficiency. For data centres and district heating, operating conditions are expected to be relatively stable, which limits this issue.

([31]) The same ratio is sometimes called SEI, system efficiency index. See K. Berglöf. “Method and guidelines to establish System Efficiency Index during field measurements on air-conditioning and heat pump systems”. SP Arbetsrapport :2014:38. The same reference also provides information on the real performance of cooling heat pump installations.

Technical box – Other COP formulas

Besides the Carnot-like formula reported in the main text, other more elaborated formulas exist. A second approximation also used in the online calculator mentioned in this chapter is related to the Lorenz COP. The Lorenz COP takes into account the inlet-outlet temperature difference at the evaporator and condenser circuits through the so-called logarithmic mean (temperatures must be expressed in K):

$$T_{LM,H} = \frac{\left(T_{out,H} - T_{in,H}\right)}{\ln\left(T_{out,H}/T_{in,H}\right)},$$

$$T_{LM,C} = \frac{\left(T_{out,C} - T_{in,C}\right)}{\ln\left(T_{out,C}/T_{in,C}\right)}.$$

In these formulas, the “in”/”out” subscripts refer to inlet and outlet values respectively, while “C”/”H” subscripts refer to the cold (evaporator) and hot (condenser) circuits. In the Carnot COP formula reported in the main text it is assumed $T_H = T_{out,H}$ and $T_C = T_{in,C}$. The Lorenz COP has an expression similar to the Carnot COP, but with T_H replaced by $T_{LM,H}$ and T_C replaced by $T_{LM,C}$. In practice, the Lorenz-like COP formula takes the form:

$$\mathrm{COP}_L\left(T_{LM,C}, T_{LM,H}\right) = \eta_L \, \mathrm{COP}_{id}\left(T_{LM,C}, T_{LM,H}\right) = \eta_L \, \frac{T_{LM,H}}{T_{LM,H} - T_{LM,C}}.$$

Consequently, while the Carnot-like formula only depends on the temperature of the fluids entering the HP, the Lorenz-like formula also depends on the temperature of the fluids leaving the HP. The inlet-outlet temperature difference can indeed significantly affect the HP performance, especially for HP adopting transcritical cycles. When using the Lorenz COP instead of the Carnot COP for the same HP, a different efficiency parameter (typically lower) should be used: in the authors' experience, one-stage HPs roughly correspond to η_L = 30 %, while transcritical HPs correspond to η_L = 40 % or higher.

More complex formulas, including for example also the effect of heat exchanger sizing, can be found in the literature. See, e.g., J.K. Jensen *et al.*, “Heat pump COP, part 2: Generalised COP estimation of heat pump processes”, Proceedings of the13th IIR-Gustav Lorentzen Conference on Natural Refrigerants, 2018).

4.6. Overview of commercially available heat pumps

The HP market currently offers a large availability of models. Considering the large scale of upcoming data centres, it is considered interesting to focus on large-scale HPs (with thermal power above 1 MW), though smaller sizes might also be used either for smaller data centres or for modular applications (i.e., combining multiple HPs in parallel).

A recent overview of large-scale HPs in Germany was published by Agora Energiewende, in collaboration with Fraunhofer IEG ([32]). This study shows a large availability of HP models, both in terms of sizes and temperature ranges:

- Sizes up to about 60 MW exist.
- Temperature supplies up to 250 °C are available, though most consolidated applications are closer to 100 °C. In particular, many producers offer HPs with supply temperature of the order of conventional DH temperatures, for all sizes.
- The range of admissible source temperatures is not investigated in detail in the report, but it is explained that many models are compatible with low temperature sources similar to air cooling or colder.

To achieve such a variety of sizes and conditions, different refrigerants and compressor solutions are exploited by producers.

Some producers of large-scale HPs also emphasise the machine flexibility in terms of minimum part load and modulation speed. The feasibility of variations of 25% of power in tens of seconds is claimed for machines based on turbo-compressors at sizes of 10 MW or above. This can be highly beneficial to provide ancillary services to the electricity grid. A minimum partial load of 50% is also common.

As discussed in the previous chapter, the range of source temperatures provided by liquid cooling can be much higher than air, possibly up to 50-70°C. Such a temperature range on the source side is more common for high-temperature heat pumps, some of which are anyway compatible also with district heating temperature (e.g., source temperature range 35-60°C, sink temperature range 75-130°C for a HP with a power of about 500 kW).

4.7. Heat pump types and configurations

This section provides some details about the most relevant heat pump components (refrigerants, compressors), as well as the efficiency impacts (system COP and PUE) of possible configurations for heat recovery from data centres to district heating.

4.7.1. Heat pump refrigerants

The working fluid selection is crucial in high-temperature heat pump design as it must align with process requirements to ensure optimal efficiency. However, the choice of refrigerant also involves environmental concerns, especially in

([32]) Agora Energiewende, “The roll-out of large scale heat pumps in Germany”, 2023.

consideration of its global warming potential (GWP). Compliance with the evolving F-gas regulation has to be taken into account.

The F-gas regulation aims to phase out F-gases, and for future-proof cooling and heat pump installations refrigerants with a GWP below 150 should be chosen. Converting heat pumps designed for banned refrigerants to new refrigerants is often possible, but results in poorer performance. It is also likely that refrigerants that contribute to PFAS (formed by leaking F-gases and which can contaminate groundwater via precipitation) will be banned. This leaves mainly natural refrigerants as future-proof alternatives. Heat pumps with ammonia refrigerant have been used in district heating networks since the 1990s and the technology is very mature and has a high COP, although with a limitation of about 90 °C towards the heat sink (condenser). Heat pumps with carbon dioxide (CO_2) refrigerant are still largely under development for small-scale machines, but are commercially available with some notable examples for large-scale machines. The choice of refrigerant can also be combined with cycle complexity, possibly introducing multiple compression stages for high temperature lifts. There is, however, a trade-off between energy efficiency and system complexity, as more compression stages and advanced features provide higher COP but increase the cost of the heat exchanger network. This aspect is addressed e.g. by Wallerand et al. ([33]), who also identify propane as one of the most advantageous working fluids (both economically and thermodynamically) for industrial applications.

4.7.1.1. Refrigerant for high-temperature heat pumps and IEA-HPT Annex 58

Concerning high-temperature HPs, hydrofluorocarbons (HFC)s such as R-245fa or R-365mfc are commonly used as working fluids today. However, their GWP is comparatively high, and in view of the prevention of global warming, these working fluids will be restricted in the foreseeable future. Synthetic refrigerants, including hydrofluoroolefins (HFOs) and hydrochlorofluoroolefins (HCFOs), have become key alternatives. HFOs such as R-1234ze(Z) and R-1336mzz(Z) enable high-temperature subcritical operation (up to 160 °C) with low GWP. Natural refrigerants such as water (R-718), ammonia (R-717), CO_2 (R-744) and hydrocarbons (R-600, R-601) provide further low-GWP options. For example, water is particularly suited for mechanical vapour recompression (MVR) steam applications at temperatures above 150°C, whereas CO_2 is ideal for transcritical hot water production up to 120°C. Ammonia is widely used for industrial heat pumps but is generally limited to about 110°C or lower, due to compressor constraints. Besides thermal suitability and environmental compatibility, further aspects which have to be considered when selecting a working fluid are safety,

([33]) A. S. Wallerand *et al.*, “Optimal Heat Pump Integration in Industrial Processes”, Applied Energy **219**, 68-92 (2018).

efficiency, availability and others. Table 11 gives a recent IEA-HPT overview ([34]) of the properties of present and future working fluids including its critical temperature T_{crit}, critical pressure p_{crit}, ozone depletion potential (ODP) relative to R-11, GWP and safety group classification (SG).

Table 11 – Refrigerants for high-temperature heat pumps.

Overview of suppliers and HTHP technology, sorted after maximum capacity.

Type	Working fluid	Description	T_{crit} (°C)	p_{crit} (bar)	ODP (-)	GWP (-)	SG
Natural	R-718	Water	373.9	220.6	0	0	A1
	R-717	Ammonia	132.3	113.3	0	0	B2L
	R-744	Carbon dioxide	31.0	73.8	0	1	A1
HC	R-601	n-Pentane	196.6	33.7	0	5	A3
	R-601a	Isopentane	187.8	33.8	0	4	A3
	R-600	n-Butane	152.0	38.0	0	4	A3
	R-600a	Isobutane	134.7	36.3	0	3	A3
	R-290	Propane	96.7	42.5	0	3	A3
HFO	R-1336mzz(Z)	1,1,1,4,4,4-Hexafluoro-2-butene	171.3	29.0	0	2	A1
	R-1234ze(Z)	cis-1,3,3,3-Tetrafluoro-1-propene	150.1	35.3	0	<1	A2L
	R-1336mzz(E)	trans-1,1,1,4,4,4-Hexafluoro-2-butene	130.4	27.8	0	18	A1
	R-1234ze(E)	trans-1,3,3,3-Tetrafluoro-1-propene	109.4	36.4	0	<1	A2L
	R-1234yf	2,3,3,3-Tetrafluoro-1-propene	94.7	33.8	0	<1	A2L
HCFO	R-1233zd(E)	1-chloro-3,3,3-Trifluoro-propene	166.5	36.2	0.00034	1	A1
	R-1224yd(Z)	1-chloro-2,3,3,3-Tetrafluoro-propene	155.5	33.3	0.00012	<1	A1
HFC	R-365mfc	1,1,1,3,3-Pentafluorobutane	186.9	32.7	0	804	A2
	R-245fa	1,1,2,2,3-Pentafluoropropane	154.0	36.5	0	858	B1
	R-134a	1,1,1,2-Tetrafluoroethane	101.1	40.6	0	1300	A1

A critical environmental consideration addressed in the IEA-HPT Annex 58 report is the formation of trifluoroacetic acid (TFA) from the atmospheric degradation of certain HFOs (e.g., R-1234yf). TFA is a persistent, mobile substance classified within the group of per- and polyfluoroalkyl substances (PFAS). Although current concentrations in the environment are considered low, TFA's persistence raises concerns about long-term accumulation in aquatic ecosystems. This has led to regulatory discussions in Europe about potential restrictions on PFAS-containing

([34]) B. Zühlsdorf et al., "High-Temperature Heat Pumps", IEA-HPT Annex 58 Final Report, 2024, https://etkhpcorderapi.extweb.sp.se/api/file/2874.

substances and a proposal of the European Chemicals Agency (ECHA) to regulate PFAS, covering nearly 10,000 different chemicals. While Annex 58 notes that the current risk to human health from TFA is minimal, it acknowledges that future policies targeting PFAS could impact the viability of certain synthetic refrigerants. As a result, there is increasing interest in natural refrigerants and in minimising leaks or improving refrigerant recovery practices to reduce environmental emissions ([35]).

This overview shows on the one hand that refrigerants are subject to multiple constraints, on the other that multiple solutions are available on the market and that development is ongoing.

4.7.2. Heat pump compressors

The compressor is central to a compression heat pump, as it elevates the pressure of the gaseous refrigerant. It hence highly influences the overall efficiency of the heat pump system. The most used compressor types are:

- Piston compressors: Common for small-to-medium-scale applications, providing high pressure ratios.
- Screw compressors: Widely used in industrial refrigeration, suitable for large capacities.
- Scroll compressors: Compact but limited in capacity.
- Turbo compressors: High-speed oil-free machines ideal for large-scale applications. The compressors also exhibit quick regulation features.

Emerging designs such as ejector-assisted cycles and rotational Joule-based compressors aim to reduce irreversibility in expansion processes, improving system COP by up to 25% in simulations.

The compressor type also affects the rangeability and flexibility of the machine. Turbo-compressors are emerging as a high-performance application for large-scale machines for temperatures suitable for district heating and good flexibility.

4.7.3. Heat pump configurations and data centre PUE

A high COP depends less on compressor or component choice and more on ensuring the heat pump setup matches the temperature and power needs of the machine working area. Even with components with the highest classification according to the Energy Directive 2017/1369 (EED), a cooling system (and heat pump) can still perform poorly if used at the boundaries of its expected working

([35]) ECHA news, “ECHA publishes PFAS restriction proposal”. Available at: https://echa.europa.eu/-/echa-publishes-pfas-restriction-proposal (visited August 2025).

area ([36]). For complex systems, such as heat recovery systems for data centres, it is therefore of great importance that the procurement sets measurable requirements for the system's properties, in addition to requirements for products according to EED. Moreover, when the market demands different properties for heat pumps than those required before, the industry initially tends to use heat pumps with suboptimal properties since development costs are high and the market is initially small. For example, heat pumps adapted for temperatures below 25°C will have problems ([37]) with high losses and lubrication at higher temperatures of the heat source.

However, innovative operating temperatures, while being challenging for available heat pumps, can open new opportunities. For example, while in the case of heat recovery from air cooling the HP configuration is trivial (low-temperature heat source connected to HP evaporator, high-temperature heat sink connected to HP condenser), in the case of heat recovery from liquid cooling more complex scenarios can arise. In particular, for high waste heat temperatures of the order of 60°C and efficient district heating networks with a return temperature of 50°C or less, a “temperature crossing” between the two flows can occur. In this case, in principle, a partial direct heat recovery (i.e., passive heat recovery through a common heat exchanger) can be carried out before the indirect heat recovery (i.e., the active heat recovery based on the heat pump). In practice, a direct pre-heating of the source circuit could be done, before implementing the final temperature boost via heat pump. Similar cases are described for example in the previously cited IEA-HPT Annex 58 report or in the LIFE4HeatRecovery project ([38]). In this case the overall COP can be significantly enhanced. For example, assuming that waste heat from cooling is available at 60°C and that the DH supply and return temperatures are 75°C and 45°C respectively, one could preheat DH from 45°C to about 54°C while cooling waste heat from 60°C to 56°C (35 % of waste heat directly recovered), using then the HP for the final heating stage up to 75°C, while further decreasing the waste heat temperature down to 50°C. The overall effective COP (neglecting circulation consumptions for the pre-heating stage) would then be enhanced by about 20-40 % (depending on the actual HP COP) with respect to a simple heat pump configuration.

Other possible configurations include the use of storage (also useful to introduce flexibility) and the partial local reuse of heat (see again theLIFE4HeatRecovery project).

([36]) Belivs, 2018. ”Utformning av kylsystem för butiker utifrån bästa tillgängliga Teknik”. Available at: Energieffektiva livsmedelslokaler - Kylteknik | RISE.

([37]) J. Lindhe, P. Tiljander, M Larsson, et.al., “Challenges and potentials of using a local heat pump in a 5GDHC solution: results from field and laboratory evaluations”, Energy, 2024. https://doi.org/10.1016/j.energy.2023.129807.

([38]) Deliverable “Database of waste heat recovery solutions”, Action C6.1 of the LIFE4HeatRecovery project, funded by LIFE programme, contract number LIFE17 CCM/IT/000085. Available at https://www.life4heatrecovery.eu/Resources/ (visited August 2025).

It is also useful to relate the HP COP/EER to the PUE. The PUE is the ratio between the total electricity consumption and the electricity consumption related to IT equipment, $\mathrm{PUE} = E_{el,tot}/E_{el,IT}$. In turn the total electricity consumption is the sum of the consumption related to IT equipment plus other electricity consumptions, among which the HP electricity would provide a major contribution. Moreover, the electricity consumption for the IT equipment is also the amount of heat to be absorbed by the HP, i.e., $E_{el,IT} = E_{th,C}$. Assuming that a HP/chiller needs to be installed for cooling purposes only (as in the case of air cooling for legacy data centres) one can develop the following relations:

$$\mathrm{PUE} = \frac{E_{el,tot}}{E_{el,IT}} \geq \frac{E_{el,IT} + E_{el,HP}}{E_{el,IT}} = 1 + \frac{E_{el,HP}}{E_{el,IT}} = 1 + \frac{E_{el,HP}}{E_{th,C}} = 1 + \frac{1}{\mathrm{EER}}$$

One can then see that a large EER (and hence a large COP) is crucial to keep a low PUE. For example, in order to have PUE ≤ 1.2, one should have EER ≥ 5.

However, **if the HP is installed mainly for heat recovery applications** (as in the case of liquid cooling, where heat rejection towards the external environment could be done without a HP), **it would be fair to allocate the HP consumptions outside the data centre**. In this case, only the cooling consumption "upstream" of the heat pump (e.g., the ones for the circulation of direct-to-chip liquid cooling) should be included in the PUE, with a strong improvement in its value. The HP performance would then only be relevant for the heat recovery efficiency.

A complementary perspective is to include energy recovery directly into the data centre efficiency analysis. While the PUE is the most common indicator of data centre efficiency, it only provides insight into how efficiently a data centre delivers power and cooling to its IT load, but it does not account for energy that may be reused outside the facility. To capture this aspect, the **energy reuse factor (ERF)** is introduced, representing the share of the total energy consumption that is exported and used beneficially, for example for district heating or industrial processes. Since PUE and ERF measure different aspects, the **energy reuse effectiveness (ERE)** metric has been developed to combine both. ERE adjusts the PUE by considering the fraction of energy that is reused, according to the relation:

$$\mathrm{ERE} = (1 - \mathrm{ERF}) \cdot \mathrm{PUE}\,.$$

This means that a data centre with significant energy reuse can achieve a much lower effective energy indicator compared to its PUE alone. For instance, a facility with a PUE of 1.4 (now including HP electricity consumption) and an ERF of 0.2 achieves an ERE of 1.12, reflecting the positive impact of energy reuse on the overall sustainability of the operation. Using a heat pump for waste heat utilisation for increasing COP, **ERF increases** and **ERE decreases** linearly with it. Allocation of HP electricity consumption within data centre boundaries or within district heating boundaries can also depend on contractual agreements and component ownership.

4.8. Connecting data centres and district heating

While some internal reuse of heat can be considered, the main waste heat recovery application for data centres is expected to be district heating. An analysis of the corresponding potential hence requires taking into account typical district heating operating conditions. Moreover, it is useful to understand how new network generations (with lower operating temperatures) can impact on the overall system efficiency.

Comments on the role of storage and connection distance will also be included. Finally, non-technical issues will be addressed, such as waste heat recovery contracts, heating and cooling as a service, and tariff structures.

4.9. Overview of data centres and district heating

The different categories of data centres have different conditions for heat reuse. The category which are the most suitable and easiest are the enterprise and crypto data centres where the owner is in full control of the IT equipment and its cooling system. These data centres are often of a decent power level and can benefit from the scale effect for doing technical installation.

The temperature rejected from the data centre is heavily dependent on cooling technology of the IT equipment, where air cooling yields the lowest temperature, while direct-to-chip and immersion cooling are in a similar range (see previous chapter for details). Another aspect is the temperature at which the chip that is cooled is operating; a higher excess temperature results in a decreased need of heat pumps to reach the desired temperature. For an immersion cooling system, it is shown that an increase of 20°C results in an IT-power consumption increase of 5%, due to the current leakages that occur when transistors are exposed for higher temperatures ([39], [40]). There is a trade-off between these two aspects that needs to be considered when designing the system integration, especially in areas with limited power supply.

4.10. Waste heat reuse in the proximity of the data centre

Using surplus heat without raising the temperature of the heat source could lead to lower capital expenditures. In such situations, allocating excess heat to nearby

([39]) P. Taddeo *et al.*, “Experimental and numerical analysis of the thermal behaviour of a single-phase immersion-cooled data centre”, Appl. Therm. Eng. **234**, 121260 (2023).

([40]) Stockholm - heat recovery city - DCD (visited August 2025).

buildings may be more effective than using it in a larger district heating infrastructure. Today's traditional data centres usually have air-cooled systems, with about 20-40°C waste heat (depending on location and season). An example of the disposal of such heat is greenhouses ([41]). Greenhouse cultivation with heat from data centres distributed with hot air and without heat pumps is probably the most commercially viable alternative from traditional data centres. Other applications such as snow melting, fish farming and industrial processes have potential, but are often dependent on local conditions and still in the pilot phase. In general, it is rare for buyers of such low-temperature heat to be located near a data centre. It is probably more common for properties with heating systems that require about 40-55°C to be located adjacent to data centres. An example is Danfoss' headquarters in Denmark, which uses heat pumps to recover waste heat from its own data centre, to distribute it to surrounding buildings via a water-borne circuit ([42],[43]).

4.11. Waste heat reuse for district heating

District heating is an efficient and sustainable solution for heating homes, buildings and industrial facilities. By distributing heat from a central source through a network of insulated pipes, district heating can use surplus energy from power plants, waste incineration or industrial processes. This reduces the need for individual heating systems and contributes to lower carbon emissions and better resource efficiency. As energy systems become more integrated, district heating plays a key role in the green transition, especially in urban areas where heat demand is high and opportunities for sector integration are significant.

The required supply temperature for district heating varies by grid generation: third-generation systems typically demand 70–100°C, while fourth-generation systems operate at lower supply temperatures (maximum 60–70°C), facilitating greater integration of data centre waste heat. Fourth-generation systems also employ efficient large-scale heat pumps and smart controls to maximise use of low-grade waste heat and renewable energy ([44]).

Fifth-generation district heating and cooling (5GDHC) systems use ambient temperature grids (6–40°C) and decentralised heat pumps in individual buildings, enabling efficient recovery and reuse of low-grade waste heat. These systems allow buildings to exchange thermal energy, supporting decarbonisation, lowering energy losses, and improving energy efficiency. Optimised 5GDHC

([41]) H. M. Ljungqvist et al., “A realistic view on heat reuse from direct free air-cooled data centres”, *Energy Convers. Manag.: X* **20**, 100473 (2025).

([42]) https://www.grontsamhallsbyggande.se/2024/06/24/danfoss-och-hpe-atervinner-overskottsvarme-fran-datacenter/ (visited August 2025).

([43]) https://community.hpe.com/t5/the-cloud-experience-everywhere/hpe-and-danfoss-unveil-heat-recovery-solution-for-enterprise/ba-p/7216660 (visited August 2025).

([44]) Lund et al., 2021, https://www.irena.org/Innovation-landscape-for-smart-electrification/Power-to-heat-and-cooling/8-Fourth-generation-DHC-systems (visited August 2025).

configurations can deliver up to 40% energy savings, with average conversion rates of 59% across building types ([45], [46]).

District heating is today the most common method for distributing waste heat from data centres due to its established infrastructure, ability to serve thousands of households and viable economic model where energy companies purchase the recovered heat. Integrating data centre waste heat into district heating systems increases energy efficiency, reduces greenhouse gas emissions, lowers water consumption and noise, and enhances the sustainability profile of data centres[47]. Existing and new district energy networks are critical for transferring recovered heat to commercial, residential and industrial buildings ([48]).

Integrating data centre waste heat into district heating networks can significantly reduce energy system costs and carbon emissions, as demonstrated in Denmark, though technical, economic and policy barriers persist. Seasonal fluctuations in heating and cooling demand can be mitigated by combined cooling and power systems that adjust output to match needs.

Case studies illustrate the impact of large-scale heat recovery: Facebook's data centre in Odense, Denmark supplies up to 100,000 MWh of waste heat annually to the city's district heating, powered by local renewable energy ([49]). In Stockholm, AWS data centres recover air-cooled waste heat via heat pumps, upgrading low-temperature heat for district heating and providing enough energy to heat tens of thousands of homes, while reducing reliance on fossil fuels ([50], [51]).

To convey the heat generated to the district heating system three different systems are involved. Firstly, the chip cooling system, which can be air, on-chip or an immersion system. Secondly, the facility cooling system, which for on-chip and immersion is a separate loop of liquid. Thirdly, the heat pump system, which upgrades the data centre heat to meet the temperature district heating supply temperature.

How this series of systems is dimensioned and which technologies are used strongly affects the output temperature of the waste heat from the data centre, which can vary in the range of 20–65°C (see previous chapter). By making a

([45]) Chicherin, S., 2025; https://www.mdpi.com/2227-9717/13/6/1730 (visited August 2025).

([46]) Sweden Data Centers Excess Waste Heat, District Heating Grid, Rack Workload Densities, Heat Pump Efficiency (COP) (visited August 2025).

([47]) Harnessing Excess Heat: Lessons from Stockholm's Open District Heating - EnergiRaven (visited August 2025).

([48]) https://www.districtenergy-digital.org/districtenergy/q1_2025/MobilePagedArticle.action?articleId=2037059#articleId2037059 (visited August 2025).

([49]) Copenhagen Centre of Energy Efficiency, 2023; https://www.irena.org/Innovation-landscape-for-smart-electrification/Power-to-heat-and-cooling/31-Waste-heat-recovery-from-data-centres (visited August 2025).

([50]) https://stockholmdataparks.com (visited August 2025).

([51]) Sweden Data Centers Excess Waste Heat, District Heating Grid, Rack Workload Densities, Heat Pump Efficiency (COP) (visited August 2025).

system design that results in the uppermost part of the temperature range, one can reduce the heat pump size and the amount of power needed. In addition, different chips have different operating case temperatures, ranging about 50-100°C.

In order to have any waste heat to deliver to the district heating system there have to be load and users of the data centre, which are highly variable according to the type of data centre. The most stable data centre load is considered to be the crypto mining data centres, which generally run their equipment in full load the year around. Secondly would be Enterprise data centres which follow their user pattern with peak and low usage points in the load, often these can be predicted and taken into consideration as variation in the waste heat generation. The most challenging ones are HPC centres, which are often represented by a binary operation, either they are in idle mode, or they are doing full capacity work which in particular occurs for training of AI models.

With data centre loads and waste heat profiles that are varying a lot, it might be a challenge for the heat pump to be able to ramp up and down fast enough. A possible solution is properly sized thermal storage that can meet this fluctuation and unload the heap pump while ramping. With a poorly dimensioned heat upgrade system the data centre owner would need to invest in a secondary cooling system in order to reject the waste heat that the heat recovery system cannot absorb. During the low load periods of the district heating system operation, it is important to ensure that there is enough cooling power for the data centre so that owner can benefit from reduced cost of installing a cooling system that is only used a couple of days or hours each year.

Apart from technical aspects, waste heat recovery to district heating always involves issues related to contractual aspects and general agreements among the DH utility and the waste heat owner. These are problems are known to the DH industry and addressed in similar contexts. Critical aspects are mainly related to allocation of investments and ownership boundaries, as well as to contract duration and related guarantees. The risks involved are sometimes addressed through the involvement of third parties, like energy service companies (ESCo's) and the introduction of heat-as-a-service contracts: this typically introduces extra costs for the DH utility, but avoids the need to deal with construction/management. This "servicification" approach is, for example, described in the REWARDHeat project ([52]). General articles for waste heat contracts were instead discussed in the ReUseHeat project ([53]), while Finnish

([52]) REWARDHeat, D3.4, "Business models at REWARDHeat demonstrators". Available at https://www.rewardheat.eu/Project-deliverables/ (visited August 2025).

([53]) ReUseHeat, https://www.reuseheat.eu/wp-content/uploads/2022/09/ReUseHeat-Handbook-For-Increased-Recovery-of-Urban-Excess-Heat.pdf (visited August 2025).

Energy made available a rather detailed contract template ([54]). While these barriers remain, awareness and solutions are spreading in the sector.

In summary, district heating networks offer an efficient, scalable and economically attractive solution for using data centre waste heat, supporting energy system decarbonisation and circular economy objectives. Technological innovation and operational optimisation can further enhance the feasibility and effectiveness of these systems. At the same time, typical issues and barriers related to waste heat recovery applications apply to the case of data centres as well.

4.12. Overall estimates for data centre waste heat recovery to district heating

The district heating demand in Europe was 548.6 TWh/y in 2023 ([55]), roughly corresponding to 12 % of the overall European space and hot water heating demand. A large majority of existing networks operates at high temperatures (order of 90 °C), with some low-temperature solutions emerging. However, as mentioned in chapter 2, the overall electricity consumption for data centres in Europe is about 55 TWh/y, distributed among different data centre types and cooling technologies. In particular, a 10 % share of high power density racks – and hence of liquid cooling with relatively high waste heat temperatures – can be estimated.

Due to operating temperatures, the recovery of data centre WH towards DH is typically expected to require HPs (see chapter 4), though liquid cooling solutions either yield high COP values or allow to avoid these machines at all (possibly using WH for pre-heating only).

The detailed matching of data centres with DH would need spatial information about relative distance and temporal information about relative availability/demand profiles. As mentioned in chapter 2, open maps of data centre exist, though with limited reliability. Similarly, EU-level open maps of district heating areas are available ([56]). In principle, by matching these resources at GIS level, it would be possible to obtain, for example, an approximate count of how many data centres are located within a certain distance from a DH network.

Assuming – as a theoretical exercise – that 50 % of data centres were within a reasonable distance from DH networks and that only 10 % of data centres exploited liquid cooling, so that a rough average COP of 3 could be expected,

([54]) https://energia.fi/wp-content/uploads/2024/01/Model-contract-for-heat-trade-20240109-1.pdf (visited August 2025).

([55]) DHC Market Outlook 2025, executive summary available at https://api.euroheat.org/uploads/Market_Outlook_2025_f4b28fcf89.pdf (visited August 2025).

([56]) Pan-European Thermal Atlas (PETA) maps, part of Heat Roadmap Europe project, are available at https://heatroadmap.eu/peta4/ (visited August 2025).

one would find that a recovery of the corresponding WH could cover about 10 % of the current DH demand (i.e., order of 50 TWh/y). In the case of temporal mismatch or priority given to other sources, this number would be lower.

Considering the expected increase of data centres (see chapter 1), much stronger than the expected growth of DH demand, one can expect that the available WH from data centres could cover 20-40 % of the overall DH demand within 2030 (i.e., 150-300 TWh/y, assuming a DH demand growth of 5 % per year). While these estimates should be taken with much caution, they clarify the great potential of heat reuse in this sector.

4.13. Flexibility and sector coupling

Flexible electricity use and sector coupling are increasingly vital concepts in the transition to more sustainable energy systems. Data centres play a unique role in this development since they can contribute to both concepts. By leveraging both flexibility and sector coupling, data centres can become active participants in balancing the grid, reducing emissions and supporting the broader energy transition.

4.13.1. Flexible operation to support the power grid

The increasing integration of renewable energy sources into power grids introduces variability and uncertainty in electricity supply. Simultaneously, the digital economy drives the expansion of data centres, which are large, energy-intensive facilities. This dual trend presents a unique opportunity: data centres can play a significant role in grid stabilisation through flexible operation. By aligning their energy consumption with grid conditions, data centres can act as controllable loads, contributing to grid reliability, efficiency, and sustainability ([57]).

Data centres consume significant amounts of electricity, often operating continuously to meet service-level agreements (SLAs). Traditionally, their energy demand has been considered inflexible. However, advancements in workload scheduling, virtualisation and distributed computing have enabled new models of operation that allow data centres to adjust their power usage dynamically ([58]).

Key characteristics that make data centres suitable for grid support include:

- High controllability: Workloads can be shifted in time or location.
- Energy storage integration: Battery systems can provide demand response and frequency regulation.

([57]) Energiforsk 2025: *Datacenter som flexibilitetskapacitet*. Report 2025:1089.

([58]) Swedish Energy Agency (2022): Smart integration of power grids, microgrids and data center.

- Geographical distribution: Multi-site data centres can balance loads across regions.
- Thermal inertia: Cooling systems can be modulated without immediate impact on operations.

Flexible data centre operation can be achieved through several mechanisms:

- Demand response participation: Data centres can participate in demand response programs by reducing or shifting loads during peak demand periods by; load scheduling, load shifting, workload migration.
- Integration with renewable energy: Data centres can align their operations with renewable generation profiles, such as solar and wind. Techniques include forecast-based scheduling using weather and generation forecasts to plan workload execution.
- Battery energy storage systems and onsite generation (e.g., solar PV, fuel cells): Enhance flexibility by providing backup power during grid disturbances, offering ancillary services and reducing grid dependency during peak hours.

Implementing flexibility requires intelligent control systems that balance operational constraints with grid needs such as real-time monitoring, optimisation algorithms and coordination platforms for multi-site data centre orchestration. Advanced techniques such as reinforcement learning, model predictive control, and federated optimisation are being explored to enhance responsiveness and scaleability.

Several pilot and commercial implementations have demonstrated the feasibility and benefits of flexible data centre operation such as Google's carbon-aware computing ([59]) where they are shifting workloads based on carbon intensity of local grids. Microsoft's grid-interactive data centres ([60]) are participating in demand response and on frequency regulation markets. These examples highlight the potential for data centres to become active participants in energy systems. Even so, a relatively small share of data centres is today known to be active on flexibility markets or ancillary services. Nevertheless, the increasing share of the total electricity power used by data centres makes them an important player to be active in such markets. An interview with the planning modelling supervisor[61] at ERCOT (Texas grid, USA) states that the grid operator's ability to dynamically manage data centre loads provides essential flexibility for grid stability.

([59]) Carbon-Aware Computing for Datacenters – Google

([60]) Grid-interactive data centres: enabling decarbonisation and system stability - Microsoft

([61]) Renewable Energy World (2025), https://www.renewableenergyworld.com/power-grid/understanding-ercots-strategy-for-grid-transformation-to-meet-the-energy-demands-of-the-future/ (visited February 2026).

Despite good conditions for flexible operation of certain types of data centres, SLAs may counteract their ability to participate in grid services and align financial models with flexibility benefits. Future work could focus on:

- Standardisation of interfaces between data centres and grid operators.
- Finalisation of the work on the EU network code on demand response and clarity of the rules for all actors before national enforcement.
- Development of flexibility metrics and certification schemes.
- Integration with edge computing and decentralised architectures.

Flexible data centre operation represents a transformative approach to grid support, leveraging digital infrastructure to enhance energy system resilience and sustainability. As the energy transition accelerates, data centres can evolve from passive consumers to active grid partners, unlocking synergies between digitalisation and decarbonisation.

4.13.2. Sector coupling for data centres

Sector coupling refers to the strategic integration of energy consuming sectors such as heating, cooling, transportation, industry and with the electricity sector to improve overall system efficiency and enable decarbonisation. Data centres, as critical infrastructure with high and growing energy demands, are increasingly recognised as key players in sector coupling strategies. By linking data centre operations with other sectors, it is possible to unlock synergies that support grid flexibility, renewable energy integration and circular economy principles. Data centre sector coupling involves the coordinated interaction between data centres and other sectors through shared energy systems, infrastructure and control mechanisms.

Sector coupling can include areas such as:

- Power-to-X integration: Using electricity for heating (Power-to-Heat), hydrogen production (Power-to-Gas), or other applications.
- Waste heat use: Recovering and repurposing heat generated by data centres for district heating or industrial processes.
- Smart grid interaction: Enabling bidirectional communication and control between data centres and energy networks.
- Use of methane waste from landfills or industry sites.
- Mobility and transport: Supporting electric vehicle (EV) charging infrastructure through co-location and energy management.

Data centres produce substantial amounts of low-grade heat, which can be captured and reused in nearby buildings or district heating systems. This coupling

reduces the need to separate heating sources and enhances overall energy efficiency. Heat exchangers and heat pumps enable effective transfer of thermal energy into heat recovery systems, such as district heating networks. DH networks can, in turn, use seasonal storage systems to save energy surplus for demand peaks, improving load balancing.

Coupling data centres with renewable energy sources and energy storage systems supports grid decarbonisation and resilience. Solar PV and wind turbines can power data centres directly by onsite generation. Backup power and grid services such as peak shaving and frequency regulation can be achieved by battery storage. At a later stage, the surplus electricity from renewables can be used to produce hydrogen, with data centres acting as anchor loads.

Data centres are often equipped with a UPS which is often modular and scaleable, a battery system and energy management system that could be used for power balancing. By co-locating a data centre with charging stations for EVs, the battery capacity and energy management system can be shared between the two and create synergies, where the car battery can be used as additional storage capacity both for up and down frequency regulation. With vehicle-to-grid, the EV batteries can also be used for ancillary services. Data centres with large energy capacity can co-host slow-charging or fast-charging stations, which can further be integrated with logistics and autonomous vehicle operations.

The benefits of sector coupling with data centres as drivers are mainly to reduce waste heat in general and improve use of energy resources such as carbon reduction by enabling an increased use of renewable energy and reduced fossil fuel dependency. A shared infrastructure and services over the different sectors will result in a lower operation cost and optimised economic for both parties. In the long run it will enhance system stability and responsiveness to demand fluctuations.

Possible areas for sector coupling and emerging case studies are to leverage cold climates and district heating for efficient heat reuse, especially in the Nordic cold climate as part of a smart city infrastructure. Furthermore, as a green hydrogen hub where data centres' electricity demand is used to stabilise hydrogen production.

Sector coupling with data centres as drivers has its challenges and considerations to be taken into account:

- Technical complexity: Requires advanced control systems and interoperability standards.
- Regulatory alignment: Policies must support cross-sector integration and value stacking.
- Investment and planning: High upfront costs and long-term coordination are needed.

- Data privacy and security: Especially when integrating with public infrastructure.

As digitalisation and electrification accelerate, sector coupling involving data centres will become increasingly important. Emerging technologies such as AI-driven energy management, modular data centres and integrated energy systems will further enhance the potential for cross-sector collaboration. Policy frameworks, research initiatives and public-private partnerships will be essential to scale sector coupling solutions and realise their full benefits.

4.13.3. Key takeaways from flexibility and sector coupling

Data centre sector coupling represents a transformative approach to energy system design, enabling more sustainable, efficient and resilient infrastructure. By bridging digital and physical systems, data centres can become central nodes in the future energy landscape, supporting climate goals and economic development.

The ability of data centres to offer flexible use of electricity depends to a large extent on the type of data centre. Certain data centres prioritise availability and are unwilling to compromise this requirement, whereas others may provide increased flexibility in exchange for financial incentives. The potential for sector coupling and enhanced system flexibility can be outlined as follows:

- Enterprise data centres can be considered suitable for sector coupling, since they have ownership of both the IT-hardware and its cooling system. With proper planning of the infrastructure, it can greatly support sector coupling such as waste heat use.
- Cryptocurrency and HPC data centres are suitable for sector coupling when doing batch processing, which can be paused and resumed.
- Cryptocurrency data centres are also suitable as a flexible load in the power grid, in some cases more suitable than AI data centres (see Chapter 4.15).

When deploying data centres for sector coupling and flexibility services, it is essential to undertake careful planning to determine appropriate UPS configuration, battery storage capacity and cooling technology. This approach enables optimal integration with other systems.

4.14. Overall techno-economic estimates

The feasibility of waste heat recovery projects significantly depends on cost aspects, both related to **investments** and to **operation**. Data centres exploit different cooling technologies, which affects both the resulting waste heat temperature and the type of involved thermal circuits.

From a cost perspective, the boundary considered here includes the interface between the external data centre cooling and the district heating. The internal cooling system of the data centre would indeed be installed and operated even in the absence of any waste heat recovery application, so its costs must be allocated on the data centre itself. The two key components to be considered are hence a waste heat recovery **substation** (typically including heat pumps, heat exchangers, common hydraulic components like pumps and valves, electrical components and a control system) and the **pipeline** needed to connect the data centre with the existing district heating network (with a pipe size depending on the delivered power and with a length depending on the connection distance). Below, orders of magnitude for the corresponding costs are provided, distinguishing between investment/capital costs (CAPEX) and operational costs (OPEX).

4.14.1. General considerations on cost estimates

For waste heat recovery projects, similarly to renewable energy projects, the major cost is the initial investment cost (though in the presence of heat pumps electricity costs also constitute a large share). For a comparison with other energy sources, it is useful to look for indicators representing the average energy unit cost, e.g., in €/MWh. To properly account for investment costs, it is important to include depreciation effects, so that investments sustained at the project start are annualised and distributed over the project lifetime. Summing these annualised investment costs with operation costs for a typical year and dividing such total by the estimated annual energy production, one obtains a reference unitary cost. This is typically expressed through the **levelized cost of heat (LCoH)**, explained in detail in the following technical box for interested readers. For a project to be competitive, LCoH should be lower than the **heat tariffs of current district heating applications**, which can be assumed **to be around 100 €/MWh**. For example, in 2021 the average tariffs for Denmark, Germany, Sweden and the Netherlands were in the range 85-98 €/MWh ([62]).

([62]) Sigurd Næss-Schmidt et al., “District heating tariffs in Europe”, Copenhagen Economics (2021).

Technical box – Levelized Cost of Heat (LCoH)

In order to compare investments related to heat production, a commonly used indicator is the LCoH (See for example Yang et al., "Seasonal thermal energy storage: A techno-economic literature review", *Renew. Sust. Energy* Rev. **139**, 110732, 2021). It is calculated in a similar way to the levelised cost of electricity (LCoE) as

$$\mathrm{LCoH} = \frac{\sum_{i=0}^{N} C_i/(1+r)^i}{\sum_{i=0}^{N} E_{th,i}/(1+r)^i} = \frac{C_0 + \sum_{i=1}^{N} C_i/(1+r)^i}{\sum_{i=1}^{N} E_{th,i}/(1+r)^i},$$

where r is the interest rate, N is the lifetime of the project in years, C_i is the cost sustained in the i-th year, and $E_{th,i}$ is the thermal energy (heat) delivered in the i-th year. It is usually assumed that the project sustains all investment costs at year zero, i.e., $C_0 = C_{inv}$ = CAPEX, and that from year 1 onwards there are only constant operation costs, i.e., $C_i = C_{op}$ = OPEX for $i = 1, \dots, N$. Similarly, it is typically assumed that energy production is zero in year zero (during system installation), while it remains constant from year 1 onwards, i.e., $E_{th,0} = 0$, $E_{th,i} = E_{th}$ for $i = 1, \dots, N$. Under these assumptions, the formula can be simplified by introducing the so-called annuity $a(r, N) = r/[1 - (1+r)^{-N}]$ (S. Frederiksen and S. Werner, "District heating and cooling", Studentlitteratur, 2013), which depends only on the interest rate and the number of years. One finds:

$$\mathrm{LCoH} = \frac{C_{inv} a + C_{op}}{E_{th}}.$$

The operating costs are also usually split into an operation and maintenance (O&M) part, calculated as a fraction of the investment cost, and into an energy-related part, e.g., related to electricity consumption for HPs. Electricity consumption can be typically estimated as a fraction of the delivered thermal energy, through an effective COP (by similarity with HPs), i.e., $E_{el} = E_{th}/\mathrm{COP}$. Then, $C_{op} = r_{O\&M} C_{inv} + p_{el} E_{th}/\mathrm{COP}$, where $r_{O\&M}$ is the considered O&M fraction and p_{el} is the electricity price. Since investment costs are often parameterised to the delivered power, it can also be useful to express the delivered thermal energy as $E_{th} = P_{th} h_{eq}$, where h_{eq} are the equivalent operation hours. One finally gets

$$\mathrm{LCoH} = \frac{C_{inv}(a + r_{O\&M}) + p_{el} E_{th}/\mathrm{COP}}{E_{th}} = \frac{c_{inv}(a + r_{O\&M})}{h_{eq}} + \frac{p_{el}}{\mathrm{COP}}.$$

where $c_{inv} = C_{inv}/P_{th}$ is the specific cost per unit of power and $c_{el} = p_{el}/\mathrm{COP}$ is the electricity related cost component. This highlights the dependence of LCoH on specific investments costs, financial parameters, maintenance parameters, operation time, electricity price and electric efficiency.

For components with different lifetimes the above simplified LCoH formula does not hold (a similar, but more complex version could be used). In addition, more detailed cost splits in terms of O&M costs are possible. However, these refinements do not strongly affect the overall estimates for the case of heat recovery projects.

As explained in the previous technical box, the LCoH can be expressed in simplified form as a function of a few key parameters

$$\mathrm{LCoH} = \frac{c_{inv}(a + r_{O\&M})}{h_{eq}} + \frac{p_{el}}{\mathrm{COP}},$$

i.e., the specific investment cost c_{inv} (expressed e.g., in €/kW), the annuity a (essentially a depreciation rate), the operation and maintenance rate $r_{O\&M}$, the number of equivalent operation hours h_{eq}, the electricity price p_{el}, and the COP. The last ratio in the equation can be seen as the electricity-related cost component of the LCoH, i.e., $c_{el} = p_{el}/\mathrm{COP}$. For HP applications this component can be significant, while for other applications (e.g., piping) it can even be neglected (since a very high effective COP would be involved, see below). In the following paragraphs, the size and variability of these parameters are discussed separately, leaving the discussion of specific investments to dedicated sections about HPs and pipelines.

Currently, typical **interest rate** values, usually given as **WACC** (weighted average cost of capital), are in the range 6-8 %. The **lifetime** of thermo-hydraulic components like heat pumps is typically around 20-25 years, as reported by the Danish Energy Agency ([63]), while DH pipelines have a longer lifetime. It is hence interesting to assess the variability of the annuity a for these ranges, as reported in Table 12.

Table 12 – Annuity variability

Variability of the annuity (a) as a function of WACC (r) and lifetime (N). See text for details.

Annuity variability(a)		Lifetime (N)		
		20	**25**	**30**
WACC (r)	**6.0 %**	8.7%	7.8%	7.3%
	7.0 %	9.4%	8.6%	8.1%
	8.0 %	10.2%	9.4%	8.9%

Concerning the **O&M cost fraction**, a typical assumption for industrial plants is $r_{O\&M}$ = 3 % (a more detailed breakdown for HPs can be found in Danish Energy Agency catalogue), while lower values are expected for DH pipe networks.

Electricity costs can vary significantly between countries and over time. A significant rise was observed in 2022, mostly due to the Ukraine war. Prices have now decreased close to previous levels, but forecasts for future years remain uncertain (while price reductions might occur thanks to higher renewable shares, international factors might still cause price increases). Finally, electricity prices can be strongly affected by particular types of contracts like power purchase

([63]) Danish Energy Agency, "Technology Data Catalogue", https://ens.dk/en/analyses-and-statistics/technology-catalogues (visited August 2025).

agreements (PPA) used for the renewable energy sector. Considering the relevance of electricity for data centres, one might expect operators to secure favourable contracts (or to choose countries with lower prices). However, if the heat recovery system is owned by the DH operator, the electricity contract would follow different negotiation paths. In general, a non-negligible variability can be expected, but probably in an interval on average slightly lower than for common industry. Hence, **values in the range 100-150 €/MWh might be reasonably considered**.

According to Eurostat ([64]), the average non-residential EU-27 electricity price (excluding VAT) in the second semester of 2024 was 190 €/MWh. While this is a much higher price than the aforementioned range, in many northern Europe countries prices are much lower: for example, 128 €/MWh in Denmark and 85 €/MWh in Sweden. Countries like Germany (236 €/MWh) and Italy (227 €/MWh) are instead slightly above average.

Concerning the COP, its dependence on temperature conditions was discussed in chapter 4 showing how the temperature of the source (i.e., of the data centre waste heat) can roughly range between 20°C and 60°C, depending on whether air cooling or most innovative liquid cooling technologies are used. However, for the temperature of the sink, a conventional DH supply temperature of 90°C was addressed, together with a lower temperature of 75°C as representative of a more efficient network. **It is useful to combine electricity price and COP into a single parameter representing electricity-related unitary costs,** $\boldsymbol{c_{el} = p_{el}/\mathrm{COP}}$. Considering reasonable electricity prices and the COP ranges identified in chapter 4 ([65]), one then finds the variability expressed in Table 13. From the table it can be seen how, under the less favourable conditions considered (heat recovery from air cooling with COP = 3 and electricity price of 150 €/MWh), the electricity-related contribution to the LCoH is about 50 €/MWh, i.e., a significant amount. However, under the more favourable considered conditions (heat recovery from liquid cooling with a COP = 7 and electricity price of 100 €/MWh), the electricity-related contribution to the LCoH is less than 15 €/MWh. The electricity-related component would become negligible in the case of direct heat recovery through a heat exchanger (feasibly only if the waste heat temperature is higher than the DH supply temperature).

([64]) Eurostat, Electricity prices for non-household consumers: https://ec.europa.eu/eurostat/databrowser/view/nrg_pc_205/default/table?lang=en&category=nrg.nrg_price.nrg_pc (visited August 2025).

([65]) Here a slightly narrower COP range is conservatively considered, i.e., 3-7, with the highest COP related to liquid cooling applications. The selected range avoids extreme values, where the approximate COP calculation of chapter 4 is less reliable.

Table 13 – Electricity-related cost variability

Variability of the electricity-related cost component (c_{el}) as a function of COP and electricity price (p_{el}). See text for details.

Electricity-related cost variability (c_{el}) [€/MWh]		Electricity price (p_{el}) [€/MWh]		
		100	**125**	**150**
COP	**3.0**	33.33	41.67	50.00
	5.0	20.00	25.00	30.00
	7.0	14.29	17.86	21.43

This discussion highlights the typical cost structure expected for these projects. The following section focuses on specific investment costs for HP substations and DH pipe networks.

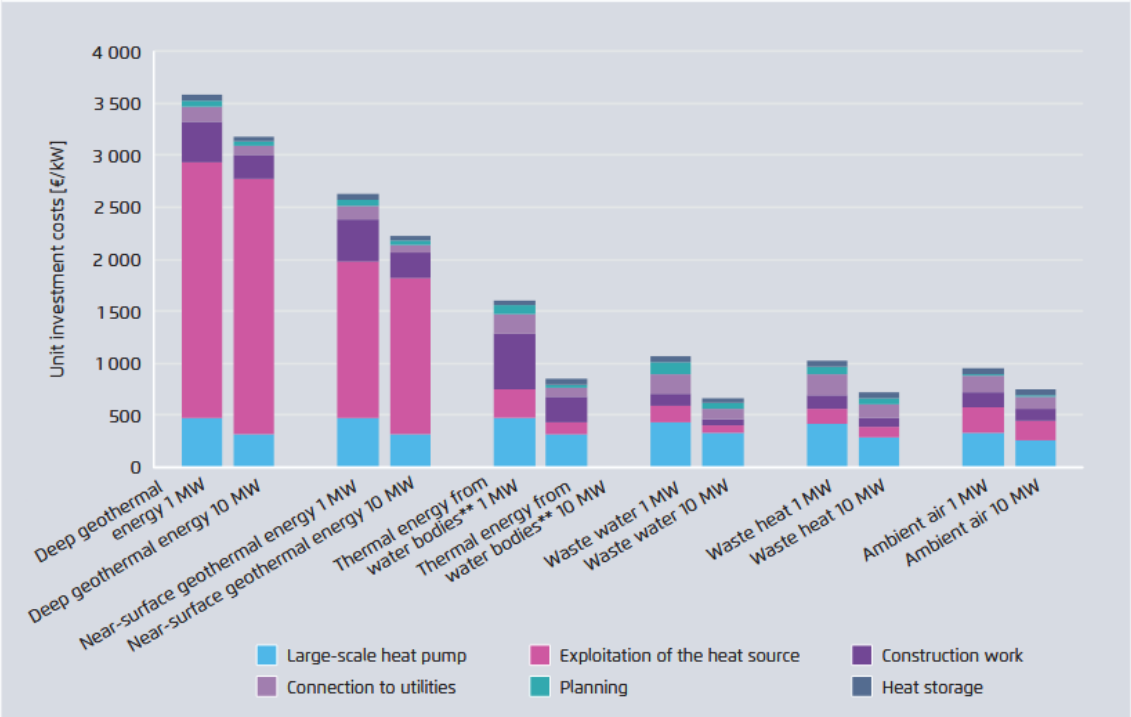
Concerning **equivalent operation hours**, h_{eq}, due to the continuous operation of data centres, very high values could be achieved for this parameter in principle. For example, if the cooling system operates on average at 50 % of its nominal capacity for all the time, one would obtain 50 % × 8760 h/y = 4380 h/y. Considering an average load factor in the range of 50-90 %, this would correspond to around 4000-8000 h/y in terms of availability. However, the actual use of the waste heat recovery system depends not only on the source availability, but also on the sink load. District heating systems have a much lower demand in summer and, depending on the DH size, part of the waste heat might not be absorbed (possibly also due to competition with other low-cost heat sources like waste-to-energy plants). Taking this finto account, it can be interesting to consider a more conservative range of operation hours, i.e., 2000-6000 h/y, with a reference mid-range value of 4000 h/y.

4.14.2. Waste heat recovery substation costs

The economic analysis presented in this section is based on a study of the available literature regarding investment costs for large-scale heat pumps used in industrial settings and district heating. Additionally, it includes a series of assumptions related to costs not directly available, such as those for design, project management, and safety.

There are several documents available online that provide summary information on investment costs for large-scale heat pumps. However, the values found vary significantly, and the descriptions do not often clarify whether they refer to the machine or the entire system. A first useful overview of unit investment costs (in € per kW of thermal power) for heat pump systems is shown in Figure 10. This highlights the variability due to source type and to HP size. The figure also shows a split by different cost items.

Figure 10 – Heat pump unit investment costs.



Source: "The roll-out of large-scale heat pumps in Germany", 2023, Fraunhofer / Agora (2024), Fig. 31.

Impact of the heat source and the other system components on the investment costs of large-scale heat pump projects.

The Danish Energy Agency (DEA) catalogue mentioned earlier provides cost data for excess heat HPs for 3 representative sizes, namely 1, 3, 10 MW (with slightly different COP values, namely about 4.0, 4.5, and 5.0 respectively). For all cases a lifetime of 25 years is estimated with reference to 2025. The heat pump system investment costs are estimated to be 1,320, 910, and 710 €/kW respectively (lower costs for larger sizes). The cost for equipment is estimated to be about 75-80 %, the remaining investment cost being attributed mostly to grid connection and installation. The costs for O&M are split into variable and fixed parts. Assuming 4000 h/y of operation for the variable part, everything can be converted into €/kW, yielding O&M costs of, respectively, 1 %, 3 %, and 10 % depending on the size (with higher O&M costs for larger sizes). If fewer operation hours are assumed, the O&M costs would decrease. The investment cost is roughly in line with the Agora report's cost for a waste heat HP shown in Figure 10 (1000 €/kW for a 1 MW HP, 700 €/kW for a 10 MW HP).

In a paper by Pieper et al. ([66]) a more detailed split of costs is given, showing different cost items taken from about 40 case studies for different heat pump applications. Case studies refer to pre-2018 data and, to make them comparable

([66]) Pieper *et al.*, "Allocation of investment costs for large-scale heat pumps supplying district heating", *Energy Procedia* **147**, 358–367 (2018).

with current data, inflation estimates should be taken into account. Eurostat provides harmonised indices of consumer prices ([67]), where the compound inflation rate from 2018 to date can be estimated at around 25 % (with minor fluctuations among EU countries). Using this correction, the order of magnitude for the cost of excess heat HP applications estimated by Pieper is again in line with the estimates reported above.

As most of the above data refer to German or Danish installations, the Eurostat "level price index" ([68]) could be used to allow for comparison with other European countries. This database also provides a split per price category. Looking at the "Price level indices for machinery and equipment" (2023) in the database, it emerges that Danish prices are 12 % above the EU average, while German prices are in line with the latter (1 % below the average).

The above estimates do not include costs for buildings or land, assuming they are already in the availability of the considered facilities. Minor structural costs for HPs (e.g., containers of platforms) are typically included.

As a conclusion, a **reference investment cost of about 1000 €/kW** is suggested with, with lower costs for installations beyond 1 MW. An overall variability of +/- 30 % is also suggested (700-1300 €/kW), both taking into account size variations and uncertainties found in specific projects.

4.14.2.1. Thermal energy Storage

The above costs do not explicitly include **thermal energy storage** (TES) solutions, which might however be valuable to enhance system flexibility. Cost data for large-size hot water tanks are available in the DEA catalogue as well. Cost data are provided as a function of volume in m^3, as well as a function of storage capacity in MWh, assuming a temperature difference of 55 K with 90 % availability. The estimated unit cost was 3000 €/MWh for a size of 175 MWh (i.e., about 3000 m^3) in 2015 (about 175 €/m^3). According to Eurostat HICP, the EU-27 price increase from 2015 to date was about 30 % (19 % for Denmark). This would lead to a present cost of almost 4 M€/GWh (about 225 €/m^3). Smaller tank sizes and other variants (e.g., installation complications, need for special foundations, different temperature ranges) can strongly increase these costs. In particular, for data centre applications and looking towards DH networks operating at lower temperatures, temperature differences of 30 K might be more reasonable than the 55 K considered in the DEA catalogue. For sizes relevant to data centres (e.g., 10 MWh, able to charge/discharge a 1 MW power for 10 h, volumes of the order of 300 m^3), it is suggested here to consider investment costs

([67]) Eurostat, harmonized indices of consumer prices (HICP): https://ec.europa.eu/eurostat/web/hicp (visited August 2025).

([68]) Eurostat, Comparative price levels for investment: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Comparative_price_levels_for_investment (visited August 2025).

of the order of 500 €/m^3, which for a temperature difference of 30 K and 90 % availability would correspond to about 16,000 €/MWh.

As a final comment concerning TES solutions, it is worth mentioning the continuous research and development in the field, with possible alternatives to water tanks. A relevant example is provided by phase change materials (PCM). Such solutions, while promising and useful when space availability is a significant issue, are however not yet cost-competitive for large-scale applications.

4.14.2.2. HP substation LCoH

Putting together the different cost components, one can assess the LCoH of the HP substation. Using the mid-range values estimated above, namely c_{inv} = 1000 €/kW, a = 8.6 % (7 % WACC, 25 y lifetime), $r_{O\&M}$ = 3 %, c_{el} = 25 €/MWh (p_{el} = 125 €/MWh, COP = 5), and h_{eq} = 4000 h/y (50 % load factor, all waste heat recovered), one obtains the following mid-range value for the HP substation: LCoH = $c_{inv}(a + r_{O\&M})/h_{eq} + c_{el}$ = **54.00 €/MWh**.

The large variability of all parameters can obviously have a significant impact on this value. For example, under the worst conditions among the considered ones (high-end specific investment cost c_{inv} = 1300 €/kW, high-end annuity a = 10.2 % due to high WACC 8 % and low lifetime 20 y, limited operation hours h_{eq} = 2000 h/y, high-end electricity component c_{el} = 50 €/MWh due to high-end electricity price 150 €/MWh and low-range COP 3 as for air cooling), one would obtain LCoH_{subs} = 135.80 €/MWh, which is above common DH tariffs even without considering DH connection costs, and hence would make the project unfeasible economically. However, higher profitability would be reached under the best identified conditions (low-end specific investment cost c_{inv} = 700 €/kW, low-end annuity a = 7.3 % due to low WACC 6 % and long lifetime 30 y, high operation hours h_{eq} = 6000 h/y, low-end electricity component c_{el} = 14 €/MWh due to low-end electricity price 100 €/MWh and high-end COP 7 as for liquid cooling), where one would obtain LCoH_{subs} = 26.30 €/MWh. The very broad obtained **range 26-136 €/MWh** demonstrates the importance of properly analysing local conditions in order to assess the feasibility of a given heat recovery project.

4.14.3. Pipeline costs

The costs related to the pipeline for connecting to the DH network consist almost only of investment costs. Some minor operating costs can be related to pumping consumptions and thermal losses.

Investment costs

Investment costs for DH pipes can vary significantly. First of all, pipe costs depend on pipe size, which in turn depends on the power to be delivered. Pipe price also depends on insulation level, which must be chosen according to operating temperature. Moreover, significant differences in overall cost can depend on the city area where the network is installed (installation works in city centres are much more expensive than in city outskirts). Finally, country differences are also present, related to the degree of development of the local DH industry or to local urban structures and policies.

A general reference for pipe cost can be found in the well-known textbook by Frederiksen and Werner, "District heating and cooling" ([69]). It provides a general approximate formula to estimate the installation cost per unit length as a linear function of pipe diameter, i.e., $c_{pipe} = c_0 + c_1 D$, where D is the diameter in meters, c_0, c_1 are coefficients, and c_{pipe} is the final cost (installation included) per trench meter. The text also provides cost data for different city areas. For "outer city areas" (typically more relevant for large data centres) and for conventional pre-insulated steel pipes ([70]), reported values are about ([71]) 400/600/800 €/m for D = 100/200/300 mm, which correspond to c_0 = 200 €/m, c_1 = 2000 €/m^2. Since these costs were collected around 2007, present value estimates can be obtained by correcting for inflation: assuming a 2 % inflation rate until 2015 and then a 3 % inflation rate (Eurostat), an overall increase factor of about 1.6 can be estimated, finally yielding c_0 = 320 €/m, c_1 = 3200 €/m^2. According to Frederiksen and Werner's figures, the cost variability for inner city areas and park areas is about +/- 20 %.

The required pipe diameter can be determined according to standard design rules. Taking as a target two representative applications of 1 MW and 10 MW respectively and a supply-return temperature difference of 40 K, the resulting flow rates are about 21.5 m^3/h and 215 m^3/h. Assuming a design velocity of 2 m/s (Frederiksen and Werner), one would obtain nominal diameters of 65 mm (DN65) and 200 mm (DN200). The corresponding linear costs (using the above linear approximation with inflation-adjusted coefficients) are then approximately 530 €/m and 960 €/m. Considering all other sources of variability/uncertainty, the **500-1000 €/m range can be considered reasonably representative for DH piping costs** (towards the higher end for large-scale data centres). To obtain overall investment costs, this must be multiplied by the connection distance. Connections of 1 km and 5 km would then fall in the ranges of 0.5-1.0 M€ and 2.5-5.0 M€ respectively. In terms of unit cost, this corresponds to the values reported in Table 14.

([69]) S. Frederiksen and S. Werner, "District heating and cooling", Studentlitteratur, 2013.

([70]) For 5GDHC non-insulated plastic pipes can be used. While these pipes might be cheaper, larger diameters due to smaller supply-return temperature differences are required, so that no large cost reductions are anyway expected.

([71]) Original figure in SEK. It was assumed 1 SEK = 0.1 €.

Table 14 – Unit pipe costs

DH pipe line costs considering different transmission powers and distances (see text).

Unit cost [€/kW]		Power [MW]	
		1	**10**
Distance [km]	**1**	500	100
	5	2500	500

Operational costs

Pumping consumption can significantly depend on the type of control used for the network. For moderate distances (e.g., around 1 km), the experience of the district heating sector suggests that the electric energy used for pumping is typically no more than 2 % of the thermal energy being delivered. Using the heat pump terminology, this would correspond to an effective COP of 50 (ratio of thermal to electric energy). This was, for example, estimated in the FLEXYNETS project ([72]), where dependence on operating conditions was also taken into account.

Thermal losses in district heating depend on the type of pipe used, as well as on the operating temperatures. Their relative value with respect to delivered heat also depends on the general operating conditions (i.e., how much the pipe is used). For networks operated with high continuity (more common in northern climates) and with good insulation levels, thermal losses can be as low as 5 %, while for less performing networks losses can reach 20 % or more. Considering the continuous availability of data centre waste heat and its suitability to cover base load, thermal losses of about 10 % can be considered a rather conservative assumption.

As deduced from the previous estimates, pumping consumption and thermal losses can impact the final LCoH by a few percentage points, corresponding to their relative value with respect to the delivered heat.

4.14.3.1. Pipeline LCoH

As mentioned above, the electricity-related cost component for pipelines is expected to be small and most of the costs are related to investments. At the same time, pipe costs could even be negligible in special situations (small distance between data centre and DH network) or could be allocated to different projects (e.g., in cases where the construction of the new pipeline also enables additional user connections). For simplicity, only the case of a 10 MW data centre

([72]) Deliverable 3.2, "Integration of substations into DHC networks" (2018) of the FLEXYNETS project, Horizon 2020 programme, grant agreement No. 649820. Available at http://www.flexynets.eu/en/Results (visited August 2025).

with a 5 km new connection line is considered here with an investment cost of 5 M€. Using then c_{inv} = 500 €/kW and mid/range values for the other parameters, namely a = 8.6 % (7 % WACC, 25 y lifetime), $r_{O\&M}$ = 3 %, and h_{eq} = 4000 h/y (50 % load factor, all waste heat recovered), and neglecting electricity costs c_{el} = 0 €/MWh, one obtains the following value for the connection line: LCoH = $c_{inv}(a + r_{O\&M})/h_{eq}$ = **14.50 €/MWh**. This value would be slightly increased by including pumping costs and thermal losses.

4.14.4. Summary on cost estimates and possible revenue streams

The numbers reported in the previous sections confirm the interesting potential of heat recovery from data centres. At the same time, they make clear that LCoH is highly sensitive to input parameters. In particular, waste heat recovery from air-cooled data centres is economically viable only under favourable boundary conditions (e.g., low electricity price and low connection distance), while highly efficient liquid cooling systems can be profitable even under less favourable boundary conditions.

While the above analysis focused on costs, it is useful to highlight that the overall economic profitability also depends on additional revenue streams for DH operators beyond selling heat. These may include, for example, provision of cooling services to data centre cooling or participation in the electricity market (see chapter 6) by offering ancillary services enabled by HP flexibility.

Finally, it is worth mentioning the investment volume which could be involved by the large-scale application of heat recovery solutions from data centres to district heating, combining the costs estimated in this chapter with the estimates reported at the end of chapter 5. According to the previous sections, some representative values for investments costs are 1000 €/kW for HPs and 500 €/kW for connections (though a large variability was identified). Conservatively ([73]) assuming a uniform distribution of data centre waste heat along the year (8000 h/y of equivalent operation), the current 50 TWh/y estimated in chapter 5 would correspond to 6.25 GW of peak thermal power, yielding about 9.4 billion euros of investments. Similarly, the 150-300 TWh/y estimated for 2030 would correspond to 18.75-37.5 GW of peak thermal power and hence to an investment of €28-56 billion. Once more, while much uncertainty is implicit in these numbers, the possible significant impact of this sector is evident.

([73]) Assuming a smaller number of operation hours would lead to a higher peak power estimate for the same amount of waste heat, and hence to a higher investment volume.

Technical box – Estimate of a reference LCoH value

Combining the estimates reported in previous sections, one can estimate a **reference LCoH value of 68.50 €/MWh**. This is valid for the following assumptions (explained and justified in detail in the previous sections):

- A medium-efficiency liquid cooling system and a relatively efficient DH network (e.g., waste heat at 40°C, DH supply temperature at 75°C, heat pump COP = 5),
- An electricity price of 125 €/MWh,
- A 10 MW size HP,
- A connection line of 5 km.

Given the significant sensitivity of LCoH to input parameters, this estimate can be considered an exercise to identify an order of magnitude for a realistic case. Considering the qualitative threshold of 100 €/MWh for common DH tariffs, this reference value shows the relevance of considering heat recovery from data centres.

4.15. Examples of business cases and best practices

This chapter presents a selection of business cases and best practice examples that illustrate how data centres can actively contribute to sustainable energy systems. Business cases and examples are provided for waste heat recovery, various sector coupling initiatives and grid-support services. The examples, drawn from different European countries and beyond, highlight both technological diversity and contextual factors (such as climate conditions, regulatory frameworks, and the maturity of local district heating networks) that shape implementation strategies.

While detailed financial or contractual information is often confidential, the cases presented provide valuable insights into successful cooperation models, technical configurations and enabling conditions. Some of the cases present more detailed information because bilateral interviews or oral exchanges were carried out between the authors of this report and parties concerned, while other cases solely rely on publicly available information.

Together, these examples demonstrate the growing potential of data centres not only as digital infrastructure but also as integral components of the clean energy transition.

4.15.1. Business cases for district heating and other sector coupling initiatives: opportunities and barriers

Sector coupling is a wide concept referring to the strategic integration of different energy using sectors and the electricity sector. Energy using sectors include for example the heating, cooling, transport fuel and industry sectors. The purpose of sector coupling is to improve overall system efficiency and enable decarbonisation. In the following business cases and examples, different sector coupling examples are presented including:

- Waste heat recovery from data centres used for district heating.
- Waste heat recovery from natural gas sources used for electricity generation that, in turn, is used to power data centres.
- Data centres enabling new production sites for renewable energy.

The feasibility of using waste heat from data centres for district heating largely relies on specific local factors. Key considerations include climate conditions (such as ambient temperature), regulatory environment and infrastructure aspects like the development level of local district heating networks.

4.15.1.1. Waste heat recovery from the Leonardo supercomputer (CINECA, Tecnopolo Bologna, Italy)

CINECA ([74]) is a centre of excellence in the Italian and European ecosystem for supercomputing technologies, with offices in multiple Italian cities. The operational headquarters is located in Bologna, at the Casalecchio di Reno site. Moreover, Bologna hosts the CINECA Leonardo supercomputer located at the DAMA Tecnopolo. Both sites are in separate city zones and close to branches of the Bologna district heating network (less than 1 km away). However, while heat recovery projects had already started, internal heat recovery was prioritised due to the significant local heating loads. This offered a better economic convenience, avoiding the need to build new connection lines to the DH network. While the Casalecchio di Reno site also offers interesting potential (about 2.5 MW electric power), here the focus is on the larger Tecnopolo site. As it is a public structure, a significant amount of information is openly available ([75]).

- The Leonardo supercomputer has an electric power of about 10 MW, expected to grow in the future. The total data centre area is about 13,000 m^2, with about 2700 m^2 dedicated to server rooms. The yearly electricity consumption is about 48 GWh, of which about 40 GWh is devoted to IT appliances. The PUE is about 1.20 and the WUE about 0.23.

([74]) https://www.cineca.it (visited August 2025).

([75]) Technical data available at: https://cineca.amministrazionetrasparente.cineca.it/archiviofile/cineca/CINECA_Rendicontazione%20Direttiva%20EED%202023-1791.pdf (visited August 2025).

- The installed cooling technology consists of liquid cooling. The waste heat temperature is about 45 °C.
- Waste heat will be recovered to supply the Tecnopolo area, of about 100,000 m^2, with an internal network. The heat recovery target is 8 MW. At the moment, only a first phase has been completed, with a heat recovery < 100 kW (30 MWh/y in 2024).
- The internal Tecnopolo network will mostly exploit radiant panels for space heating. This will allow for low operational temperatures and no additional heat pumps are planned.
- While no detailed economic information was collected for this report, costs are expected to be towards the low range estimated in Chapter 7, thanks to the absence of heat pumps.
- No information on possible usage for flexibility services is currently available.

4.15.1.2. Integration of the Avalon 3 data centre into the A2A district heating network (Milan, Italy)

A2A is an Italian utility managing multiple DH networks in Italy. In Milan A2A owns multiple DH networks covering different city areas, and is expected to sign multiple agreements with local data centres. A first agreement has been signed with Retelit, for the Avalon 3 data centre ([76]).

- Avalon 3 is a colocation data centre, with a surface of 2500 m^2 and an electric power of 3.2 MW. It is one of the multiple data centres managed by Retelit in Milan.
- The installed cooling technology consists of air cooling. The waste heat temperature is hence about 25 °C.
- Waste heat will be mostly recovered to district heating, with a thermal power of 2.5 MW and through a heat pump. The plant is set to be operational at the beginning of 2026. A yearly heat production of about 15 GWh is expected (corresponding to 6000 h/y of equivalent operation).
- The connected DH network is the Milan West network, a high temperature network reaching up to 110°C in winter. However, the heat pump supply temperature will be limited to 90°C. Direct injection into the network supply pipe is expected to be feasible by mixing with higher temperature flow from other sources (also in consideration of the large peak power of the Milan network, above 200 MW).

([76]) https://www.gruppoa2a.it/it/media/comunicati-stampa/milano-primo-progetto-italia-recupero-calore-data-center (visited August 2025).

- No detailed information is currently available on costs, which are expected to be within the average range estimated in Chapter 7.
- The data centre is close to the existing network, so no significant extra costs for connection occurred.
- No information on possible usage for flexibility services is currently available.

4.15.1.3. High-temperature waste heat reuse from the Qarnot cloud HPC centre (Brescia, Italy)

Beside the Milan case, A2A also signed an agreement with a data centre in Brescia ([77]).

- This case is related to a cloud HPC centre managed by Qarnot. Two steps are planned: the first phase, already operational, includes 30 QBx computational units, the second phase (Qarnot2) will include a large scale-up.
- The installed cooling technology consists of liquid cooling. The waste heat temperature can hence reach up to 65 °C.
- Waste heat will be mostly recovered to district heating. Thanks to the high waste heat temperature, heat pumps will be avoided, limiting, however, the heat recovery to a pre-heating of the district heating return flow. The first phase of the project allows for a thermal power of 100 MW and a yearly heat output of 800 MWh (a highly continuous operation is indeed expected, around 8000 h/y). The second phase of the project is expected to reach a power of 2 MW and a yearly heat supply of 16 GWh.
- The connected DH network is the Brescia network, a large network with more than 22,000 customers. It is operated at high temperature, exceeding 100 C in winter. As mentioned, however, heat recovery will only be used for pre-heating up to 65°C.
- No detailed information is currently available on costs, which are expected to be towards the low range estimated in Chapter 7, thanks to the absence of a heat pump.
- The data centre is close to the existing network, so no significant extra costs for connection occurred.
- No information on possible usage for flexibility services is currently available.

([77]) https://www.a2acaloreservizi.it/it/media/comunicati-stampa/a2a-calore-green-brescia-primo-data-center-teleriscaldamento (visited August 2025).

4.15.1.4. ReUseHeat project: data centre waste heat recovery for a new residential area (Braunschweig, Germany)

In the city of Braunschweig (Brunswick), an innovative project has transformed a data centre's waste heat into a sustainable source of district heating. Developed within the EU Horizon 2020 ReUseHeat project, this initiative brings together Veolia and the local utility BS ENERGY.

The idea emerged when a new residential neighbourhood was being planned next to a newly built data centre. Recognising the opportunity, BS ENERGY designed a system that captures the data centre's waste heat and integrates it into a nearby 4th-generation district heating network. This approach benefits both sides: the data centre improves its cooling efficiency, while local households receive climate-friendly heat.

The process is straightforward yet technically advanced. The water circulating through the data centre's cooling system reaches about 25°C and is channelled to a nearby energy station. There, a CO_2 heat pump (Thermoeco, Engie) increases the temperature to around 70 °C (return from district heating at 40°C), making it suitable for residential heating. The cooled water then returns to the data centre to absorb more heat, creating a closed and efficient loop.

The recovered heat supplies roughly 600 housing units, covering around 48,000 m² of living space. The system primarily meets the base load demand, while any peak needs are supported by Braunschweig's existing high-temperature district heating network. In total, the system distributes heat through about 2.75 km of new district heating pipes and includes connections to the wider network (a well-established urban system supplying roughly 800 GWh of heat per year with an installed thermal capacity of around 320 MW and a pipeline length of about 260 km) for flexibility.

With a coefficient of performance (COP) of 3.6, the heat pump produces around 2,300 MWh of heat per year, of which 1,750 MWh comes directly from the data centre's waste heat. This results in annual energy savings of around 1,280 MWh and CO_2 reductions of approximately 300 tonnes compared to conventional gas heating.

The project is operated under a 10-year heat supply contract between BS ENERGY and the property developer, which is later transferred to individual homeowners ([78]).

([78]) 19GDECA-Desc-ReUseHeat-BraunschweigDE.pdf and Heat recovery from data centres - Brunswick, Germany - Euroheat & Power (visited August 2025).

4.15.1.5. Feasibility and barriers for waste heat use at 1&1 data centres (Germany)

The company 1&1, one of Germany's largest telecommunications and internet services provider, operates around 1,200 technical sites of different sizes in the country, including five large data centres of about 1,000 m² each, located in Berlin, Munich, Mainz, and North Rhine-Westphalia.

Cooling is primarily managed through conventional split cooling systems, with average cooling temperatures around 25°C. Unfortunately, the waste heat temperature of the cooling cycles, although not specifically monitored, is too low for direct integration into district heating networks, without additional technologies, such as heat pumps. Discussions with potential partners confirmed that feasibility is limited, largely due to the low temperature, the need for additional infrastructure, and the distance to existing district heating grids.

Flexibility services within the data centres are mainly linked to server optimisation for 5G mobile networks, rather than energy system services such as demand response or load shifting.

Current sustainability activities of the company focus on improving efficiency in cooling operations, experimenting with raising operating temperatures up to 26°C, and reducing electricity consumption. Other measures include optimising heat flows and adopting decarbonisation strategies such as replacing old lighting systems with LEDs.

The main barriers encountered are technical rather than regulatory. Temperature levels, grid connections and the placement of heat pumps are decisive factors. Legislation could become a driver if waste heat use were made mandatory, but currently, the company sees no regulatory obligations in this area.

However, from 1 July 2026, new German regulations will oblige data centres that begin operations to demonstrate an energy reuse factor (ERF, defined in DIN EN 50600-4-6) of at least 10 %. This threshold rises to 15 % in July 2027 and 20 % in July 2028, calculated on an annual average. Exemptions are envisaged in cases where operators have already reached agreements with municipalities or district heating companies for the later utilisation of waste heat. The overarching aim of the legislation is to steer new data centres towards locations where their waste heat can be effectively recovered and fed into local energy systems, but the practical and technical hurdles, as identified by 1&1, remain: municipal district heating networks in fact typically operate with flow temperatures in the range of 65 to 90°C. By contrast, the waste heat from cooled 1&1 data centres remains significantly lower, even during peak operation. As a result, supplementary heat pumps are necessary to raise the temperature to a level suitable for integration into district heating systems.

4.15.1.6. Waste heat reuse from the NTT FRA4 data centre in cooperation with Mainova AG (Hattersheim am Main, Germany)

The NTT FRA4 data centre campus in Hattersheim am Main, located about 15 km southwest of Frankfurt, stands out as a leading example of how data centre waste heat can be effectively integrated into a local district heating network. The project is a collaboration between NTT Global Data Centers, Mainova AG, and the municipality of Hattersheim, demonstrating how digital infrastructure can contribute to local energy transitions and decarbonisation efforts.

FRA4 is a large-scale, multi-building campus designed with sustainability in mind. Once fully operational, it will provide up to 82 MW of IT capacity across 32,900 m² of IT space. From the outset, the facility was planned with heat recovery via high-efficiency heat pumps, allowing the continuous waste heat from servers to be captured and reused. The cooling water from the data centre, typically around 30°C, serves as the heat source for the system.

Mainova has constructed an energy centre directly connected to the data centre, where two large high-temperature heat pumps raise the waste heat to 70–75°C—sufficient for district heating supply. This recovered heat is then fed into a newly built local heating network serving both a new residential area and an existing housing district (Hugo-Hoffmann-Ring). The network extends over roughly 1.1 km and also integrates existing gas CHP units and peak-load boilers to ensure reliability during periods of exceptionally high demand or maintenance.

To further reduce the project's carbon footprint, the system is powered in part by rooftop photovoltaic panels installed on the data centre, providing around 10 % of the electricity required for the heat pumps.

Each year, approximately 5.2 GWh of useful heat—corresponding to about 2.5 MW of heating capacity—will be delivered to more than 600 homes. The recovered heat is available year-round, ensuring a constant, low-carbon heat supply for the community. This initiative is expected to avoid around 570 tonnes of CO_2 emissions annually compared to conventional gas heating, while achieving a primary energy factor of about 0.32 ([79]).

4.15.1.7. Circular heat use from the LUMI supercomputer for local district heating (Kajaani, Finland)

LUMI (Large Unified Modern Infrastructure) is one of the most powerful and environmentally efficient supercomputers in the world, located in Kajaani, Finland, hosted by CSC. It is designed for high-performance computing and artificial intelligence workloads, supporting cutting-edge research in fields such

([79]) NTT Global Data Centers liefert Abwärme und Mainova bringt diese auf Heiztemperatur and Frankfurt 4 Data Center | NTT (visited August 2025).

as climate science, cancer diagnostics and materials physics. LUMI is built on the HPE Cray EX architecture and powered by AMD EPYC CPUs and AMD Instinct MI250X GPUs, with a peak performance exceeding 531 petaflops and 10 MW. The data centre runs entirely on green renewable power and repurposes excess heat to warm nearby buildings and the DH system for Kajaani. Here heat pumps are used to meet the supply DH temperature. LUMI is part of the AI-factory and will be expanded with a further 30 MW ([80])([81]).

4.15.1.8. Heat recovery from the Facebook data centre to the Fjernvarme Fyn network (Odense, Denmark)

Facebook's hyperscale data centre in Odense recovers up to 100,000 MWh of waste heat annually. The heat is transferred to the Fjernvarme Fyn district heating system, supplying residential heating. The facility is powered by 100% renewable energy, mainly from local wind farms ([82]).

4.15.1.9. Stockholm Data Parks: large-scale integration of data centre heat into the district heating grid (Stockholm, Sweden)

A public-private initiative where data centres (e.g., Bahnhof) supply waste heat to Stockholm's district heating grid. Capable of heating over 120,000 homes, reducing CO_2 emissions by up to 100,000 tons per year. Operators receive financial compensation for heat delivery, improving ROI and sustainability alignment ([83]).

4.15.1.10. Digital heat strategy for urban decarbonisation in HafenCity East (Hamburg, Germany)

Waste heat from a local data centre is used to heat residential and public buildings. A digital heat strategy enables up to 70% emission reduction, with strong municipal-private collaboration ([84]).

([80]) https://www.itpro.com/infrastructure/inside-lumi-one-of-the-worlds-greenest-supercomputers (visited August 2025).

([81]) https://csc.fi/en/our-expertise/high-performance-computing/lumi-supercomputer/ (visited August 2025).

([82]) https://www.irena.org/Innovation-landscape-for-smart-electrification/Power-to-heat-and-cooling/31-Waste-heat-recovery-from-data-centres (visited August 2025).

([83]) https://eutech.org/harnessing-waste-heat-from-data-centers-driving-sustainability-and-compliance-in-europe/ (visited August 2025).

([84]) https://eutech.org/harnessing-waste-heat-from-data-centers-driving-sustainability-and-compliance-in-europe/ (visited August 2025).

4.15.1.11. Data centres reducing natural gas flaring through onsite power use (Crusoe Energy, USA)

The Cambridge Digital Mining Industry Report ([85]) gives an example where data centres are used to reduce natural gas flaring in the USA (*Crusoe Energy mitigating natural gas flaring to power computation*).

4.15.2. Business cases for flexibility services supporting the power grid

Data centres supporting the power grid is a relatively new practice and descriptions of business cases are therefore more prevalent in popular science newspaper articles and journalistic web articles than in academic papers.

However, academic reports and papers have reported, in general terms, that the integration of flexible loads (such as some types of data centres) offers a short-term strategy for addressing structural transformations in power grids. For example, Duke University published a report analysing "a new area of electricity demand" in the USA ([86]). Key results include e.g. that 10% of the USA's current aggregate peak demand could be integrated with a low load curtailment rate of 0.25%.

For grid stabilisation, several measures or systems are available: for example fast frequency reserve (FFR), frequency containment reserve (normal FCR-N, and disturbance FCR-D), and frequency restoration reserve (automatic aFRR, and manual mFRR). All of the systems are more or less united across Europe and all of them with different response time, power capacity and duration time. The majority of data centres participating in grid service stabilisation are part of the FCR-D up reserve, which is activated when the grid frequency deviates 0.1–0.5 Hz below the target value of 50 Hz (in European grids). This disturbance occurs when the demand for electricity is higher than the current electricity generation. The measure to raise the grid frequency back to normal is to decrease the demand for electricity, e.g. by decreasing the power use of data centres.

The two most common ways for data centres to be active as grid stabilisation is by:

- Activating the UPS system going from grid power supply to battery power supply. This setup is mostly used by enterprises and co-location data centres.

([85]) Neumueller et al 2025: Cambridge Digital Mining Industry Report: Global Operations, Sentiment, and Energy Use.

([86]) Duke University (2025): Rethinking Load Growth–Assessing the Potential for Integration of Large Flexible Loads in US Power Systems.

- Shutting down servers or pausing their work, which is mostly used by HPC and crypto mining data centres.

In particular, the FCR cooperation consists of 12 transmission grid operators (TSOs) in nine countries across Europe ([87]). This enables replication of the same system and qualification of the markets for data centre operators with several sites in these countries.

4.15.2.1. UPS-based frequency and fast reserve participation by Conapto (Stockholm, Sweden)

Conapto is a co-location data centre in Sweden that participates in fast frequency reserve (FFR) and frequency containment reserve (FCR) markets. Uses UPS systems with lithium-ion batteries to provide rapid response to grid frequency deviations. Monetises backup capacity while supporting grid stability and decarbonization goal ([88]).

4.15.2.2. Load flexibility through compute curtailment by HIVE Digital Technologies (Sweden)

HIVE, a Canada-based data centre owner with facilities in several countries, supports Sweden's energy infrastructure by providing demand and frequency response services, helping to balance the national grid and enhance its reliability. By participating in the FCR-D up market by direct shut down of their compute power use, which is done without installation of additional infrastructure ([89]).

4.15.2.3. Grid-interactive UPS systems providing fast frequency response (Eaton and Microsoft, Ireland and Nordics)

Eaton's grid-interactive UPS systems deployed in Microsoft data centres. Provides fast frequency response (FFR) and under frequency load shedding (UFLS). Enhances grid resilience, supports renewable integration, and generates new revenue stream ([90]).

([87]) ENTSOE: Frequency Containment Reserves (FCR). https://www.entsoe.eu/network_codes/eb/fcr/

([88]) https://www.datacenterdynamics.com/en/news/vertiv-brings-grid-supporting-ups-to-conapto-data-centers-in-sweden/ (visited August 2025).

([89]) https://www.hivedigitaltechnologies.com/operations/sweden/ (visited August 2025).

([90]) https://www.ercot.com/files/docs/2025/07/10/Eaton-Data-center-A-Good-Grid-Citizen.pdf (visited August 2025).

4.15.2.4. Dynamic grid support and energy arbitrage using Liebert® EXL S1 UPS (Vertiv, United Kingdom and EU)

Liebert® EXL S1 UPS with dynamic grid support with static and dynamic frequency regulation together with peak shaving and energy arbitrage. Enables participation in ancillary services markets and demand-response programs ([91]).

4.15.2.5. ECO Data Centre: reuse of excess heat for sustainable pellet production (Falun, Sweden)

The excess heat of ECO DC, a Swedish co-location data centre, is used to support the power plant's pellet production, which has a positive impact on the local community by supporting the development of sustainable energy solutions ([92]).

4.15.2.6. GothNet co-location data centre supplying waste heat to a local greenhouse (Gothenburg, Sweden)

GothNet part of the Gothenburg Energy consortium is operating a co-location data centre that is heating a nearby greenhouse of 120 m^2 that is used for supporting the sufficiency by near produced herbs and vegetables ([93]).

4.15.2.7. Genesis Digital Assets: crypto mining heat recovery for social and agricultural integration (Boden, Sweden)

GDA operates a crypto mining data centre in Boden that is conveying its heat to a greenhouse of 300 m^2, which is used for local farming of vegetables and herbs. It runs buy a societal integrating company with the aim to introduce people to meaningful work and everyday occupation, which result in a society with less exclusion ([94]).

4.15.2.8. Demand-side response supporting grid stability by Riot Platforms (Texas, USA)

Riot Platforms operates large-scale data centres and bitcoin mining applications. The Cambridge Digital Mining Industry Report ([95]) describes the business case

([91]) VERTIV WHITE PAPER: How to Maximize Revenues from Your Data Center Energy Storage System with Grid Interactive UPS.

([92]) https://ecodatacenter.tech/sustainability-data-center (visited August 2025).

([93]) https://www.goteborgenergi.se/i-var-stad/artikelbank/goteborgsk-chili-och-ingefara-pa-spillvarme-fran-datacenter (visited August 2025).

([94]) https://www.sunderbyfolkhogskola.se/projekt/dc-farmers/ (visited August 2025).

([95]) Neumueller et al 2025: Cambridge Digital Mining Industry Report: Global Operations, Sentiment, and Energy Use.

for Riot Platforms where their data centres in Texas are offered as a demand-side response resource for the ERCOT power grid. It is stated that the data centre operations possess a greater ability to rapidly modulate power consumption than traditional industrial loads. A specific example is given where the power consumption was curtailed during a heat wave, contributing to system stability by offering these grid ancillary services.

4.15.2.9. Flare-gas powered HPC mini data centre (Gothenburg, Sweden)

Data centres can also use otherwise wasted energy sources to support a more sustainable energy system. For example, a mini plant of shipping container size has been started in Gothenburg (Sweden [96]) to verify that a HPC data centre (AI computations) can be powered by heat from flare gas. The pilot is located at a landfill site where methane gas is flared to, among other things, reduce its environmental impact by releasing CO2 instead of methane. The company Texel uses the heat from the flaring to generate electricity in a Stirling engine, and the electricity is used for powering a local data centre. The data centre is connected wirelessly to the internet, but the electricity production is off-grid meaning a much cheaper, faster and easier startup process. This is a small facility showcasing the possibilities by using landfill gas for generating electricity off grid. The electricity could be used for other purposes that do not need a power grid connection. There are many landfills in the EU where this technology can be used to generate electricity from otherwise wasted energy.

4.15.2.10. Renewable micro-grids enabled by data centres (Zambia)

Data centres can support a sustainable energy system in more ways than stabilising the frequency of the power grid. For example, they can also support renewable microgrids. A concrete business case is described in Zambia ([97]), where new renewable energy sites have been put in place thanks to co-locating a crypto-mining data centre. In these rural communities, there is no connection to a larger power grid. Without the mining facility, it would not be economically feasible to produce solar electricity here. These sites are usually relatively small in scale but important for the local community.

([96]) TogetherTech (2025): TEXEL Turns Flare Gas into Green AI (https://www.togethertech.com/en/case/texel-turns-flare-gas-green-ai)

([97]) BBC World Service 2025: Bitcoin in the bush – the crypto mine in remote Zambia.

4.15.2.11. DCFlex initiative: advancing data centre flexibility for grid integration (international consortium)

The DCFlex initiative is a pioneering effort to demonstrate how data centres can play a vital role in supporting and stabilising the electric grid while enhancing interconnection efficiency. It aims to drive a cultural, taxonomic, and operational transformation across the data centre ecosystem, offering a comprehensive blueprint for stakeholders including utilities, market operators, technology innovators, and policymakers. Through four focused workstreams, DCFlex explores the technical, economic and regulatory dimensions of data centre flexibility. It contains 4 workstreams:

1. Grid-informed Flexible Data Centre Designs
   Enabling future data centres to become grid resources through flexible & efficient designs and operational practices.
2. Transformational Utility Programme
   Exploring market and programme structures that advance data centre flexibility.
3. Grid Planning for Operational Flexibility
   Equipping the utility industry planning practices to embrace large flexible loads
4. Data Centre Informed Energy Supply
   Guiding the development of energy supply resource development.

The consortium consists of about 40 members including companies like Google, Meta, Microsoft, Nvidia ([98]).

4.16. Recommendations

4.16.1. Policy background: Heat recovery and flexibility measures from data centres in the EU legislation

4.16.1.1. EU governance of the heating and cooling sector

Currently, there is no single EU policy tackling the heating and cooling sector. The 2016 H&C strategy is now outdated, and a major revision is expected for 2026. Policies tackling H&C are scattered across EU legislation (i.e. from the Energy Efficiency Directive and the Renewable Energy Directive to the Energy Performance of Buildings Directive).

([98]) https://dcflex.sf.epri.com/ (visited August 2025).

Specifically, the revised Energy Efficiency Directive (EED 2023/1791) and Renewable Energy Directive (REDIII, 2023/2413) aim to increase the share of renewable and recycled energy and to improve the energy efficiency of energy systems. For district heating and cooling, this means progressively stricter requirements for primary energy performance and a growing integration of renewable energy and waste heat/cooling.

The EED defines “efficient district heating and cooling” based on minimum and progressively increasing shares of renewable energy, waste heat/cooling and high-efficiency cogeneration (EED art. 26). REDIII complements this framework by acknowledging that waste heat and cool supplied through efficient systems can be counted towards energy efficiency and renewable energy targets. Thus, heat recovery from data centres can directly contribute to the EU’s overall energy and climate objectives when integrated into efficient district heating and cooling networks.

4.16.1.2. Waste heat from data centres – definition and policy relevance

The EU framework defines waste heat as “unavoidable by-product heat” that would otherwise have been cooled (RED art. 2.9). In order for waste heat from data centres to be counted in statistics and targets, it must be delivered to an efficient district heating or cooling system. Heat from waste incineration in CHP is generally considered co-generated heat (cogeneration) and is therefore not covered by the waste heat definition.

Data centres generate significant amounts of low to medium temperature excess heat, in line with growing digitalisation, AI and rack power. When this heat is upgraded to usable temperatures and integrated into efficient grids, it contributes to meeting both EED and REDIII targets.

4.16.1.3. Reporting and planning requirements

The EU, via the Commission Delegated Regulation C/2024/1364 on the first phase for establishing an EU-wide scheme to rate the sustainability of EU data centres ([99]), requires data centre operators to measure and report energy performance, renewable energy share, heat recovery, cooling performance and water use. When establishing new data centres, the heat recovery potential should be assessed already in the location and design phase, including the possibility of supplying heat to nearby district heating networks. Additionally, under Article 12 of the recast Energy Efficiency Directive, data centres with

([99]) EU, ”Data centres in Europe”, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13818-Data-centres-in-Europe-reporting-scheme_en (visited August 2025).

installed power > 500 kW will also be required to publish information on their energy performance and sustainability.

For larger district heating networks (e.g. > 5 MW), planning requirements according to the EED apply (EED, Art. 26): actors must show how they increase the share of renewable/recycled and meet efficiency criteria over time, with a clear policy anchored to the EED/RED. Article 26 (paragraphs 6 and 7) of EED also explicitly requires data centres with a power input exceeding 1 MW utilise the waste heat or other waste heat recovery applications unless they can show through an "installation level cost-benefit analysis" that it is not technically or economically feasible.

According to Articles 25.6 of the EED, Member States must ensure that municipalities with more than 45,000 inhabitants develop local heating and cooling plans. These must be designed to ensure a cost-effective, secure and climate-sustainable supply and use of heating and cooling, in line with the "Energy Efficiency First" principle. The plans must, among other things, identify the potential for increased energy efficiency, integration of renewable and recycled energy and opportunities for sector coupling and flexibility in the energy system. This means that municipalities have a clear responsibility to analyse and plan how heating and cooling systems can contribute to both national and European energy and climate targets.

Beside general EU directives, energy efficiency planning and reporting for data centres are also promoted through dedicated EU initiatives, see, e.g., DG ENER ([100]) and JRC ([101]).

4.16.2. Recommendations for future actions

In the following, a list of recommendable actions to foster the heat reuse and flexibility potential of data centres is provided. Recommendations were developed mostly from a broad energy sector perspective and are listed in non-prioritised order.

- Integration of heat mapping and feasibility study requirements into data centre planning and permitting processes.
    - *Identified barriers*. Fragmented coordination between data centre developers and district energy operators at early development stages; data centre site is decided without heat recovery analysis (Chapters 5, 7).

([100]) EU, "Assessment of next steps to promote the energy performance and sustainability of data centres in EU", https://op.europa.eu/en/publication-detail/-/publication/29aa14ed-a4ba-11f0-a7c5-01aa75ed71a1/language-en (visited October 2025).

([101]) JRC, "2025 Best Practice Guidelines for the EU Code of Conduct on Data Centre Energy Efficiency", https://publications.jrc.ec.europa.eu/repository/handle/JRC141521 (visited October 2025).

- *Relation with current EU policy*. EED's provisions on planning for heating/cooling and efficient DHC; REDIII's enabling conditions for recovered heat. Partly already addressed by EED Art. 26, which requires cost-benefit analysis for waste heat reuse.
- *Recommended actions*. Promote the development of guidelines and best practices for feasibility studies for heat recovery and connection options. Promote the exploitation and expansion of data centre and district heating area maps, for an easier matching. Develop standard memorandum-of-understanding and agreement templates for heat contracts, leading to faster negotiation.

- Support for temperature-efficient solutions: introducing temperature considerations into data centre cooling systems, criteria for efficient DHC networks, and municipal heating and cooling plans (including building level).
  - *Identified barriers*. High district heating supply/return temperatures hinder integration of low/medium temperature waste heat (Chapter 5).
  - *Relation with current EU policy*. EED's transition towards efficient DHC; link to EPBD and system efficiency.
  - *Recommended actions*. Require local temperature roadmaps (towards 4th/5th generation DHC) and promote development of connection guidelines with temperature ranges and incentives for low return temperatures from buildings. Promote liquid cooling in synergy with the driver given by increasing server power, e.g., provide conditional incentives for liquid cooling (depending on connection to district heating and available power).
- Creating market arrangements that value recovered heat and flexibility.
  - *Identified barriers*. Limited compensation for system services (such as fast-ramping of large heat pumps or installation of thermal storages) and for avoided emissions (Chapters 6–7). Risks for the district heating operator about long-term waste heat availability.
  - *Relation with current EU policy*. EED's energy efficiency first, REDIII targets, electricity market design for demand response.
  - *Recommended actions*. Set tariffs/credits for (i) recovered heat into efficient grids and (ii) grid services from heat pumps/storage (demand response, ancillary services). Promote contract templates. Identify risk-mitigation mechanisms for waste heat recovery, e.g., providing guarantees/insurances for district heating operators installing waste heat recovery measures with internal investments. Support third-party solutions like energy service companies.

- Further standardisation of reporting metrics and transparency across data centres.
    - *Identified barriers*. Additional possible indicators (e.g., pPUE, ERE) with respect to PUE/ERF/WUE; potential issues on how HP electricity consumption is allocated (Chapters 4, 6).
    - *Relation with current EU policy*. EU reporting requirements for data centres (2024) on efficiency, renewables and heat recovery. New European data centre database and related reporting requirements.
    - *Recommended actions*. Promote harmonised definitions (clear system boundaries for PUE when HP is primarily used for heat recovery). Promote collection on possibly (anonymised) open data on data centre consumption and temperature profiles (at hourly level) to better assess matching with district heating load and flexibility potential. Evaluate reporting of possible flexible load of data centres to increase flexibility awareness. Classify data by data centre type (also in terms of usage, e.g., HPC) and cooling technology (e.g., liquid cooling).
- Support for large-scale heat pumps.
    - *Identified barriers*. High initial CAPEX for HPs/substations/pipes, sensitive economics (electricity price, operating hours) (Chapter 7). Evolving regulation about refrigerants.
    - *Relation with current EU policy*. State aid rules, EU taxonomy, financing instruments for energy efficiency and renewable energy.
    - *Recommended actions*. Support diffusion of innovative financing mechanisms (e.g., blended finance). Publish guidelines including reference cost benchmarks. Promote synergies for electricity contracts between data centres and district heating operators (when the latter owns the heat pumps). Promote best practices for procurement requirements: e.g., imposing equipment (e.g., heat pumps) to be adequately sized for the entire operating area (in terms of temperature and power range) and have an appropriate average COP, as well as imposing COP follow-up monitoring. Clarify F-gas and PFAS regulation and support research on compliant refrigerants.
- Development of guidelines for location and co-location of data centres near heat loads.
    - *Identified barriers*. Long connection distances drive up LCoH; weak connection between physical planning and thermal networks (Chapters 5, 7).
    - *Relation with current EU policy*. EED's planning for heating/cooling; REDIII's system integration.

- *Recommended actions*. Integrate heat demand/supply mapping in spatial planning; prioritize data centre locations within economically reasonable distances to existing/planned DHC.

- Develop guidelines and support knowledge activities for implementing demand-side flexibility services from data centres in Europe.
  - *Identified barriers.* The main barrier for data centres to implement flexibility services is uncertainty. One example is that the prospective income from ancillary services and flexibility markets are different in different regions of Europe, and it can differ within a country. The income is also highly volatile. For example, the compensation for being active on the FCR-D market in Sweden decreased by a factor of 10 in 1–2 years. In addition, not all data centre types are suitable for flexibility services.
  - *Relation with current EU policy.* EED's energy efficiency first, REDIII targets, electricity market design for demand response. Flexibility services are not regulated by the EU; they are made on a voluntary basis.
  - *Recommended actions.* There is a need for two-way knowledge transfer between power grid operators and data centre operators on what is technologically possible to achieve on flexible data centre operation. Promote more formal collaborative efforts between data centre operators and power grid operators. There are today informal discussions going on between power grid operators and data centre operators to increase knowledge and together find ways on how these actors can work together on increasing the grid capacity by e.g. flexible power use while assuring both grid stability and datacentre operational stability.

- Standardised technical connection rules for DHC integration.
  - *Identified barriers*. Tailor-made technology and contracts extend lead times and costs (Chapters 5, 7).
  - *Relation with current EU policy*. EED's guidance on efficient DHC and best practice framework.
  - *Recommended actions*. Develop model rules (temperature range, metering, quality of supply) and standardised commissioning for connections between data centres and district heating. Promote research and guidelines for replicable heat pumps configurations, classifying standard cases (e.g., heat recovery from air cooling or from liquid cooling).

- Promotion of transparent cost of service methods.
  - *Identified barriers*. Varying cost/benefit sharing between data centre operators and DHC companies (Chapter 7).

  - *Relation with current EU policy*. The principles of energy efficiency first and cost transparency.
  - *Recommended actions*. Recommend reference methodology for LCoH calculation (handling electricity price hedging, assumptions about operating hours and heat losses). Promote development of guidelines or templates for business and financial models in the sector.

- Strengthening cross-sectoral governance and data exchange.
  - *Identified barriers*. Siloed decisions between ICT, energy companies and urban planners (Chapters 5–6).
  - *Relation with current EU policy*. Integrated planning according to EED; digitalisation initiatives. Amended version of Art. 24 of EED, about establishing coordination frameworks among sector stakeholders.
  - *Recommended actions*. Establish collaboration platforms (ICT–DHC–networks, grid operators) with shared datasets (temperatures, capacities, load profiles). Concerning flexibility, promote replication of similar systems and market qualification rules, e.g. starting with data centre operators with multiple sites in countries involved in the FCR cooperation.
- Exploring the development of “green calculation” certificates or similar guarantee of origin.
  - *Identified barriers*. Even though feasible only for selected data centre types (e.g., HPC, where computing job scheduling is feasible), there is an unexploited demand flexibility potential (chapter 6).
  - *Relation with current EU policy*. EU reporting requirements for data centres (2024) on efficiency, renewables and heat recovery. New European data centre database and related reporting requirements.
  - *Recommended actions*. Enforce classification of data centres to identify them with schedulable tasks/jobs. Support the clarification of the impact of electricity costs on overall computing costs (e.g., requiring data centres to declare the share of electricity costs with respect to overall costs). Support the investigation of carbon-aware scheduling and of pricing schemes enabling tariffs related to the day period and the type of exploited electricity, possibly yielding corresponding certificates.

Heat recovery from data centres is a priority in EU energy and climate policy. By integrating surplus heat into efficient district energy systems, stakeholders can contribute to energy efficiency and renewable energy targets while strengthening system flexibility and resource efficiency. Governance, planning and business

models should be adapted to the requirements and opportunities established by the already existing directives and by the ongoing EU initiatives.

5. Task 5: Analysis of the Open Public Consultation Results for the Strategic Roadmap

The analysis of the Open Public Consultation results for the Strategic Roadmap, developed under Task 5 of this study, has already been published separately on the ENTEC2 webpage as of January 2026. The document can be accessed and downloaded in PDF format through the link provided under the “Study Planning” section, which redirects to the European Commission Publications Office. For ease of reference, the direct link to the publication is provided below:

https://op.europa.eu/en/publication-detail/-/publication/9a04385e-f356-11f0-b9bc-01aa75ed71a1/language-en

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696,

– via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.



Publications Office
of the European Union