ukai

AI, Energy and Infrastructure

Building a Green AI Superpower

Contents

ukai

Executive Summary

This report argues that the UK's main AI opportunity is to become a world leader in Green AI, i.e., artificial intelligence that is energy efficient, affordable, deployable at scale, and aligned with long-term economic and environmental sustainability. Green AI is not a niche concern or a constraint on growth. It is a source of competitive advantage in a world where energy, infrastructure capacity, and public trust are becoming decisive factors in where AI investment and value are realised. More importantly, the market for technology that delivers efficiency is global and growing; this demand could provide an important hedge to future market corrections.

As AI systems scale, their dependence on electricity, grid capacity, cooling, connectivity, and skilled labour is growing rapidly. At the same time, the UK faces high and volatile energy prices, long grid connection times, fragmented infrastructure planning, and increasing competition for global AI investment. These pressures mean that the success of the UK AI sector now depends less on algorithmic breakthroughs alone and more on whether the surrounding systems are fit for purpose. Without action, there is a real risk that the UK will continue to generate AI innovation while losing deployment, investment, and economic value to other jurisdictions.

UKAI, as the trade body for AI businesses in the UK, has focused on this challenge because it is fundamental to unlocking sustainable growth for the sector and, thus, realising AI's potential to improve the lives of people in the UK. Through two industry roundtables with AI companies, infrastructure providers, investors, and energy specialists, a consistent message emerged: the constraints on AI growth are systemic, and they cut across infrastructure, markets, innovation, and coordination. This report translates those insights into a set of practical, industry-informed recommendations.

The report is structured around four mutually reinforcing chapters.

Chapter 1, Integrated Infrastructure, argues that AI infrastructure must be treated as a connected system rather than a collection of isolated assets. It highlights the need for stronger cross-government coordination, clearer national planning signals for AI and

compute infrastructure, earlier integration of energy and connectivity into project design, and greater transparency around environmental performance. Recommendations include establishing a dedicated delivery unit for major AI infrastructure, aligning planning reform with grid and skills constraints, and embedding community benefit and legitimacy into infrastructure development.

Chapter 2, Fairer Pricing, focuses on energy markets as one of the most binding constraints on AI deployment. While clean generation costs are falling, AI operators remain exposed to high and unpredictable prices driven by marginal pricing, network congestion, and system inefficiencies. The chapter sets out how targeted market reform – alongside investment in grid reinforcement, storage, and flexibility – could reward efficient, flexible users and reduce risk for long-term investment. It also explores the role of long-term power arrangements and public institutions in stabilising prices with fewer blunt subsidies.

Chapter 3, Targeted Innovation, makes the case that the UK should lead in making AI more energy-efficient, rather than competing on sheer scale. It highlights opportunities in hardware design, software optimisation, system-level efficiency, and application-layer innovation. Recommendations include prioritising energy-efficient performance as a criteria for public funding, requiring application-level energy and carbon metrics, and supporting modular and distributed compute models that reduce waste and overbuild.

Chapter 4, Scalable Systems, addresses how these reforms are turned into impact. It argues that ecosystems supported by shared standards, testbeds, coordinated procurement, and skills pipelines are essential to scaling implementation. The chapter sets out how the Government can act as a steward of ecosystems, how investment risk can be reduced through standardisation and clarity, and how the UK can create a significant export opportunity in developing sustainable AI systems.

Taken together, the report sets out a practical pathway for the UK to lead globally in Green AI, driving economic growth and positive social impact. With a clear mission from the Government, informed by industry and supported by academia, the UK can align infrastructure, pricing, innovation, and ecosystems to unlock growth while strengthening energy resilience and public trust.

Introduction

Artificial intelligence is increasingly recognised as a general-purpose technology with the potential to transform productivity, public services, scientific discovery, and economic growth. However, the success of AI depends far less on algorithms alone and far more on the systems that surround them: energy, infrastructure, markets, skills, and institutions. For the UK, the challenge is no longer whether it can invent or adopt AI, but whether it can deploy it at scale in ways that are economically competitive, environmentally sustainable, and publicly legitimate.

UKAI, as the trade body representing AI businesses across the UK, has become increasingly focused on this question because it sits at the heart of the sector's future growth. Our members range from early-stage startups to established scale-ups, infrastructure providers and the businesses that use AI, but they share a common concern: the constraints holding back AI deployment in the UK are now structural rather than technical. High and volatile energy prices, fragmented infrastructure planning, unclear signals for investment, and slow pathways from innovation to deployment are shaping where AI models and tools are built, trained, and operated. If these issues are not addressed, the UK risks seeing the economic value of its AI capability realised elsewhere.

To explore these challenges in depth, UKAI convened two industry roundtables, bringing together AI companies, infrastructure operators, investors, energy specialists, and policy experts. These discussions were grounded in real commercial experience rather than abstract theory. Participants were asked not only what is broken, but what would practically unlock investment, accelerate deployment, and improve system performance. The insights from these conversations form the backbone of this report and its recommendations.

The objectives of this report are therefore fourfold. First, to identify what is required to make the AI industry in the UK successful in practice, not just in principle. Second, to clarify where the UK can credibly be a world leader, rather than attempting to compete on every dimension at once. Third, to set out policy recommendations that are informed directly by industry experience and investment realities. And finally, to offer a clear, coherent proposal for what needs to be done by government, regulators, and industry together to unlock sustainable growth.

The report is structured around four interlocking chapters. Chapter 1, Integrated Infrastructure, examines how AI depends on connected systems of energy, planning, digital connectivity, skills, and legitimacy, and why treating these elements in isolation has repeatedly slowed delivery. Chapter 2, Fairer Pricing, focuses on energy markets, arguing that high and volatile electricity prices are now one of the most binding constraints on AI infrastructure investment, and that targeted market reform is essential to restore competitiveness. Chapter 3, Targeted Innovation, explores how the UK can lead in making AI more energy efficient across hardware, software, and applications, by directing innovation towards system performance rather than raw scale. Chapter 4, Scalable Systems, addresses the question of how these elements are brought together, making the case for ecosystems that enable implementation at scale through standards, procurement, skills, and coordinated investment. The analysis in each of the chapters concludes with clear, practical, and realistic recommendations that cover both quick impact actions and structural priorities.

Taken together, these chapters advance a central argument: the UK should aim to be a world leader in Green AI, not simply as a branding exercise, but as a source of durable competitive advantage. Green AI, in this context, means AI that is efficient, affordable, deployable, and trusted; AI that strengthens the energy system rather than destabilising it; and AI that delivers real economic and societal value per unit of resource consumed. This is an area where the UK's unique combination of strengths – world-class research, a credible regulatory environment, deep infrastructure expertise, and strong public institutions – can be brought to bear.

Achieving this ambition will require the Government to ensure that its mission for AI is clearly articulated and comprehensively delivered, with consistent direction provided across policy areas that are too often treated separately. It will also require close collaboration between industry and academia, with UKAI playing a convening role: translating industry insights into policy proposals, helping align incentives, and supporting the development of shared solutions.

Delivering the change needed will be complex and challenging, but the prize for success is significant. If the UK can align infrastructure, pricing, innovation, and ecosystems around a coherent vision for Green AI, it can unlock growth in its AI sector and improve lives across the country while setting a global benchmark for how advanced technologies are deployed responsibly and cost effectively at scale. This report charts a clear path to achieving that vision.

Chapter 1 – Integrated Infrastructure

Building the connected foundations for sustainable AI

1. Introduction: Why AI Infrastructure Fails When Built in Silos

AI is often described as the next "general purpose technology".[1] But that description can be unhelpful if it encourages vague thinking. In practical terms, AI becomes useful when compute, data, and people come together to deliver measurable improvements in productivity, public services, scientific discovery, and decision making.

Compute is the capability to run AI models. That capability is delivered through data centres and networks, and increasingly through smaller systems closer to users. Compute has physical requirements: power, cooling, space, connectivity, and maintenance. It also has institutional requirements: planning consent, safety standards, workforce pipelines, finance, and trust.

When governments and investors talk about AI infrastructure, the focus tends to fall on the visible assets: new data centre sites, new grid connections, and new megawatts of power procurement. Those elements are necessary, but they are not sufficient.

1 https://www.oecd.org/en/publications/is-generative-ai-a-general-purpose-technology_704e2d12-en.html

In practice, the most common reasons for underperformance or delay are multifaceted systemic failures, as explored below

Energy delays: grid connection queues that last for years. Uncertain access to firm power during peaks. Price volatility that makes operating costs unpredictable.

Planning delays: local opposition triggered by distrust, opaque information, or a perception that benefits are captured elsewhere. Fragmented decision making across local authorities, regulators, and infrastructure owners.

Digital delays: fibre routes that are not resilient. Latency that makes services unusable. Inadequate backhaul that turns a data centre into an isolated island of compute.

Resource shortages: limited access to resources and inaccurate future forecast for demand, particularly of water, required for construction and running of data centres.

Skills shortages: a shortage of technicians, electricians, mechanical engineers, cooling specialists, and security staff. A lack of practitioners who can turn compute into working applications in health, energy, manufacturing, and public administration.

Demand failures: capacity built on speculative forecasts rather than anchored by real users. Subsidies that encourage low value use and discourage efficient design.

Ownership failures: infrastructure physically located in the UK but governed by commercial decisions made elsewhere, with limited value retained locally.

These delays and failures are closely interconnected. For example, if grid connection queues last for years, investors demand subsidies, councils face pressure, communities become sceptical, and businesses seek compute overseas. If planning consent is fragile, operators prioritise speed over quality, which increases public opposition. If skills are missing, projects become more expensive and delayed, and operational resilience suffers.

Therefore, if the UK wants to fulfil its potential to be a world leader in Green AI that is both competitive and sustainable, it must stop treating data centres, power, planning, skills, and adoption as separate policy areas. They must be designed and delivered together.

The rest of this chapter offers a detailed diagnosis of the problems that can undermine attempts to create a connected AI infrastructure and proposes solutions that remedy those problems holistically. Section 2 builds on the analysis started in this introduction by setting out all the components of a connected AI infrastructure and explaining why and how they must be designed as a full system. Section 3 sets out how planning to address system constraints collectively creates advantages and leads to the generation of assets that can be exported to other countries that face similar challenges. Section 4 provides a worked example of how a data centre can be planned and delivered as an integrated system. Section 5 focuses on green energy, affordability, and hydrogen as system components. Section 6 highlights issues related to financing, ownership, and value chapter.

This chapter also focuses on the Planning and Infrastructure Act and the Great British Energy Act, exploring the opportunities that new legislation is creating to improve the UK's AI infrastructure. Finally, the chapter concludes with policy recommendations that are divided into quick impact and longer term, structural priorities. Taken together, these recommendations show how the vision of creating an integrated AI infrastructure to help the UK achieve world-leading status in Green AI can be realised in practice.

2. Connected AI Infrastructure – Designing the Full System

Connected AI infrastructure is not a slogan or an abstract ambition. It should be placed at the centre of the Government's ambitions to turn the UK into a global AI leader. It starts from the premise that AI only creates value when a full system is in place, and that failure in any one component can undermine the entire investment. For the UK, this systems approach is the only viable way to build AI infrastructure under conditions of high energy prices, constrained grids, complex planning, and heightened public scrutiny.

At its core, connected AI infrastructure means that energy, planning, digital connectivity, water, skills, demand, and capital are treated as a single design problem rather than sequential hurdles. Instead of asking whether a site can host a data centre, the connected approach asks whether a place can sustainably host compute as part of a wider economic and social system. Any connected Green AI infrastructure must take account of the following components.

Energy Infrastructure as a Foundation

Energy is the most visible and most contentious dependency of AI infrastructure. Data centres require electricity that is both abundant and predictable. For operators, volatility is as damaging as scarcity. For policymakers, poorly planned demand can exacerbate peak pricing and undermine public confidence.

In a connected model, energy planning begins before any commitment to a specific site. It considers generation, storage, transmission, and flexibility as a combined system. Behind the meter renewable generation and storage are not treated as optional extras but as core components of risk management. Long-term power purchase agreements are used to stabilise prices and reduce exposure to wholesale market shocks. Where appropriate, alternative energy sources such as hydrogen are considered as tools for firm power, backup, and seasonal balancing.

This approach changes the political economy of AI infrastructure. Instead of seeking special energy discounts that create visible winners and losers, projects reduce their own exposure to the grid and contribute to local energy resilience. Energy becomes an enabler of competitiveness rather than a source of controversy.

Planning, Consent, and Legitimacy

Planning is often framed as a barrier to infrastructure delivery. In reality, it is a proxy for legitimacy. Infrastructure that lacks social licence will face delays, legal challenges, and political risk, regardless of its technical merits.

Connected delivery treats planning as a partnership process rather than a final hurdle. Local authorities are engaged early, with access to clear and verifiable information on energy use, water consumption, emissions, and community impact. Proposals are shaped to include visible local benefit, whether through skills investment, heat reuse, supply chain opportunities, or direct community funding.

Crucially, connected infrastructure requires national and regional spatial strategies so that local decisions are not taken in isolation. Without this context, councils are asked to weigh local impacts against national ambitions they did not help define, which is neither fair nor effective.

Recent moves to devolve further powers to mayors and combined authorities create a significant opportunity to strengthen this approach. With greater control over strategic planning, housing, transport, skills, and elements of energy and economic development, mayors are increasingly well placed to act as convenors of connected infrastructure delivery at a regional level. Where mayoral authorities can articulate clear spatial strategies for growth – identifying where energy, digital connectivity, skills, and demand can be aligned – planning decisions become less reactive and more anticipatory.

This shift allows infrastructure to be planned at the scale at which its impacts and benefits are actually felt. This reduces the burden on individual councils and improves public confidence that developments form part of a coherent regional vision rather than a series of isolated decisions.

Digital Connectivity as Critical Infrastructure

Compute without connectivity is stranded capacity. Yet fibre routes, latency requirements, and resilience are often only considered late in the development process.

A connected approach integrates digital connectivity planning alongside energy and land use. This includes securing multiple fibre routes to ensure resilience, understanding latency requirements for target workloads, and guaranteeing compatibility with the public sector and regulated networks where sensitive data is involved. Edge computing is considered where proximity to users materially improves performance or security, reducing the need for constant long distance data movement.

By treating connectivity as part of the core infrastructure stack, the UK can avoid building isolated islands of compute that struggle to attract users.

Water and Cooling as Design Inputs

Public debate around data centre water use is often shaped by overseas examples that do not reflect UK practice. Nonetheless, water remains a real constraint in specific locations – particularly the East and South East – and cumulative impacts matter.

Connected delivery requires cooling strategies to be chosen at the design stage, informed by local conditions. Closed loop and dry cooling systems are prioritised where feasible, with clear targets for water usage effectiveness. Transparent monitoring and reporting are built into operations, allowing concerns to be addressed with evidence rather than assertion.

By addressing water openly and early, projects reduce the risk of opposition driven by misunderstanding or mistrust.

Skills and Workforce as Enabling Infrastructure

AI infrastructure is built and operated by people. The shortage of skilled technicians, engineers, and practitioners is already a limiting factor in delivery.

In a connected model, skills planning is embedded from the start. Developers work with further education colleges, universities, and training providers to create clear pathways into construction, operations, and AI application roles. Apprenticeships, mid-career reskilling, and retention programmes are treated as investments in resilience rather than corporate social responsibility.

This approach also addresses a wider economic challenge. By creating visible, skilled jobs linked to AI infrastructure, communities are more likely to see such projects as opportunities rather than impositions.

Demand and Economic Value

Not all compute demand is equal. Some applications deliver significant productivity gains, public service improvements, or scientific advances. Others consume substantial amounts of energy to generate little economic or social value.

Connected delivery, therefore, requires a disciplined approach to demand. Capacity is anchored by real users with defined needs, rather than speculative forecasts. Pricing, procurement, and governance are designed to encourage high value use and discourage waste.

This focus on value is also an environmental strategy. When compute is scarce or expensive, efficiency matters. When it is cheap and subsidised, waste proliferates.

Capital, Ownership, and Sovereignty

Finally, connected infrastructure considers who owns and controls the assets on which the AI economy depends. Ownership structures affect where profits flow, how decisions are made, and how strategic risks are managed.

This is not an argument against international investment. It is an argument for clarity and balance. For sensitive public sector workloads, trusted governance arrangements are essential. For the wider economy, ensuring that some infrastructure is UK-anchored helps retain value and resilience.

Using the Planning and Infrastructure Act to Enable Connected AI Infrastructure

The Planning and Infrastructure Act[2] provides an important opportunity to reset how nationally significant digital and energy infrastructure is delivered in the UK. While the Act is not written specifically with AI in mind, its emphasis on faster decision making, clearer national direction, and stronger coordination across infrastructure types aligns closely with what AI infrastructure actually requires.

If applied deliberately, the Act could help move the UK away from fragmented, asset-by-asset approvals and towards the connected delivery model for Green AI set out in this report.

2 https://www.legislation.gov.uk/ukpga/2025/34/enacted

From Isolated Planning Decisions to System Delivery

As this report argues, one of the persistent weaknesses in the UK planning system has been the way decisions are made in isolation. The Planning and Infrastructure Act creates scope to rebalance this. By strengthening the role of nationally significant infrastructure planning and clarifying how national priorities should be weighed, it allows the Government to make the case that certain forms of AI infrastructure are not merely local developments, but components of a wider national system.

For AI infrastructure, the key shift is conceptual. A data centre should not be treated as a warehouse with an unusually large electricity demand. It should be treated as part of a combined digital and energy system that underpins economic productivity, public services, and national resilience. Where that framing is adopted, the Act provides tools to support more integrated assessment and faster resolution of trade-offs.

Giving National Weight to Digital and Compute Infrastructure

A central challenge for AI infrastructure has been the mismatch between national ambition and local burden. Local planning authorities are asked to approve developments that serve regional or national needs, while shouldering the political risk of energy use, visual impact, or perceived environmental harm.

The Planning and Infrastructure Act can help address this imbalance by reinforcing the role of national policy statements and strategic priorities in planning decisions. If AI infrastructure is clearly recognised within national infrastructure policy – alongside energy, transport, and water – it becomes easier for decision makers to weigh local impacts against national benefit in a transparent way.

This does not remove the need for good design or community engagement. On the contrary, it raises the bar. But it does mean that councils are not left to adjudicate national industrial strategy by default.

Accelerating Consent for Integrated Projects, Not Just Speed for its Own Sake

Speed is often presented as the primary benefit of planning reform. For AI infrastructure, speed matters, but integration matters more.

The Planning and Infrastructure Act offers a route to faster delivery precisely because it can support joined-up projects. Where a proposal brings together a data centre, behind the meter renewables, storage, grid reinforcement, and digital connectivity as a single, coherent scheme, the Act provides a framework in which those elements can be assessed together rather than sequentially.

This matters because many delays in AI infrastructure do not arise from objections to the data centre itself, but from unresolved dependencies. Grid connections lag behind building consent. Energy infrastructure is approved years after digital infrastructure. Fibre routes are negotiated separately, creating bottlenecks late in the process.

By encouraging or requiring developers to present integrated infrastructure propositions, the Act can help surface and resolve these dependencies earlier. The result is not simply faster approval, but fewer stalled projects and fewer unpleasant surprises.

Raising Standards While Reducing Uncertainty

A common fear is that faster planning means weaker scrutiny. For AI infrastructure, the opposite should be the goal: clearer standards applied more consistently.

The Planning and Infrastructure Act creates space for the Government to set clearer expectations on issues that currently generate controversy and delay, including energy sourcing, water use, resilience, and community benefit. Where those expectations are explicit, developers can design to them from the outset, and communities can see how decisions are being made.

In practice, this could mean

- Clear national guidance on acceptable cooling strategies and water reporting for data centres.
- Expectations around energy efficiency, use of low-carbon power, and contribution to system resilience.
- Stronger links between planning consent and skills commitments, apprenticeships, and local economic benefit.

The advantage of this approach is predictability. Developers know the rules. Local authorities have a framework to rely on. Communities see consistency rather than ad hoc decision making.

Enabling Regional Coordination and Cluster Development

AI infrastructure works best when it is clustered. Skills, supply chains, energy assets, and demand reinforce one another. Yet the UK planning system has historically struggled to support this kind of regional coordination.

The Planning and Infrastructure Act can help by reinforcing the role of strategic and regional planning, particularly where combined authorities or devolved governments are involved.

Where regions can articulate clear propositions – for example, AI and energy clusters aligned to offshore wind, hydrogen import, or major public sector datasets – the Act provides a mechanism for those propositions to carry real weight in planning decisions.

This is particularly important for moving beyond a small number of headline growth zones. Over time, the UK will need many sites of compute and edge infrastructure – not all of them nationally significant on their own, but collectively critical. Therefore, a planning framework that recognises and supports regional systems rather than isolated projects is essential. We explore the question of how to deliver these clusters in Chapter 4, Scalable Systems.

Aligning Planning Reform With Energy and Grid Reform

Perhaps the most important opportunity lies in alignment. Planning reform alone will not deliver AI infrastructure if grid connection reform, energy market reform, and skills policy remain disconnected.

The Planning and Infrastructure Act should therefore be seen as one component of a broader delivery agenda. Used well, it can provide the spatial and legal backbone that allows other reforms to land. Used in isolation, it risks accelerating building approvals without solving the underlying constraints.

For AI infrastructure, success will depend on the Government using the Act to support genuinely connected delivery: faster planning decisions that are matched by faster grid connections, clearer energy strategies, and credible skills pipelines.

The 2026 King's Speech: Where Next for Planning Reform in the Age of Artificial Intelligence

The next phase of planning reform must move beyond process acceleration and address how AI infrastructure is treated within the planning system as a matter of strategic national importance.

This will require the Government to make explicit choices about where AI infrastructure sits alongside energy, transport, and water within national planning policy, and to provide clearer guidance on how cumulative impacts, system benefits, and long-term demand should be assessed.

In practice, this means evolving from a model that reacts to individual applications towards one that anticipates future need, designates suitable locations in advance, and aligns planning decisions with grid investment, digital connectivity, and skills development. For AI, where lead times are long and dependencies are complex, planning reform must support earlier certainty rather than later acceleration.

Without this shift, the UK risks repeating a familiar pattern: approving infrastructure too slowly to meet future demand, then attempting to compress delivery once constraints have become acute. The opportunity now is to use planning reform to ensure that the infrastructure needed for the next decade of AI growth is planned deliberately, rather than discovered by necessity.

Implications for AI Infrastructure Policy

The Planning and Infrastructure Act offers the Government a chance to shift the debate about AI infrastructure away from narrow questions of speed and towards questions of system design.

If AI is treated explicitly as part of the UK's critical infrastructure, and if planning reform is used to support integrated energy, digital, and skills delivery, the Act could become a powerful enabler of sustainable AI growth.

If, however, it is simply used to push through isolated developments more quickly, it risks reinforcing exactly the problems that have slowed progress to date: local opposition, grid bottlenecks, and infrastructure that exists on paper but underdelivers in practice.

Used as a tool for connected delivery, the Planning and Infrastructure Act can help unlock the UK's green compute advantage. Used narrowly, it will fall short of its potential.

3. Designing for Efficiency in a Constrained System

The UK's structural constraints are often framed as disadvantages. In reality, they point towards a distinctive and potentially powerful strategy. The UK has an urgent need for energy efficiency, which drives innovation (which is the subject of Chapter 3 of this report) and requires greater collaboration (as discussed in Chapter 4).

High energy prices make waste visible. Grid constraints make poor planning costly. Contested land use makes legitimacy essential. Together, these factors push the UK towards a model of AI infrastructure that is efficient, modular, and purpose-driven.

Designing for efficiency means optimising for outcomes per units of energy, capital, and water. It means selecting model architectures and hardware that are appropriate to specific tasks rather than defaulting to the largest possible systems. It also means deploying compute closer to where it is used, reducing latency and data movement. And finally, it means building infrastructure that can adapt as technology evolves, rather than locking in a single generation of design.

This approach aligns with areas where the UK already has strengths: advanced research, system integration, regulatory capability, and sector-specific expertise. It also creates assets that are more exportable. Many countries face similar constraints. Few can afford hyperscale abundance. Therefore, efficient, well-governed AI systems will have global relevance and present near-term export opportunities.

4. A Worked Example – Delivering a Data Centre as a Connected System

To illustrate how these principles translate into practice, this section walks through the development of a mid-scale data centre designed primarily for inference and sector-specific AI services. The purpose is not to present a template, but to demonstrate a method of delivery that integrates the full infrastructure system.

The starting point is purpose. Rather than asking where a data centre could be built, the project begins by defining what the facility is for. In this case, the priority is supporting health services, energy systems, and public administration. These use cases rely heavily on inference, require high reliability and security, and benefit from proximity to users and data sources.

This definition immediately rules out certain options. Frontier model training at massive scale is not the objective. Instead, the focus is on steady, efficient operation and integration with public sector systems. That choice reduces energy intensity, narrows location requirements, and simplifies governance.

With purpose defined, the next step is to map constraints and opportunities as a single system. Grid capacity and connection timelines are assessed alongside opportunities for behind the meter generation and storage. Fibre routes and latency requirements are mapped alongside energy considerations. Water stress and cooling options are analysed in parallel with land use and planning constraints. Skills availability and training capacity are considered alongside long-term operational needs. Demand anchors, such as NHS organisations and combined authorities, are identified early to reduce speculative risk.

Only once this systems map is complete does site selection begin. The chosen location is not the cheapest parcel of land, but the place where energy, connectivity, planning, water, and skills align most effectively. The result is a site with lower overall risk, faster delivery, and higher long-term value.

Energy strategy is then developed as the first substantive workstream. Behind the meter renewables and long-term power purchase agreements provide price stability. Battery storage reduces peak grid draw and improves resilience. Hydrogen is considered as a source of backup and potentially firm low-carbon power, particularly where import and storage infrastructure is emerging. The aim is not to eliminate grid use, but to reduce exposure to volatility and constraint.

Cooling and water strategy is designed alongside the energy system. Closed loop or dry cooling is selected as the default, with clear performance targets and monitoring built in. This information forms part of the planning submission and public engagement, allowing concerns to be addressed transparently.

The facility itself is designed for efficiency and adaptability. Electrical and mechanical systems target best practice power usage effectiveness, with scope for continuous improvement. Hardware procurement prioritises performance per watt rather than peak specifications. The design allows for the integration of advanced accelerators, including photonic and neuromorphic systems, as they mature. Modular construction enables incremental expansion and reduces the risk of stranded assets.

Digital architecture is developed in parallel. Multiple fibre routes provide resilience. Edge nodes are deployed where low latency is critical. Federated learning is used where data cannot move, reducing central compute demand and strengthening data governance.

Planning and community engagement run throughout the process. Local authorities are engaged as partners. Skills commitments, apprenticeships, and local supply chain opportunities are built into proposals. Where feasible, waste heat is reused. The principle is reciprocity: the facility draws resources from the community and returns value.

Finally, success metrics that go beyond megawatts and floor space are defined. These include jobs created and sustained, technicians trained, energy cost stability, carbon intensity, water usage effectiveness, compute utilisation for priority workloads, and measurable improvements in public service outcomes.

5. Green Energy, Affordability, and Hydrogen as System Components

A sustainable AI industry depends on an energy system that is affordable, resilient, and politically durable. This requires moving beyond simplistic debates about whether green energy is cheap or expensive.

Energy affordability is a real competitiveness issue, as we discuss further in Chapter 2, Fairer Pricing. Volatility and peak exposure matter as much as average prices. At the same time, public tolerance for infrastructure that appears to raise household bills is low.

The connected infrastructure approach addresses this by changing the delivery model rather than relying on visible subsidies. Faster grid reinforcement, greater use of storage, behind the meter renewables, and demand flexibility all reduce system stress. These investments benefit multiple users, not just data centres.

Hydrogen plays a strategic role within this system. As a source of firm power and backup, it reduces reliance on gas during peak periods. As a storage medium, it supports seasonal balancing. In locations where hydrogen import and storage by ship is feasible, it can underpin clusters of energy-intensive activity, including compute, without overwhelming local grids. The value of hydrogen lies in integration. Used well, it enhances resilience and price stability. Used poorly, it becomes an expensive distraction. Policy should therefore focus on system design rather than headline capacity.

Using the Great British Energy Act to Underpin AI Infrastructure Delivery

The Great British Energy Act marked a significant shift in how the UK approaches energy system delivery. While the Government's flagship legislation is framed around clean power, energy security, and public value generally, it has important implications for the future of AI infrastructure specifically because it creates new institutional capacity to act across the energy system rather than leaving delivery entirely to fragmented market actors.

If deployed deliberately, Great British Energy (GBE) could become a critical enabler of connected AI infrastructure, particularly in a context of high energy prices, grid constraints, and public sensitivity to perceived preferential treatment for large energy users.

From Market Coordination to System Stewardship

As this chapter has shown, one of the defining challenges for AI infrastructure in the UK is not a lack of generation capacity in aggregate, but a lack of coordination. Renewable assets, grid reinforcement, storage, and demand are often developed on different timelines, by different actors, responding to different incentives.

The Great British Energy Act creates an opportunity to move beyond this fragmented approach. By establishing GBE, a publicly owned entity with a mandate to invest, partner and take a long-term view, the Act introduces the possibility of system stewardship rather than purely market coordination.

For AI infrastructure, this matters because data centres are not flexible, intermittent loads. They require firm, predictable power over extended periods. Private developers can contract for power, but they struggle to shape wider system investment in storage, flexibility, or grid reinforcement. GBE could help fill that gap.

Enabling Integrated Energy Solutions for Compute Clusters

AI infrastructure works best when energy solutions are designed around clusters rather than individual sites. Yet most energy investment decisions are still made on a project-by-project basis.

GBE could play a catalytic role in enabling integrated energy solutions for compute clusters, particularly in regions with strong renewable resources or emerging hydrogen infrastructure. This might include

- Co-investment in renewables and storage that serve multiple data centres and local users.
- Partnership with local authorities and combined authorities to align energy assets with economic development.
- Supporting behind the meter solutions that reduce grid stress while improving resilience.

Crucially, this does not require GBE to 'pick winners' in AI. Instead, GBE must invest in energy assets that support multiple strategic objectives: clean power, price stability, regional development, and industrial competitiveness.

Hydrogen and Long-term System Resilience

Hydrogen has often been discussed in abstract terms. The relevance of the Great British Energy Act is that it provides a mechanism to test hydrogen's role in real systems rather than theoretical models.

For AI infrastructure, hydrogen's value lies less in daily operation and more in system resilience. As a source of firm power and long-duration storage, hydrogen can help cover extended periods of low renewable output and reduce exposure to gas price spikes.

GBE could support this by

- Investing in hydrogen production and storage linked to ports and industrial clusters.
- Supporting pilot projects where hydrogen-backed generation provides resilience for energy-intensive infrastructure, including compute.
- Integrating hydrogen planning with grid reinforcement and storage strategies.

This kind of integration is difficult for individual data centre developers to achieve. It is precisely the kind of gap a publicly owned energy actor can fill.

Reducing Risk for Private Investment in AI Infrastructure

AI infrastructure is capital-intensive and long-lived. Investors care deeply about energy price stability, regulatory risk, and long-term system adequacy.

By taking a visible role in system investment, GBE can reduce perceived risk for private capital without directly subsidising individual firms. Where investors can see that storage, flexibility, and firm power are being built at scale, confidence in long-term operating conditions improves.

This is particularly important for UK-anchored infrastructure. Long-term confidence in the energy system makes it more viable to build and operate compute in the UK rather than defaulting to overseas locations with cheaper but less resilient power.

Aligning Public Value With Private Delivery

A central promise of the Great British Energy Act is that the public should see tangible benefit from energy investment. AI infrastructure risks falling foul of public perception if it is seen as consuming substantial amounts of power without clear social return.

GBE offers a mechanism to align energy investment with visible public value. For example

- Energy assets built to support compute clusters can also support housing, transport, and local industry.
- Returns from public investment can be recycled into grid upgrades, skills programmes, or bill reduction.
- Community benefit can be built into energy projects from the outset, rather than negotiated defensively later.

For AI infrastructure, this alignment is essential to maintaining political consent.

Limits and Risks

It is important to be clear about what the Great British Energy Act cannot do on its own. GBE is not a shortcut around grid constraints. It cannot replace the need for planning reform, skills development, or demand discipline. Nor should it be used to socialise risk while privatising reward.

If GBE is pushed into acting as a blunt instrument to lower prices for a narrow set of large users, it risks repeating the mistakes of past industrial policy. Its value lies in system level investment, not preferential treatment.

Implications for AI Infrastructure Strategy

Used well, the Great British Energy Act can provide the energy backbone for connected AI infrastructure delivery. It can help stabilise prices, improve resilience, enable integrated energy solutions and reduce investment risk, all without fuelling public backlash.

Used narrowly, it will have little impact on AI infrastructure beyond symbolism.

For the Government, the strategic opportunity is to treat GBE not as an energy-only institution, but as a partner in building the foundations of the digital economy. That requires explicit alignment between energy strategy, AI infrastructure planning, and regional development.

If that alignment is achieved, GBE could become one of the most important enablers of the UK's ambition to build an AI industry that is efficient, resilient, and publicly legitimate.

6. Financing, Ownership, and Value Capture

AI infrastructure is capital-intensive and long-lived. Therefore, financing models shape outcomes for decades.

Public support should be used selectively to de-risk early-stage projects with clear strategic value rather than subsidising inefficient operation. Guarantees or first loss capital can unlock private investment where pre-lets are difficult, provided they are tied to efficiency, transparency, and public benefit.

Ownership and governance matter, particularly for sensitive workloads. Trusted compute capacity for public services requires clear standards for security, audit, and control. This does not exclude international providers, but it does require deliberate procurement and governance. Ensuring that some infrastructure is UK-anchored helps retain value, skills, and resilience, even in an open investment environment.

Policy Recommendations

Quick Impact Priorities (0–24 months)

1 Treat AI Infrastructure as a Connected National System

AI infrastructure should be formally defined as a connected system, spanning energy, grid access, planning, connectivity, water, skills, and demand, rather than as isolated data centre assets.

To deliver this shift, the Government should adopt a shared, systems-based framing of AI infrastructure across strategy, planning, and investment decisions. Embedding this perspective in national infrastructure and AI strategies would help ensure that compute demand is considered alongside energy capacity, connectivity, water use, skills availability, and wider system constraints, providing a clearer and more coherent policy narrative for delivery bodies and investors.

This approach should also be reinforced through appraisal, assurance, and advisory processes, ensuring that decisions about AI-related infrastructure reflect system-wide impacts, dependencies, and trade-offs rather than narrow, asset-level considerations. Independent analysis and long-term system thinking will be important to maintaining consistency, credibility, and resilience as AI infrastructure scales over time.

2 Establish a Cross-government AI Infrastructure Delivery Unit

Delivering AI infrastructure at pace will require stronger coordination across government than is currently in place. A cross-government AI Infrastructure Delivery Unit would provide a focal point for aligning policy, planning, energy, and digital infrastructure decisions under clear ministerial direction. Positioned at the centre of government, such a unit could help translate strategic ambition into coordinated delivery and reduce the risk of bottlenecks emerging between departments.

The role of the unit would be to bring together relevant capabilities across government – including AI and digital policy, energy system planning, and the planning and infrastructure regime – to align priorities, timelines, and investment decisions. By operating across major delivery programmes and working closely with the Department for Science, Innovation and Technology (DSIT), the Department for Energy Security and Net Zero (DESNZ), and the Ministry of Housing, Communities and Local Government (MHCLG), the unit would provide a single mechanism for resolving issues and accelerating progress.

Crucially, this approach is about coordination rather than creating new policy. A delivery unit would help ensure that decisions are joined up, delivery risks are surfaced early, and AI infrastructure ambitions are translated into practical, investable outcomes.

3 Give National Strategic Weight to AI and Compute Infrastructure

AI and digital compute infrastructure should be explicitly recognised as strategically significant national infrastructure, providing clear national policy context for local planning decisions.

To deliver this recommendation, MHCLG could update the National Planning Policy Framework and relevant National Policy Statements to explicitly recognise AI and digital compute infrastructure as strategically significant national infrastructure, providing clear policy weight for decision making at local and regional level.

This should be underpinned by clear ministerial statements of strategic importance, signalling to local authorities, planning inspectors, and investors that AI and compute infrastructure is a national priority and should be assessed accordingly within the planning system. Combine this with reform of grid connection queues for strategically important compute.

4 Require Early Energy Integration in AI Infrastructure Proposals

AI infrastructure proposals should be expected to demonstrate integrated energy strategies from the outset, including grid capacity, storage, flexibility, and long-term power arrangements.

To implement this recommendation, DESNZ should lead the integration of energy considerations into AI infrastructure development by embedding expectations on grid capacity, storage, flexibility, and long-term power arrangements within its system planning frameworks.

The National Energy System Operator (NESO) and the Office of Gas and Electricity Markets (Ofgem) should align connection, capacity, and network guidance to require early engagement and evidence of deliverable energy solutions for major AI infrastructure proposals, reducing late-stage risk and system constraints. Together, these levers should ensure energy readiness is treated as a core enabling condition for AI infrastructure delivery rather than a downstream consideration and create a permitting fast-track for genuinely integrated schemes.

5 Introduce Standard Transparency Metrics for AI Infrastructure

Standard reporting on power usage effectiveness (PUE), water usage effectiveness (WUE), and carbon intensity should be required as a condition of public support. As an industry body, UKAI will help set up, maintain, and promote regular reporting. This data will help dispel many of the myths around the energy and water consumption associated with AI infrastructure, thus building greater consumer trust through transparency and accountability.

To support this, the Government and industry should work together to establish a small number of clear, consistent transparency metrics for AI infrastructure. Common reporting standards would improve comparability, support informed decision making, and create a shared evidence base for policy, investment, and public engagement.

Embedding these metrics across public funding, procurement, and planning processes would help normalise transparency as part of AI infrastructure delivery, while remaining proportionate and practical. UKAI can play a supporting role by coordinating industry input, maintaining reporting frameworks, and promoting disclosure as a foundation for trust in Green AI. This transparency is a first, foundational step in building performance thresholds that can then be used to incentivise greater efficiency.

Structural Priorities (2–10 years)

6 Anchor AI Infrastructure Growth to Evidence-based Demand

The Government should commission independent modelling of AI demand and economic value to distinguish high-impact use from low-value volume, guiding infrastructure scale and incentives and preventing overbuild.

Robust demand and value modelling would help ensure that public support and planning assumptions are grounded in realistic assessments of economic impact rather than headline capacity projections alone. The Government can draw on independent, forward-looking evidence to inform decisions about the scale, sequencing, and location of AI infrastructure.

Using this evidence consistently across policy, investment, and public sector adoption would improve coherence and predictability, while reducing the risk of misaligned infrastructure investment. Independent challenge and system-level analysis will be important in testing assumptions, exploring scenarios, and maintaining a balanced, resilient approach as AI demand continues to evolve.

Such evidence would allow the government to establish a public sector demand-anchored and aggregated procurement programme that builds on existing commissioned modelling and translates it into active market shaping. By creating stable, predictable demand, this approach would help de-risk regional investment. The programme could include aggregated procurement across the NHS, local government, and the research sector; anchor-tenancy models to support the development of regional clusters; and clear governance frameworks for handling sensitive workloads.

7 Align Planning Reform With Grid, Energy, and Skills Constraints

Planning reform should be explicitly aligned with grid connection reform, energy system planning, and programmes to increase the availability of infrastructure-critical skills.

The Government should lead a joined-up approach to planning, energy, and skills policy, ensuring that faster planning processes are matched by realistic delivery capacity across the wider system. Aligning planning decisions with grid availability, energy timelines, and workforce readiness will be essential to avoid creating new bottlenecks or delivery risks.

A coordinated, place-based perspective will also be important, enabling national reform to translate into outcomes that are achievable on the ground. By aligning planning, energy, and skills considerations across national and regional levels, the Government can improve delivery certainty, reduce delays, and support the timely growth of AI infrastructure in a way that is resilient and sustainable. See Recommendation 11 (below) for further details on how the availability of the required skills can be increased.

8 Shift Energy Policy Toward Shared System Upgrades

Energy policy should prioritise investment in shared system upgrades – such as grid reinforcement, storage, flexibility, and behind the meter renewables – before blanket price interventions.

Focusing on shared upgrades helps unlock capacity for AI and other growth sectors while delivering wider system benefits. To support this shift, greater emphasis should be placed on energy system investments that increase overall capacity, resilience, and flexibility, rather than relying primarily on price-based incentives to manage demand.

Aligning regulatory, investment, and delivery frameworks around these shared assets can reduce long-term costs, spread risk more fairly and strengthen energy security. By prioritising system-wide improvements, energy policy can better support the sustainable growth of AI infrastructure while accelerating progress toward Net Zero objectives.

9 Treat Digital Connectivity as Critical AI Infrastructure

Resilient, low-latency digital connectivity should be treated as critical infrastructure for AI, embedded early in planning and delivery decisions.

To deliver this, the Government should ensure that digital connectivity requirements for AI are reflected consistently across digital, infrastructure, and planning frameworks. Treating connectivity as a core enabling condition, rather than a secondary consideration, will help improve resilience, performance, and investment certainty as AI infrastructure scales.

A coordinated approach across regulation, planning, and place-based delivery will be important to ensure that connectivity investment aligns with energy, infrastructure, and skills decisions. By embedding connectivity considerations early, the Government can reduce deployment risk, support system resilience, and enable AI infrastructure to operate effectively as part of a connected national system.

10 Embed Community Benefit and Local Legitimacy Into Infrastructure Design

AI infrastructure projects should be required to demonstrate tangible community benefit and early local engagement as part of the consent process.

Early and meaningful engagement can help address concerns, build trust, and ensure that projects respond to local priorities rather than imposing solutions from the outside. To enable this, the Government should encourage a consistent expectation that AI infrastructure delivers visible local value alongside national benefit. Taking a place-based approach to community benefit (linking infrastructure delivery to jobs, skills, local investment, and wider public value) can help strengthen legitimacy and reduce opposition over time.

Additional planning requirements could be created that mandate heat re-use feasibility assessments, cumulative impact assessment standards for compute clusters, or other mechanisms to quantify and track local-value. This provides local communities and planners with a tangible way to ensure ongoing value delivery, rather than a one-off, initial approval. By embedding community considerations into planning and delivery from the outset, AI infrastructure can become a positive contributor to local economic and social outcomes, supporting long-term acceptance and sustainable growth.

11 Treat Infrastructure Skills as a Delivery–critical Asset

Skills required to build, operate and maintain AI infrastructure should be treated as core infrastructure, supported through targeted apprenticeships, reskilling, and retention programmes.

Greater coordination between skills policy, infrastructure planning, and energy system needs will be essential to avoid workforce constraints becoming a barrier to delivery.

The Government should take a more strategic view of infrastructure-critical skills, ensuring that education and training pathways are aligned with the scale, timing, and technical demands of AI infrastructure deployment.

A place-based approach can help translate national ambition into local capability, linking employers, training providers, and infrastructure programmes around shared demand. By recognising skills as a foundational asset rather than a supporting input, the Government can improve delivery certainty, strengthen local labour markets, and support the long-term resilience of AI and energy ecosystems (a topic further explored in Chapter 4).

Strategic Implications: Intelligence Through Integration

The UK does not need to build the world's largest AI infrastructure. It needs to build the most intelligent one. Intelligence, in this context, is not about model size or raw compute. It is about system design and integration. It is about recognising that AI only creates value when energy, planning, water, digital connectivity, skills, finance, and demand are aligned as a single delivery system.

This chapter has shown that where that alignment is missing, infrastructure stalls, costs rise, public trust erodes, and capacity sits underused. Where it is achieved, constraints become drivers of innovation rather than brakes on growth. Many of the barriers facing AI infrastructure in the UK (grid delays, planning conflicts, skills shortages, and volatile operating costs) are not failures of ambition, but failures of integration. Treating these dependencies separately leads to brittle outcomes; treating them as a system creates resilience.

A central implication is that efficiency must be prioritised as a strategic advantage, not a secondary consideration. In a context of rising compute demand, constrained grids, and heightened public scrutiny, the ability to deliver useful AI outcomes with less electricity, less water, less capital, and lower systemic risk will increasingly define competitiveness. This is an economic strategy shaped by real-world constraints and an approach necessary to achieve environmental goals and long-term sustainability.

When green energy is integrated properly – through long-term power arrangements, storage, flexibility, and system planning – it reduces volatility and long-term risk rather than simply shifting costs. Likewise, transparency around water use, energy sourcing, and community impact is not a regulatory burden but a condition of legitimacy. Infrastructure that is clearly planned, transparently governed, and visibly beneficial is far more likely to secure consent than projects perceived as imposed or extractive. Skills pathways, community benefit, energy resilience, and local economic value are not peripheral concerns; they are the foundations of durable delivery.

For the Government, the implications go beyond individual policy levers. AI infrastructure cannot be delivered through isolated announcements, siloed departments, or single-issue incentives. It requires sustained coordination across energy, planning, digital, skills, and finance, backed by institutions capable of stewarding systems rather than merely approving assets. Recent legislative tools, including the Planning and Infrastructure Act and the creation of GBE, provide important opportunities, but only if they are used to support integration rather than to increase speed alone.

This chapter establishes a baseline proposition: the UK's structural constraints point towards a distinctive model of AI development. One that prioritises efficiency, purpose-driven deployment, and system design over scale for its own sake. This is the foundation of what this report characterises as a Green AI approach, not defined narrowly by emissions, but by the disciplined use of energy, capital, and resources to maximise long-term economic and social value.

The next chapters build on this foundation. Chapter 2 examines the challenges around energy pricing, and Chapter 3 explores how these constraints can become a source of innovation rather than limitation, shaping new approaches to AI efficiency, deployment, and competitiveness. Together, they argue that the UK's opportunity is not to compete on volume, but to lead in building AI that is efficient, resilient, and legitimate in a resource-constrained world.

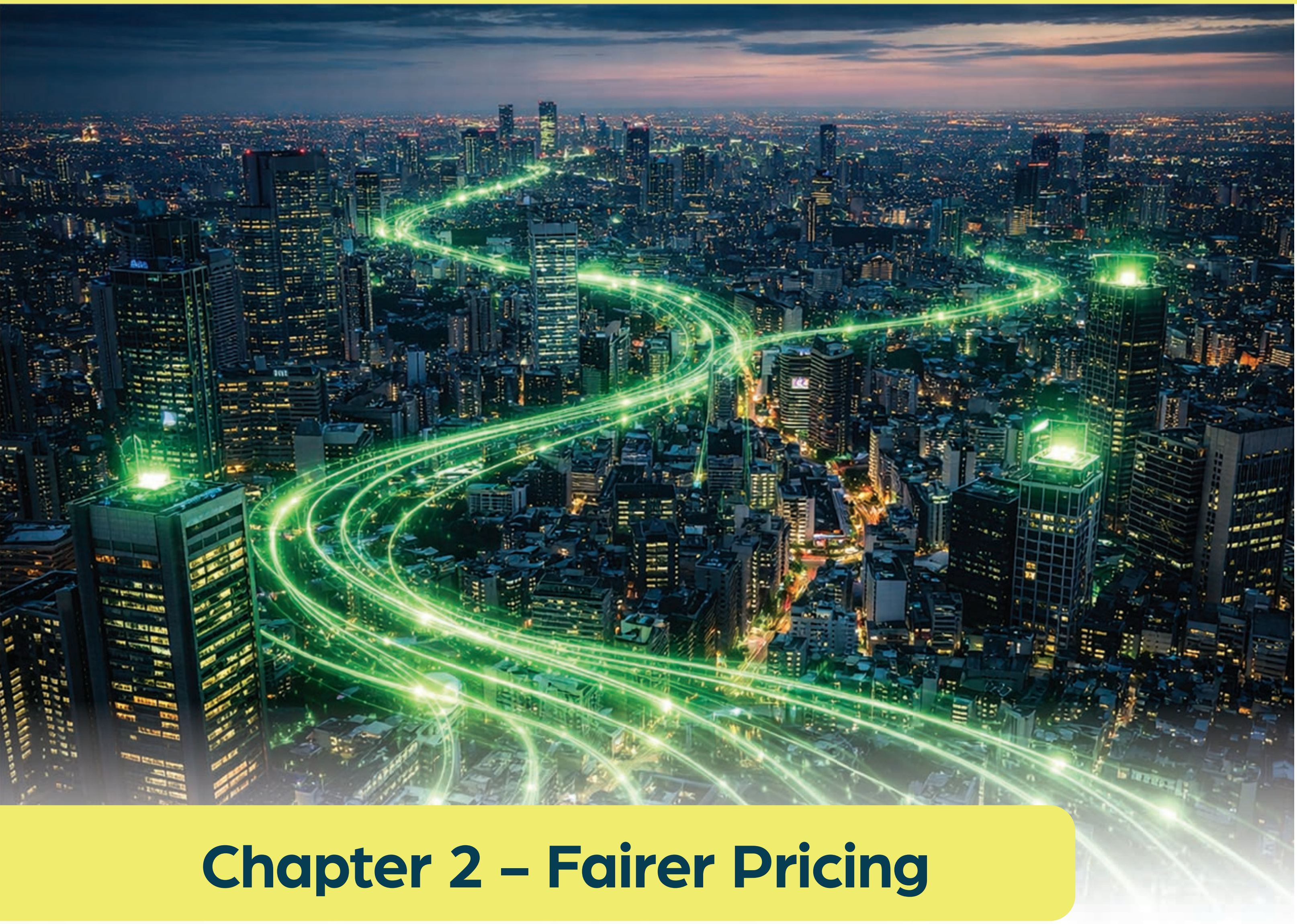
Industry Perspective

"Ark Data Centres welcome the report from UKAI that sheds valuable insight into the many issues arising from the rapid growth of AI, the impact that it is likely to have and the importance of moving forward on a well-informed and realistic basis.

The report importantly recognises the complex infrastructure that underpins digital infrastructure and delivery, and the intertwined relationships of power, network availability, data centre capacity and location. Successful AI deployment will also need to recognise the needs of those organisations, large and small, who will be either developing or supporting the widespread use of AI applications.

For the UK to fully realise the AI ambitions that the Government supports, it is crucial to understand the importance of energy pricing, energy availability and the necessary improvements that will need to be made to power allocation. The UK is currently disadvantaged in energy pricing, suggesting that it will be harder if not impossible to be fully competitive."

Allan Bosley, Director of Public Affairs
Ark Data Centres



Chapter 2 – Fairer Pricing

Reforming energy markets to reward efficiency, flexibility, and long-term value

1. Introduction: Energy Pricing is the Limiting Factor for the Growth of the UK's AI Sector

The UK has the research capability, talent, and demand required to be a global leader in AI. But its ability to achieve that ambition is increasingly constrained by high and volatile energy prices.

Investment in data centres, which support training and large-scale inference, is of critical importance to the AI industry. However, the cost and predictability of electricity are decisive factors in any data centre investment decisions. And although the UK is increasing the share of low-carbon generation in its energy mix, the nation's electricity prices remain tightly linked to global gas markets and short-term trading volatility. This misalignment discourages capital-intensive AI infrastructure investment and risks pushing strategically important compute capacity to jurisdictions with more stable and competitive energy pricing.

Addressing this challenge requires targeted reform of energy market design rather than reliance on long-term generation build-out alone. This is of critical importance to the AI industry, which relies upon existing and future investment in data centres. The Government has rightly identified these issues, and has published the UK Compute Roadmap (July 2025) and formed the AI Energy Council, both important steps to identify solutions. Ambitions for the UK to lead in AI require practical and fast solutions, however, few have yet been implemented. In UKAI's analysis of the AI Opportunities Action Plan in July 2025 our members gave the Government a score of 2 out of 5 for delivery of sustainable AI infrastructure (Recommendation 5 of the plan).

Following this introduction, the rest of this chapter explores the challenges around energy pricing in eight further sections. Section 2 sets out the current solutions for electricity market reform. Section 3 details structural approaches to reducing energy prices for strategic UK industries. Section 4 reviews the international evidence that shows how other jurisdictions have successfully reduced energy prices for data centres. Section 5 shares the perspectives of the UK AI industry and its investors on energy prices for AI infrastructure. Section 6 examines the implications of the Government's Modern Industrial Strategy for AI data centres. Section 7 considers how the UK's Energy Intensive Industries (EII) framework could be extended to incorporate data centres. Section 8 summarises the additional available mechanisms for managing energy pricing. And Section 9 makes the case for applying the UK's expertise in carbon trading to data centres.

The chapter then concludes with recommendations that can address energy challenges in the short and long terms to pioneer the development of Green AI in the UK. Before turning to all that, however, it is necessary to first detail the limitations that energy pricing issues place on investment in AI infrastructure.

How Energy Pricing is Limiting AI Infrastructure Investment

AI data centres are fundamentally different from most industrial energy users. They require large volumes of electricity on a continuous basis, operate on investment horizons of 10 to 20 years, and are highly sensitive to both price level and price volatility. In the UK, investors face electricity prices that are materially higher (3x) than those in competitor jurisdictions such as Norway and parts of the United States, combined with uncertainty about how those prices will evolve over time.[3] Even where renewable generation is abundant, prices remain exposed to gas market shocks and short-term trading dynamics. This creates a significant deterrent to investment, particularly for training workloads that require long, uninterrupted runs and predictable operating costs. As a result, companies increasingly locate large-scale training infrastructure overseas, while reserving UK sites for smaller inference or edge deployments. This dynamic risks undermining the UK's broader AI strategy by weakening the domestic compute base on which innovation, security, and economic value creation depend.

3 https://www.gov.uk/government/statistical-data-sets/international-industrial-energy-prices

How Market Design Drives High Electricity Prices

Electricity prices remain anchored to gas, not system costs

Wholesale electricity prices in the UK are largely set by the marginal cost of gas-fired generation. Under current market arrangements, the price paid for electricity in each settlement period is determined by the most expensive plant required to meet demand. In practice, this is frequently gas-fired power, even when gas provides a minority share of total generation.

As a result, electricity generated by renewables and nuclear – which have low operating costs and are largely insulated from fuel price volatility – is priced at levels that reflect gas market conditions rather than underlying system costs. This means that domestic electricity prices continue to track global gas price movements, despite the UK's electricity mix becoming progressively less dependent on gas.

The issue is not marginal pricing in itself, but the persistence of gas as the price-setting technology in a system increasingly dominated by low-marginal-cost generation.

Wholesale market exposure amplifies volatility and risk premiums

The structure of wholesale electricity markets amplifies this effect. Exposure to short-term price volatility, relatively thin forward market liquidity, and imbalance risk lead suppliers to embed significant risk premiums into prices. These premiums often persist even when physical supply conditions improve or generation costs fall.

While Contracts for Difference (CfD) have insulated newer low-carbon generation from wholesale price swings, a substantial share of electricity supply remains exposed to market volatility. This leaves consumers facing fluctuating prices, while low-cost generators can earn elevated revenues during periods of high gas prices rather than locking in long-term affordability at system level.

The result is a pricing environment shaped as much by risk management and market exposure as by the actual cost of producing electricity.

System inefficiencies are socialised through consumer bills

Grid constraints and network congestion further distort electricity prices. Where the transmission system cannot move power efficiently from generation-rich regions to demand centres, generators are paid to curtail output while higher-cost plant is run elsewhere to maintain system balance. These constraint and curtailment costs are recovered through network charges and passed on to consumers.

This masks the availability of abundant low-cost renewable power and weakens incentives for energy-intensive users to locate close to supply. It also reflects a system in which grid investment and market design have not kept pace with the rapid growth of renewable generation.

Retail market design compounds these effects. Price cap mechanisms and supplier risk management practices tend to pass through cost increases more rapidly than cost reductions, embedding additional volatility and uncertainty for households and large energy users alike.

Speed is of the Essence

A major challenge with energy and infrastructure is the time required to deliver results, but the AI industry requires urgent action; otherwise, the UK will be left behind and miss its opportunity to lead the world in the development of Green AI. This report recommends a number of solutions, prioritising those that can be delivered quickly. Many of the measures outlined here can be implemented within 18 months, as they rely primarily on regulatory and contractual change rather than new physical infrastructure. Expanding fixed-price contracting, reducing exposure to gas-linked pricing, reforming pass-through mechanisms, and piloting strategic energy arrangements for AI infrastructure could all deliver meaningful price reductions in the near term.

However, longer-term reforms will still be required to sustain these improvements. These include deeper electricity market reform, grid reinforcement, smarter management of constraints, and continued investment in firm low-carbon capacity such as storage and nuclear. Crucially, these structural measures should build on, rather than delay, near-term action.

The Role of Great British Energy

As discussed in the previous chapter, the Great British Energy Act 2025 established Great British Energy (GBE) as a new, publicly owned energy company with a mandate to accelerate clean power deployment and support the UK's transition to a more secure, affordable, and sustainable energy system. While the Act does not directly deal with the reform of wholesale electricity pricing or retail tariffs, it introduces an important new institutional actor that could influence energy pricing outcomes indirectly through investment strategy, market participation, and system design.

GBE cannot change pricing rules on its own, but it can help address some of the structural drivers of high and volatile prices by increasing the supply of low-cost generation, supporting long-term power contracts, and reducing exposure to short-term market fluctuations. Ofgem and the Government remain responsible for decisions on wholesale market design, locational pricing, and consumer protection. But, as we will explore, GBE could be an important vehicle to champion the reforms or convene key stakeholders.

As a publicly owned investor and operator, GBE has the potential to act counter-cyclically to private markets, prioritising long-term system value over short-term returns. By investing in generation assets with long lifetimes and stable operating costs, GBE could help expand the share of electricity sold under long-term arrangements rather than spot pricing, supporting greater price stability. Over time, this could reduce system-wide reliance on volatile wholesale markets and create clearer price signals for industry, data centres, and other energy-intensive users.

If aligned carefully with broader energy-market reforms, GBE could become an important lever in translating clean power into genuinely lower and more stable energy prices over time, which would significantly increase the UK's chance of becoming a world leader in Green AI technology.

2. Current Solutions for Electricity Market Reform

It is clear that the current electricity market needs reform to disentangle the accretion of adjustments that have been made over decades. However, this does not require the dismantling of the existing market overnight. It does require targeted interventions to better align prices with system costs. One of the most impactful measures would be to reduce the degree to which electricity prices are determined by gas.

Reducing the impact of gas prices on electricity

At the moment, Contracts for Difference (CfD) is a government-backed mechanism used to support low-carbon electricity generation by providing price certainty. Under a CfD, a generator is guaranteed a fixed 'strike price' for the electricity it produces over a long period, typically 15 years. If the market price of electricity falls below the strike price, the generator receives a top-up payment; if the market price rises above it, the generator pays back the difference. This stabilises revenues for generators, lowers the cost of financing new projects and protects consumers from excessive prices, while allowing electricity to continue to be traded through the market.

CfDs are already central to the UK's approach to supporting new low-carbon generation, providing long-term price certainty for investors while protecting consumers when wholesale prices are high. Current plans envisage expanding and evolving this model to cover a greater share of low-carbon generation over time, including potentially bringing some existing assets into revised contractual arrangements as they exit legacy support schemes. Alongside this, the Government has been exploring hybrid pricing models that better reflect the fundamentally different cost structures of low-carbon and gas-fired generation.

Options under consideration include separating low-carbon power from gas-based price setting, extending CfD-style stabilisation mechanisms, or piloting alternative market arrangements such as dual pricing or pay-as-bid models for renewables and nuclear. The common objective is to reduce the automatic transmission of gas price volatility into electricity prices, while continuing to provide credible, long-term investment signals for clean generation. Taken together, these reforms would help lock in the benefits of low-cost renewables for consumers without undermining future investment.

The Government can signal long-term direction and reassure investors

In parallel, the Government can act quickly to increase the availability of long-term, fixed-price electricity for strategic infrastructure such as AI data centres. This could include enabling government-backed offtake agreements, aggregating demand through public procurement mechanisms, or designating AI infrastructure as strategically eligible for long-term power contracts. By providing price certainty, these measures would materially reduce investment risk without distorting the wider market. We will explore these ideas further below.

Market reform to reduce volatility

Reducing volatility and trading-related risk premiums is another area where policy can have near-term impact. Improving transparency and liquidity in forward markets, encouraging longer-term contracting, and reviewing the role of speculative behaviour during periods of low physical scarcity would all help bring prices closer to underlying system costs. At the retail and industrial level, reforms to price-pass-through mechanisms could ensure that falls in wholesale prices are reflected more quickly and predictably in end-user bills.

Over the medium term, reform of constraint and congestion pricing will be necessary to sustain lower prices. Better locational signals, combined with protections for consumers, would allow energy-intensive infrastructure, such as that required by AI, to access cheap renewable power more directly and reduce the overall cost burden on the system.

3. Structural Approaches to Reducing Energy Prices for Strategic UK Industries

Incremental reform of the electricity market will help reduce energy prices over time, but it is increasingly clear that marginal adjustments alone may not be sufficient to unlock investment in strategically important, energy-intensive infrastructure such as AI data centres. If the UK wishes to remain competitive with jurisdictions that offer materially lower and more predictable power costs, government and financial markets may need to adopt more radical, structural approaches that reshape how risk, pricing, and strategic priority are allocated within the energy system.

Reframing Electricity as Strategic Infrastructure, Not a Commodity

A foundational shift would be to explicitly recognise electricity supply for certain strategic uses as a form of national infrastructure, rather than a purely traded commodity. The recent designation of data centres as Critical National Infrastructure (CNI) establishes a precedent for recognising their national importance,[4] but this designation currently focuses on resilience, security, and continuity rather than cost or access. A more radical extension of this logic would be to treat energy supply to CNI-designated assets as part of the infrastructure itself, justifying differentiated pricing, priority access, or tailored contracting arrangements on the grounds of national capability and economic security.

Such an approach would not require full renationalisation of the electricity system. However, it would require a clear policy decision that price formation for strategic infrastructure should reflect long-run system costs rather than short-term marginal scarcity driven by global gas markets.

---

4 https://www.gov.uk/government/news/data-centres-to-be-given-massive-boost-and-protections-from-cyber-criminals-and-it-blackouts

Creating a Split Electricity Market for Low-carbon Baseload and Marginal Generation

One of the most significant structural reforms would be to separate the electricity market into two functional components: a low-carbon baseload market and a marginal balancing market. Under this model, electricity generated by nuclear and contracted renewables would be pooled and sold at an averaged, regulated, or contract-based price reflecting their underlying costs, while gas and other peaking assets would continue to operate in a marginal market designed to manage variability and scarcity.

This approach builds on existing mechanisms such as CfD, which already remove price volatility for new low-carbon generation, but extends the principle more broadly across the system. Rather than allowing gas to set the price for all electricity, this model would limit gas's influence to the portion of the system where it is genuinely required. For energy-intensive users, such as AI data centres, this would create access to electricity priced in line with the cost of low-carbon generation, without undermining incentives for flexibility or investment in balancing capacity.

Establishing a State-backed Strategic Power Buyer

Another radical but credible option would be the creation of a state-backed or regulated entity that acts as a long-term purchaser of electricity on behalf of strategically important users. This entity would aggregate demand from sectors such as AI infrastructure, advanced manufacturing, and critical digital services, and enter into long-term power purchase agreements with low-carbon generators. Electricity would then be sold on to participating users at stable, predictable prices.

This model builds on existing precedents in the UK energy system, including the Government's role in underwriting CfDs and the use of regulated asset models in other infrastructure sectors, such as the Regulated Asset Base (RAB) model for water, currently regulated by Ofwat. It would not require the Government to operate power stations, but it would allow the state to intermediate risk, using its balance sheet and credibility to reduce the cost of capital and strip volatility out of prices faced by industry. For AI data centres, the value of such an arrangement lies less in achieving the absolute lowest price and more in securing long-term certainty that aligns with investment horizons.

Introduce a Strategic Industrial Power Tariff

The UK already differentiates between consumers in many parts of the energy system, including through network charging, climate levies, and exemptions. A more explicit strategic industrial power tariff could be introduced for assets that meet defined national criteria, such as contributing to AI capability, economic resilience, or public-sector digital capacity. Eligibility could be linked to existing designations, including CNI status, Nationally Significant Infrastructure Project (NSIP) approval, or participation in designated AI Growth Zones.

Such a tariff would not necessarily require direct subsidies. It could instead be delivered through discounted network charges, access to long-term fixed-price power sourced from low-carbon generation, or exemptions from certain policy costs currently levied on electricity bills. By framing this as an industrial policy tool rather than a consumer subsidy, the Government could align it with existing powers while maintaining fiscal discipline. This Strategic Industrial Tariff could be developed as an extension of the Government's current AI Growth Zone programme.

Using Spatial Pricing to Capture Local Energy Abundance

The Government should give serious consideration to embracing zonal pricing, paired with a deliberate place-based industrial strategy. Under such a model, areas with persistent renewable oversupply and grid constraints, such as parts of North West Scotland, would see lower local electricity prices that reflect actual system conditions. The Government could then actively steer AI data centre development, particularly for training, to these locations through planning fast-tracks, NSIP designation, grid prioritisation, and targeted tax relief. The Scottish Government has already had some success in creating 'Green Freeports', an extension of the UK-wide freeport zones, leveraging tax relief to incentivise investment in sustainable industries. The first two Green Freeports[5] were announced in 2023 in Inverness and Cromarty Firth and in the Firth of Forth.

This approach would turn a current system weakness – renewable curtailment and congestion – into an advantage, allowing energy-intensive infrastructure to monetise surplus generation directly. However, to attract data centres, these zones must provide a diversified energy mix capable of delivering firm, high-availability power with four-nines reliability to support continuous operation. This solution could also be politically sensitive at a time when consumers are very concerned about domestic energy prices, particularly if it leads to visible regional price differences. However, the zonal pricing model offers one of the clearest pathways to structurally lower prices without ongoing subsidy.

Mandating Long–term Contracting to Reduce Exposure to Spot Markets

Electricity prices remain high in part because too much demand is exposed to short-term markets and the risk premiums embedded within them. A more interventionist approach would be to require suppliers and large consumers to source a higher proportion of their electricity through long-term contracts. This would reduce the influence of spot prices on overall system costs and dampen volatility.

For AI data centres, which already favour long-term certainty, such a shift would align market design with investment reality. While this would represent a departure from the UK's historically liberalised market philosophy, it would be consistent with the treatment of other forms of long-lived infrastructure, particularly during times of national upheaval. It would also reflect the critical importance of data centres, and the sovereign AI technology they deliver, to the future of the UK and enhance the likelihood of achieving world-leading status in the development of Green AI.

5 Footnote: https://www.gov.scot/policies/cities-regions/green-ports/> Where

Long-term Contracting to Absorb Volatility

Financial markets could be used more deliberately to transform volatile wholesale prices into stable industrial inputs. One option would be the creation of a government-backed or regulated power hedging instrument that offers qualifying users a fixed electricity price in exchange for a premium. This would operate similarly to insurance, with the state's role focused on backstopping tail risks rather than setting prices.

One of the clearest messages from operators is that the UK needs to make long-term power contracting simpler, more standardised, and less risky. While corporate power purchase agreements (PPAs) exist in the UK, they are often bespoke, legally complex, and exposed to 'basis risk' (the gap between the price you thought you had fixed and the price you actually pay) arising from grid constraints and balancing arrangements. This contrasts with markets such as the Netherlands, where standardised PPA structures and deep liquidity have made long-term contracting the default rather than the exception.

Investors see long-term contracting not as a distortion of the market, but as a necessary adaptation for capital-intensive digital infrastructure. They consistently argue that if the UK wishes to attract training-scale AI infrastructure, it must normalise access to long-term, fixed-price electricity, whether through private PPAs, state-backed aggregation, or regulated offtake frameworks (where the Government acts as a credible counterparty to reduce risk for investors). It does not matter who carries the risk, but that the risk is structured, mitigated, and shared in a form that lenders and investors can finance.

Such instruments could be aligned with existing public revenue streams from CfDs or capacity mechanisms, effectively recycling periods of high market prices into protection against future volatility. For AI investors, this would convert an unmanageable risk into a priced and bankable one.

The Great British Energy Act provides a route to address this tension more intelligently. Rather than subsidising consumption, GBE can focus on lowering system costs through investment in assets that reduce volatility and congestion.

Reforming Capital and Tax Treatment to Complement Pricing Reform

While energy price itself is the primary barrier to investment in the energy-intensive infrastructure that AI requires, it is also important to recognise the role that capital and tax policy can play in reinforcing radical pricing reform. Existing tools such as the General Capital Allowance regime already allow significant first-year deductions for plant and machinery, including data centre infrastructure. More targeted use of accelerated allowances, or their extension to energy-related assets, such as on-site generation and storage, could materially improve project economics when combined with lower and more stable electricity prices.

Similarly, reforms to the Climate Change Levy and other electricity-specific charges could rebalance incentives so that electrification and digital infrastructure are not penalised relative to fossil alternatives.

Reducing Energy Prices for Data Centres: Lessons from International Market Reform

Countries that have successfully attracted large-scale data centre and AI infrastructure investment have generally not relied on a single policy instrument to reduce energy costs. Instead, they have combined market-led reforms that reshape price formation and risk allocation with targeted fiscal measures that reduce delivered cost, particularly during the early stages of investment. These approaches are instructive for the UK, where high and volatile electricity prices remain a binding constraint on AI infrastructure investment.

Market–led Pricing Reform: Changing How Electricity is Priced and Risk is Allocated

A common feature among successful jurisdictions is the deliberate reduction of exposure to short-term, gas-driven wholesale prices through long-term contracting and structural market design.

Long–term power contracts in Norway

Norway provides the clearest example of this approach. While its hydro-dominated system naturally delivers low-cost electricity, the decisive factor for data centre investment has been the prevalence of long-term bilateral power contracts. Large industrial and digital infrastructure users in Norway typically secure multi-year PPAs priced against long-run hydro costs rather than spot market volatility. Although Norway participates in the Nordic power market, spot prices are not the primary determinant of industrial electricity bills. Gas price shocks, therefore, do not propagate into the cost base of data centres in the same way they do in the UK. This outcome is market-led rather than subsidy-driven: policy stability, credible regulation, and an expectation of long-term contracting have shaped behaviour across the system.

Nuclear power pricing mechanism in France

France offers a different but equally instructive model through its regulated nuclear pricing mechanism, ARENH. By allowing suppliers and large consumers access to nuclear electricity at a regulated price that reflects production costs, France has effectively separated a large share of electricity supply from gas-driven marginal pricing. Although ARENH was not designed specifically for data centres, its impact has been to anchor electricity prices for much of the economy to low-carbon baseload costs rather than volatile fossil fuel markets. This implicit market split significantly reduced price volatility during recent energy crises, and it demonstrates how regulated access to low-cost generation can coexist with a liberalised market framework.

Mature financial markets in the Netherlands, market integration in Germany, and cost stabilisation in Texas

The Netherlands illustrates how financial market maturity can reduce effective electricity prices without direct price controls. Dutch data centre operators routinely rely on long-term corporate PPAs with wind and solar projects, fixing prices for 10 to 15 years and insulating themselves from short-term wholesale volatility. The success of this approach rests on standardised contracts, legal certainty, deep liquidity, and predictable balancing arrangements. Wholesale prices still exist, but they are largely irrelevant to investment decisions because long-term contracts dominate. The UK's PPA market is less mature and more bespoke, which increases basis risk and limits its effectiveness as a price-stabilisation tool.

Germany demonstrates how market integration can reduce costs by allowing large users to interact with the power system as participants rather than passive consumers. Energy-intensive users can lower their effective electricity costs by providing flexibility, participating in balancing markets, and reducing network charges through demand-side response. Although Germany still has high headline electricity prices, these mechanisms allow large, predictable loads to pay closer to system cost rather than retail-style prices. This approach reframes electricity pricing as a function of system value as well as consumption.

Outside Europe, Texas provides a useful comparator. The ERCOT market is characterised by extreme spot price volatility, but large data centres are rarely exposed to it. Instead, they rely on long-term physical and financial hedges that stabilise costs over time. The key lesson is that volatility itself is not necessarily a deterrent to investment if markets provide credible tools to manage it. The UK currently exhibits volatility without equivalent insulation mechanisms for strategic users.

Fiscal and Tax-based Measures: Reducing Delivered Cost Quickly

Alongside market reform, many countries have used targeted tax policy to reduce the delivered cost of electricity for data centres, particularly to accelerate early investment.

Sweden's reduction of electricity tax for large 'computer halls' from 2017 is one of the clearest examples. By cutting electricity tax by approximately 97 per cent for qualifying facilities, Sweden materially lowered operating costs and rapidly improved its attractiveness as a data centre location. Although this incentive was later withdrawn in 2023, its initial success demonstrates how quickly fiscal measures can shift investment decisions. The subsequent reversal also highlights the importance of political durability and clear public-benefit conditions for long-lived assets.

Finland followed a similar path by applying a low electricity tax category to data centres, effectively reducing their energy costs relative to other commercial users. While Finland has since announced that data centres will move back to the higher general tax category from 2026, the period during which the incentive applied was sufficient to attract significant investment. As with Sweden, the lesson is not simply that tax relief works, but that investors price in the risk of future policy reversal.

Denmark combined reduced electricity tax rates for eligible data centres with deliberate system planning. Large data centres benefited from significantly lower electricity taxes while also being integrated into national energy planning, including waste heat utilisation and grid impact assessment. This combination reduced both operating costs and political friction, reinforcing investor confidence.

Norway also used fiscal levers to complement its market structure. A reduced electricity tax rate for data centres introduced in 2016 helped reinforce the country's positioning as a 'data centre nation', even though the primary driver of competitiveness remained long-term contracting and system-cost pricing. The later proposal to remove this relief underlines again that fiscal tools are most effective when paired with stable market fundamentals.

What These Combined Approaches Reveal

Across these cases, a consistent pattern emerges. Countries that successfully reduced energy prices for data centres did not rely solely on tax exemptions, nor did they rely purely on wholesale market reform. Instead, they combined structural market arrangements that anchor prices to low-carbon costs with targeted fiscal measures that reduce delivered price and signal political intent.

Market-led solutions provided the foundation by ensuring that long-term electricity prices reflected system costs rather than short-term marginal scarcity. Tax and fiscal measures then acted as accelerants, improving competitiveness quickly and absorbing residual costs that markets alone could not eliminate. Where either element was missing, outcomes were weaker or less durable.

Implications for the UK

Countries that treat energy pricing as a strategic input rather than a passive market outcome are far more successful at attracting and retaining large-scale data centre investment. As the above examples suggest, addressing energy prices for AI data centres in the UK will require a dual approach. Market reform is necessary to reduce exposure to gas-linked pricing and volatility, whether through expanded CfD-style mechanisms, regulated access to low-carbon baseload, or a stronger framework for long-term contracting. At the same time, targeted fiscal tools, such as Climate Change Levy relief, enhanced capital allowances under the General Capital Allowance regime, or differentiated network charges, can reduce effective prices quickly and improve investment bankability.

Crucially, these interventions must be framed as part of a coherent strategy linked to existing designations such as Critical National Infrastructure or Nationally Significant Infrastructure Projects. Doing so would align energy pricing with the UK's stated priorities on AI, digital resilience, and economic growth, while providing the policy durability that investors require. Ultimately, such alignment and the additional investment it would attract would significantly increase the likelihood of the UK achieving world-leading status in Green AI.

5. Industry and Investor Perspectives on Addressing High Energy Prices for AI Infrastructure

Over the last year, UKAI has engaged with data centre operators, infrastructure investors, and AI businesses. These discussions have revealed a striking degree of alignment around the issues and diagnosis set out in earlier sections of this chapter. From an industry perspective, high and volatile energy prices in the UK are not viewed solely as the inevitable outcome of decarbonisation or scarcity, but as the product of specific market design choices that unnecessarily expose strategic infrastructure to gas-linked pricing. As a result, investors increasingly see energy pricing as the single most important constraint on the scale and type of AI infrastructure that can be viably deployed in the UK.

Energy Price Volatility as the Primary Deterrent

Investors consistently emphasise that it is not only the absolute level of electricity prices that undermines investment confidence, but their volatility and unpredictability. As this report has highlighted, data centres, particularly those supporting AI training workloads, are long-lived, capital-intensive assets with investment horizons of 10 to 20 years. Exposure to short-term wholesale price fluctuations, driven by global gas markets rather than domestic system costs, introduces risks that are difficult to hedge fully and that materially affect financing terms.

From the perspective of equity investors and lenders, the UK market is perceived as one where electricity prices are structurally misaligned with underlying generation costs, and where future policy responses to price spikes remain uncertain. This contrasts with jurisdictions such as Norway or France, where long-term price anchors exist either through market practice or regulation, and where investors can model operating costs with greater confidence.

Support for Market-led Reform That Reflects Physical System Reality

Data centre operators and their investors broadly support market reforms that align electricity prices more closely with physical system conditions. In particular, there is strong industry interest in locational or zonal pricing models that would allow regions with abundant renewable generation to offer materially lower electricity prices. Investors see this as a rational correction rather than a subsidy, as it reflects the real cost of supplying power in different parts of the system and would encourage energy-intensive infrastructure to locate where it delivers system value.

Similarly, there is widespread support for reforms that reduce the degree to which gas sets electricity prices when gas is not the dominant source of generation. Industry stakeholders often point to the expansion of mechanisms such as CfD, or to regulated access to low-carbon baseload power, as ways of anchoring prices to long-run costs rather than marginal scarcity. From an investor standpoint, these approaches are attractive because they reduce downside risk without eliminating market incentives for efficiency and flexibility.

Using Fiscal and Regulatory Levers to Reduce Delivered Cost

Alongside market reform, industry strongly supports targeted measures that reduce the delivered cost of electricity without altering wholesale prices directly. Existing mechanisms such as Climate Change Agreements, which provide substantial discounts on the Climate Change Levy in exchange for efficiency commitments, are widely used and viewed as effective. However, investors argue that these measures are insufficient on their own and should be complemented by broader recognition of data centres as strategically important energy users.

There is particular interest in extending reliefs currently available to energy-intensive industries to qualifying data centre projects, especially those supporting AI capability, public-sector digital resilience, or national security objectives. From the industry's perspective, this would be a logical extension of existing policy rather than a novel subsidy, aligning cost relief with strategic importance.

Capital allowances also play a role in shaping investment decisions, and His Majesty's Treasury (HMT) should explore their potential to incentivise long-term investment. Enhanced or accelerated allowances for energy-related infrastructure – including on-site generation, storage, and grid connections – are seen as a way to improve project economics at the margin, particularly when combined with lower and more stable electricity prices.

Critical National Infrastructure as a Foundation

While the designation of data centres as Critical National Infrastructure (CNI) has been welcomed by industry, it is widely understood as a necessary but incomplete step. Investors value the signal that data centres are nationally important, particularly in terms of security and resilience, but they note that this designation currently has little impact on operating costs or energy pricing.

Industry consensus is that CNI status should be used as a policy foundation upon which economic measures are built, rather than as an end in itself. In practical terms, this means linking strategic designation to tangible benefits such as access to long-term power contracts, priority grid connection, differentiated network charges, or participation in pilot pricing mechanisms. Without this linkage, CNI status risks being symbolic rather than transformative.

Preference for Coherent Packages Rather Than Isolated Interventions

A consistent theme across operator and investor feedback – and our report – is that isolated interventions are less effective than coherent policy packages. Tax relief without market reform may improve short-term competitiveness, but it does not address volatility or long-term risk. Market reform without transitional fiscal support may be too slow to influence near-term investment decisions. Investors therefore favour approaches that combine pricing reform, contracting mechanisms, and targeted reliefs into a single, credible offer.

Jurisdictions that have succeeded in attracting data centres are perceived as those that provided clarity and consistency across these dimensions, rather than those that relied on any single policy lever. For the UK, this implies that addressing high energy prices for AI infrastructure requires coordination across DESNZ, HMT, Ofgem, and DSIT, with a shared understanding that energy pricing is now a central component of any digital and industrial strategy. In that context, it is notable that energy pricing is currently absent from the Government's Modern Industrial Strategy (which is discussed in greater detail in Section 6 of this chapter).

Implications for UK Policy

From the industry's perspective, the UK does not need to invent entirely new mechanisms to reduce energy prices for data centres, but it does need to apply existing tools more strategically and coherently. Market-led reforms that reduce gas-driven pricing and normalise long-term contracting, combined with targeted fiscal measures that reduce delivered cost, are viewed as both credible and necessary.

Most importantly, investors stress that policy durability matters as much as policy ambition. Long-lived AI infrastructure will only be built at scale where there is confidence that energy pricing arrangements will remain stable over time. Addressing high and volatile energy prices is therefore not simply a matter of competitiveness, but of credibility in the UK's broader commitment to being a serious, long-term home for AI infrastructure. This report's vision of the UK leading the world in Green AI cannot be achieved without significant and sustained efforts to address energy pricing issues.

6. Implications of the Government's Modern Industrial Strategy for AI Data Centres

The Government's Modern Industrial Strategy addresses high energy prices primarily through targeted interventions in non-commodity costs, network charges, and investment enablers. While these measures were not designed exclusively with AI infrastructure in mind, they have clear and material implications for AI data centres, particularly those operating at scale or supporting training workloads.

The British Industrial Competitiveness Scheme (BICS) is the most significant prospective intervention. By exempting eligible businesses from policy-driven costs such as the Renewables Obligation, Feed-in Tariffs, and Capacity Market charges, BICS directly reduces the delivered cost of electricity. For AI data centres, which are large, constant electricity consumers, these non-commodity costs can represent a substantial share of total energy expenditure. If AI data centres are included within the scope of BICS, or if an equivalent mechanism is extended to them, the result would be improved long-term price predictability and an immediate reduction in operating costs. However, if eligibility remains confined to traditional industrial sectors, AI data centres will continue to face a structural disadvantage, despite exhibiting similar electricity intensity and international mobility.

The British Industry Supercharger, particularly the expansion of Network Charging Compensation, also has direct relevance for AI data centres. Network charges are a significant component of electricity bills for large users and are largely independent of how efficiently electricity is generated. For AI data centres, which often require large grid connections and operate continuously, these charges can materially affect site viability. Enhanced compensation reduces exposure to these costs and improves the economics of locating high-demand facilities in the UK, especially in regions where grid reinforcement costs would otherwise be reflected in higher charges.

The continuation and expansion of Energy Intensive Industries (EII) exemptions provides an important precedent rather than a direct benefit to AI data centres. While data centres are not currently eligible under the EII framework, the Industrial Strategy's reliance on EII-style exemptions signals acceptance of a core principle: that policy costs should not undermine the competitiveness of strategically important, electricity-intensive activity. For AI data centres, this strengthens the policy case for either expanding EII eligibility or creating a parallel framework that delivers similar outcomes, particularly given the role of AI infrastructure in productivity, national security, and digital resilience.

Measures aimed at accelerating and prioritising grid connections are particularly important for AI data centres, where delays in securing capacity can add years to development timelines and significantly increase costs. Faster grid access reduces the need for temporary generation, diesel backup, or suboptimal interim power solutions, all of which inflate effective energy prices and undermine sustainability objectives. For AI infrastructure, improved grid coordination directly affects not just cost but also deployment speed, which is a critical factor in global competition for AI investment.

The strategy's emphasis on investment incentives, including enhanced capital allowances, also has a meaningful impact on AI data centres. While capital allowances do not reduce electricity prices directly, they lower the after-tax cost of building energy-resilient infrastructure such as substations, backup generation, batteries, and cooling systems. For AI data centres designed to meet four-nines availability requirements, these energy-related capital costs are substantial. Accelerated allowances improve cash flow in the early years of a project and can partially offset the impact of higher operating energy costs, making UK-based investments more financially viable.

Success will depend on coordinated action with government support

Taken together, the Modern Industrial Strategy's energy pricing measures potentially offer partial but important relief for AI data centres. They reduce certain cost components, improve investment economics, and signal a willingness to intervene where energy pricing undermines competitiveness. However, the effectiveness of these measures for AI infrastructure ultimately depends on whether data centres are explicitly included within their scope or are able to access equivalent mechanisms. Without that inclusion, the strategy risks leaving AI data centres exposed to the same high and volatile energy prices that continue to constrain large-scale AI investment in the UK, thus decreasing the likelihood of the UK achieving its potential as a Green AI leader.

7. Extending Energy Intensive Industries (EII) to Data Centres

The UK's EII framework was created to address a longstanding competitiveness challenge: the risk that high electricity prices, driven in part by domestic policy costs and network charges, would disadvantage energy-heavy industries, relative to international competitors. Introduced in the mid-2010s, the EII regime was designed to protect sectors such as steel, chemicals, cement, glass, and paper, where electricity costs form a large proportion of operating expenditure and where exposure to global competition is acute.

Over time, the scheme has evolved and expanded. Eligible EIIs now benefit from exemptions from the indirect costs of major electricity policy levies, including Contracts for Difference, the Renewables Obligation, and Feed-in Tariffs, alongside exemptions from Capacity Market charges and increasing compensation for network charges through the British Industry Supercharger package. These measures do not alter wholesale electricity prices, but they materially reduce the *delivered cost* of electricity and have been successful in narrowing the gap between UK industrial energy prices and those in comparable European economies. For qualifying sectors, EII support has provided both cost relief and greater certainty, helping to retain investment and mitigate the risk of industrial offshoring.

The underlying logic of EII is directly relevant to data centres, particularly those supporting AI training and large-scale digital infrastructure. Like traditional EIIs, modern data centres are highly electricity-intensive, internationally mobile, and sensitive to energy price differentials. While they differ from manufacturing in output and labour profile, their exposure to electricity costs is comparable in scale and importance. In this sense, the challenge facing AI data centres mirrors the challenge that EII was originally designed to address: electricity pricing that reflects policy and system costs rather than international competitiveness or strategic value.

Extending EII-style support to data centres would not solve all aspects of the UK's energy pricing problem, particularly the exposure to gas-linked wholesale prices, but it would directly reduce non-commodity costs that inflate electricity bills. Exemptions from policy levies, capacity charges, and a greater share of network cost compensation would lower operating costs immediately and improve the bankability of long-lived AI infrastructure investments. Crucially, these are areas where the UK already has established legal and administrative mechanisms, reducing the need for entirely new policy instruments.

There are two principal ways such an extension could be achieved. One option would be to formally expand EII eligibility to include data centres or a new category of 'digital infrastructure' within the existing framework. This would deliver rapid cost relief using a familiar mechanism but would require revisiting sectoral definitions and intensity thresholds that were originally designed for manufacturing. A second, increasingly discussed, option would be to create a parallel framework that mirrors EII outcomes for strategically important electricity users, rather than modifying EII itself. Such a framework could apply to data centres designated as Critical National Infrastructure or approved through Nationally Significant Infrastructure Project processes, linking cost relief explicitly to national priorities around AI capability, resilience, and economic growth.

Both approaches face challenges. Expanding EII raises questions about precedent and scope, particularly concerns that relief designed for heavy industry could be perceived as a subsidy for the technology sector. A parallel scheme would require new policy design and coordination across departments, notably DESNZ, HMT, Ofgem, and DSIT. In either case, political durability will be critical: investors in AI infrastructure consistently stress that the stability of energy-cost arrangements over 10 to 20 years matters as much as the level of relief itself. Since 2020, national and international politics has been defined by unpredictability and extremes, particularly around energy supply and the environment.

Importantly, extending EII-style support to data centres would align with, rather than cut across, current government policy. The designation of data centres as Critical National Infrastructure signals their strategic importance but does not yet translate into economic or energy-pricing benefits. The British Industry Supercharger demonstrates a clear willingness to use exemptions and compensation to address industrial electricity costs, while AI Growth Zones and planning reforms indicate an appetite to remove barriers to AI infrastructure investment. An EII extension or equivalent framework would provide the missing link between these initiatives by directly addressing the high and volatile cost of power, which remains the primary constraint identified by the AI industry and investors.

In summary, the EII regime offers a proven model for reducing the delivered cost of electricity for strategically important, energy-intensive activities. While data centres are not currently covered, the rationale for EII support increasingly applies to AI infrastructure. Whether through the expansion of the existing scheme or the creation of a parallel mechanism, adapting the principles of EII to modern digital infrastructure would represent a pragmatic, immediately actionable step towards improving the UK's competitiveness in AI while building on established legislation and policy commitments.

8. Additional Mechanisms to Manage Energy Pricing

Beyond wholesale market reform and the Energy Intensive Industries framework, there exists a wider set of regulatory, market, and planning mechanisms that materially affect the effective price of electricity faced by data centres in the UK. These mechanisms are often less visible than headline energy prices, yet they play a significant role in shaping investment decisions, particularly for large-scale AI infrastructure. Understanding and activating these levers is essential if the UK is to address energy pricing constraints in a comprehensive and coordinated way.

Make Network Charges Dynamic

Transmission and Distribution Use of System (TDUoS) charges constitute a significant and sometimes volatile component of electricity bills for large consumers. While these network charges are intended to reflect the cost of maintaining and expanding the grid, in practice, they can penalise large, constant loads such as data centres, even where those loads are located in regions with abundant renewable generation.

Ofgem already has clear statutory authority to reform how electricity network charges are designed, allocated, and recovered, and it has exercised this power in ways that materially affect consumer costs. Recent and ongoing reforms show that network charges are not fixed reflections of physical infrastructure costs but are highly sensitive to regulatory and policy decisions. Through interventions such as the Targeted Charging Review, which restructured residual network charges, the Distribution Use of System (DUoS) Significant Code Review, and its current Cost Allocation and Recovery Review,[6] Ofgem has actively reshaped who pays network costs, how those costs are spread, and what price signals are sent to large users. These reforms demonstrate that network charging methodologies are policy tools, capable of being redesigned to support wider system objectives such as efficiency, decarbonisation, and investment. Therefore, it seems reasonable to suggest that such reforms should be considered as ways to increase the development of Green AI in the UK. Reforms that adjust how charges are allocated, how locational signals are applied, or how strategic loads are treated could reduce delivered electricity costs for data centres without altering wholesale prices.

Simplify and Incentivise Direct Connection

Private wire and direct connection arrangements represent another underutilised mechanism. By connecting directly to renewable generation assets, data centres can reduce exposure to wholesale price volatility, network charges, and certain levies. Although such arrangements are legally permissible in the UK, they remain complex, slow to implement, and subject to regulatory uncertainty. However, clearer guidance, standardised approval processes, and explicit recognition of private wire as a system benefit, particularly where it reduces congestion and curtailment, could make such arrangements viable, market-led routes to lower effective energy prices for AI infrastructure.

Building direct connections within AI clusters should be part of a coordinated strategy. Existing regional programmes (such as AI Growth Zones or Green Freeports) could be used to accelerate planning and incentivise investment, and they could also provide an ecosystem of partners within the cluster.

Data Centres Should be Treated as Part of the Solution

The Capacity Market is a mechanism that ensures electricity supply remains reliable by paying generators, storage providers, and demand-side participants to be available during periods of peak demand, rather than only paying for electricity when it is generated. Through competitive auctions held several years in advance, capacity providers secure contracts that guarantee payment in exchange for a commitment to deliver power or reduce demand when called upon by the system operator. This approach supports security of supply as the energy system incorporates more intermittent renewables, with the costs of the scheme recovered through electricity bills paid by consumers.

The Capacity Market and ancillary services markets also offer opportunities to offset energy costs, yet data centres are not currently positioned to benefit fully from them. Large data centres increasingly have the technical capability to provide demand-side

6 https://www.ofgem.gov.uk/press-release/ofgem-announces-major-review-how-costs-are-allocated-across-energy-system

response, on-site generation, battery storage, and load flexibility. However, current market design treats them primarily as passive consumers rather than as contributors to system stability. Reforming these markets to better reward flexibility and availability would allow data centres to generate revenues that reduce their net electricity costs while improving overall system resilience.

Leverage CNI and NSIP to Expedite and Coordinate Planning

Planning and infrastructure designation provide further indirect levers. Where data centres are designated as Nationally Significant Infrastructure Projects, the Government has greater scope to coordinate planning decisions, grid reinforcement, and infrastructure delivery. While NSIP status does not automatically reduce energy prices, it creates a framework in which bespoke energy arrangements – such as integrated generation, storage, and connection solutions – become easier to justify on national interest grounds. This approach could help align energy cost outcomes with strategic objectives around AI, sovereignty, and digital resilience.

Harness Carbon Trading to Deliver Sustainable Compute

The UK is internationally recognised for its leadership in climate governance, carbon markets, and the use of market-based mechanisms to support environmental responsibility, all of which emphasises the country's potential to lead in the development of Green AI. Nevertheless, commitments to Net Zero are often seen as being contrary to the interests of the energy-intensive data centres that power the AI industry. Carbon trading has proven to be a successful way to balance the needs of broader industry with its environmental responsibilities, and the UK could pioneer leveraging these mechanisms to create a more sustainable AI industry.

The UK has pioneered carbon trading

Carbon trading reduces emissions by pricing CO_2 equivalent emissions and enabling the trading of allowances or credits. Its core benefit is efficiency: emissions reductions occur where they are cheapest, while investment is steered toward cleaner technologies through transparent price signals rather than prescriptive regulation. Well-designed carbon markets also improve predictability for investors, reward innovation and create liquid financial instruments that can be integrated into wider energy and commodity markets. For governments, carbon trading can align decarbonisation with competitiveness by using markets to allocate risk and capital rather than relying solely on subsidies. The UK has played a central role in the development of carbon trading, from early participation in the EU Emissions Trading System to the creation of the UK ETS and the use of market-based mechanisms such as Contracts for Difference. London has become a global hub for carbon, energy, and environmental commodity trading, supported by deep financial markets and regulatory expertise.

Rewarding energy-efficient data centres

Applied to data centres, carbon trading could help reduce effective energy pricing by changing how power is procured and used rather than simply adding cost. If data centres can earn credits or reduce obligations by consuming electricity during low-carbon periods, locating near clean generation, or contracting long-term low-carbon power, carbon pricing becomes a tool to *optimise cost*, not just penalise emissions. In this way, carbon markets could reinforce load-shifting, long-term PPAs, and firm low-carbon supply, lowering exposure to volatile, gas-linked prices. Over time, integrating demand-side carbon signals with power markets could reduce system costs that are currently socialised, indirectly benefiting large, flexible users such as data centres.

Incentivise more transparent reporting

From a sustainability perspective, carbon trading would provide a clearer and more credible framework for managing the environmental impact of rapidly growing AI infrastructure. Rather than relying primarily on voluntary offsets or certificates, data centres could operate within a regulated system that rewards genuine emissions reductions and system-friendly behaviour. This would improve transparency around Scope 2 emissions, incentivise investment in on-site and nearby clean energy, storage, and flexibility, and reduce the risk of greenwashing. For operators, it would translate sustainability from a reputational objective into a core operational and financial discipline.

The UK is building from a position of strength

The UK is well placed to lead on this agenda because it already combines deep experience in carbon markets, sophisticated energy regulation, and a strong financial services sector. Potential future models include sectoral baseline-and-credit schemes for data centres, grid-intensity-linked carbon obligations tied to real-time consumption, or hybrid systems where large electricity users trade credits based on when and where they consume power. Getting started would not require immediate inclusion of data centres in the UK ETS; instead, the UK could pilot reporting-linked trading schemes, voluntary-but-standardised credit markets, or regional trials aligned with AI Growth Zones. This incremental approach would allow market design to mature while positioning the UK as an innovator in demand-side carbon pricing for the digital economy, strengthening its position as a leader in the development of Green AI.

9. Building Carbon Trading for Data Centres: Getting Started

Consistent measurement standards

The first step toward carbon trading for data centres is building the foundations that make demand-side carbon pricing workable. This begins with consistent, high-quality measurement and reporting of electricity consumption and associated CO_2-equivalent emissions at the facility level. The UK already has much of this capability through smart metering, energy reporting, and emerging data centre transparency requirements, but it would need to be standardised and aligned with carbon accounting methodologies suitable for trading rather than disclosure alone.

Defining pilot projects

From there, government and industry would need to agree the scope and objective of an initial scheme. Early activity is most likely to take the form of a pilot or sandbox, rather than a mandatory programme. This could focus on large data centres above a capacity threshold, or on facilities located in designated AI Growth Zones or energy-constrained regions. The aim at this stage would be learning: understanding how carbon signals interact with energy pricing, load-shifting, long-term power contracts, and grid constraints.

A critical early step would be designing a mechanism that avoids double-counting with existing upstream carbon pricing on generators. This means carbon trading for data centres would likely be structured as a baseline-and-credit or intensity-based system, rather than a classic cap-and-trade model. That design choice allows carbon pricing to reward better-than-average performance without re-pricing emissions already covered elsewhere.

The Government must lead a coordinated approach

No single organisation can deliver this alone. Progress would require coordination across government, regulators, industry, and financial markets. Within the Government, DESNZ would lead on policy design, ensuring alignment with the UK ETS, electricity market reform, and net zero objectives. DSIT would play a key role in defining which data centres are strategically important and ensuring the scheme supports AI capability rather than deterring investment. HM Treasury would be central to decisions about fiscal treatment, interaction with existing carbon pricing, and whether any incentives or safeguards are required during early phases.

Ofgem's role would be indirect but important, particularly in ensuring that any carbon trading signals align with network charging, balancing arrangements, and system operation, rather than cutting across them.

Outside the Government, UKAI – as the industry body for the entire AI sector, representing data centres and AI businesses – would be an essential partner in the process of shaping workable rules. Equally important are carbon market institutions, including exchanges, registries, and verification bodies, many of which are already based in London and have experience operating emissions and environmental commodity markets.

Some of the infrastructure and expertise already exists

Many hyperscale operators already operate internal carbon pricing systems, effectively simulating carbon markets within their own organisations. At the same time, voluntary markets for renewable energy attributes, long-term PPAs, and corporate offsetting have created the operational infrastructure needed for trading, verification, and settlement.

On the policy side, the UK has experience using regulatory sandboxes and pilot programmes in both energy and financial services. A similar approach could be adopted here: a voluntary, time-limited pilot that allows participating data centres to earn and trade credits based on carbon-intensity performance or system-friendly behaviour, without immediate compliance obligations.

Carbon trading produces better data and efficiency

If designed well, the benefits would be shared rather than concentrated. Data centre operators would gain new tools to manage both carbon exposure and energy costs, turning flexibility, location choice, and long-term clean power procurement into tradable value. This could reduce effective energy pricing for operators that behave in system-beneficial ways.

The energy system would benefit from better demand-side signals, encouraging consumption when the grid is clean and unconstrained and reducing pressure during high-carbon or high-stress periods. This helps lower system costs that are currently socialised across all consumers.

Enabling long-term planning and incentivising investment

For data centre operators and investors, the main incentive to accept carbon trading is control. Done properly, carbon trading does not just add a cost; it creates a mechanism to manage risk, reduce long-term energy exposure, and improve asset value in ways that are otherwise difficult or unavailable.

The most immediate advantage is greater predictability of costs. As has been repeatedly stressed throughout this chapter, energy price volatility is one of the biggest risks facing data centre investments. A carbon trading framework that rewards low-carbon, firm, and system-friendly electricity consumption allows operators to convert operational choices – such as long-term PPAs, location near clean generation, or flexible load management – into tradable value. This can offset carbon obligations and smooth overall energy costs, making long-term financial modelling more reliable and improving financing terms.

A second incentive is access to capital. Investors are increasingly constrained by climate-related risk requirements and sustainability mandates. Data centres that operate within a regulated carbon framework with transparent pricing and compliance mechanisms are easier to underwrite, insure, and finance. Carbon trading provides a recognised, auditable way to manage emissions exposure, reducing regulatory and reputational risk and aligning assets with institutional capital requirements.

Third, carbon trading can differentiate assets competitively. In a world where AI infrastructure is globally mobile, facilities that can demonstrate lower carbon intensity and participation in credible carbon markets become more attractive to hyperscalers, enterprise customers, and governments. This can translate into higher utilisation, longer contracts, and lower risk premiums, all of which matter more to investors than short-term operating cost.

Accepting carbon obligations early, through pilots or voluntary schemes, gives operators a seat at the table to shape future regulation. It allows them to influence design, secure transitional protections, and develop internal capabilities before participation becomes widespread or mandatory. From an investor's perspective, this reduces the risk of abrupt policy change and creates a strategic advantage over competitors who are forced to adapt later.

Finally, the UK economy would benefit from reinforcing its position as a hub for carbon markets, climate finance, and digital infrastructure, creating new financial products and services while supporting AI investment rather than pushing it offshore. Carbon trading for the data centre market and products could have a global application, making this an important new area where the UK can lead.

In short, the incentive is not compliance for its own sake, but the opportunity to turn sustainability from a constraint into a financial and strategic lever that improves resilience, competitiveness, and long-term returns.

The Government provides the framework for the market, financial institutions manage the detail

The Government's role is to set the rules, ensure integrity, and manage systemic risk. This includes defining eligibility, setting baselines or performance thresholds, ensuring compatibility with the UK ETS, and protecting competitiveness during early phases. Together with UKAI, the Government also has a convening role, bringing together energy, AI, and financial policy communities that do not always interact.

Financial institutions would provide the market plumbing. Banks, exchanges, clearing houses, and insurers would structure products, manage risk, provide liquidity and help participants hedge exposure. London's existing strengths in carbon trading, energy derivatives, and environmental finance mean this capability already exists; what is missing is a new class of demand-side asset to trade.

An opportunity for the UK to seize in 2026

Carbon trading for data centres is not an abstract or distant idea, but a logical extension of existing UK strengths in energy markets, carbon pricing, and financial innovation. The next steps lie in piloting, not mandating; in collaboration, not imposition; and in designing mechanisms that help data centres reduce both emissions and cost exposure. If approached in this way, the UK could move first, learn fastest, and set the template for how digital infrastructure is integrated into carbon markets globally.

Policy Recommendations

Quick Impact Priorities (0–24 months)

1 Anchor Electricity Prices for Strategic Infrastructure through Long-term, Low-carbon Contracting

Establish long-term, fixed-price electricity as the default for strategically important data centres, reducing exposure to gas-driven spot markets and volatility. Rather than redesigning the entire market, this approach focuses on accelerating access to stable power using existing contractual tools.

Implementation should combine mandatory long-term contracting for qualifying large loads with government-backed aggregation or underwriting to reduce basis risk. A state-backed strategic power buyer or hedging vehicle could pool demand and contract with low-carbon generators, recycling CfD-style revenues to dampen volatility. Delivery requires coordination between DESNZ, Ofgem and Great British Energy to standardise contracts, normalise long-term PPAs, and ensure these arrangements are financeable and rapidly deployable.

2 Reduce Delivered Electricity Costs for Data Centres Using Established Fiscal and Levy Frameworks

Extend proven industrial relief mechanisms to AI data centres, without waiting for wholesale market reform. This reflects how non-commodity costs materially undermine competitiveness even when wholesale prices fall.

Implementation should adapt existing frameworks rather than invent new ones. Energy Intensive Industries (EII) relief or an equivalent parallel scheme should be extended to qualifying data centres, alongside exemptions from Capacity Market charges and enhanced network cost compensation. Capital allowance reform should explicitly cover energy-related infrastructure such as substations, storage and direct connections. HM Treasury, working with DESNZ and DSIT, should link eligibility to strategic designations (CNI, NSIP, AI Growth Zones) to ensure durability and legitimacy. Additional criteria may incentivise 'system friendly behaviour' such as proximity to supply, flexibility provision and efficiency thresholds. Any new mechanism should include suitable sunset mechanisms to ensure this programme is fiscally sustainable.

3 Treat Data Centres as Active System Participants, not Passive Energy Consumers

Enable data centres to offset energy costs by contributing flexibility, capacity, and resilience to the electricity system. This reflects the chapter's argument that pricing reform should reward system value, not just consumption.

Implementation requires reform of the Capacity Market and ancillary services so that data centres can be incentivised to earn revenues for demand-side response, on-site generation and storage. In parallel, Ofgem should reform network charging methodologies to be more dynamic and locational, reducing penalties for constant loads located near abundant generation. Clear market access rules and simplified participation pathways would allow operators to monetise flexibility, lowering net energy costs while improving system resilience.

As a short-term solution, practical mechanisms such as Demand Side Response (DSR) can demonstrate near-term feasibility by enabling flexible compute loads to support grid balancing. This helps incentivise market change by showing that AI and data-centre infrastructure can operate as a flexible, grid-responsive asset rather than a fixed source of demand.

Structural Priorities (2–10 years)

4 Reframe Electricity for Strategic Digital Infrastructure as National Infrastructure, Not a Traded Commodity

This recommendation establishes a durable policy foundation by explicitly recognising electricity supply to strategic digital infrastructure as infrastructure in its own right. It addresses the structural misalignment between long-lived AI assets and short-term marginal pricing.

Implementation requires a formal policy shift that allows differentiated pricing, priority access, and tailored contracting for strategically designated assets. Electricity for CNI-designated data centres should be priced against long-run system costs rather than marginal gas scarcity. This does not require renationalisation, but it does require clear cross-government alignment so that market rules, regulation, and planning decisions consistently reflect strategic intent. This further underlines the importance of having sovereign infrastructure for both compute and energy supply.

Any differentiated pricing or access must be based on clear, non-arbitrary criteria such as Critical National Infrastructure designation and national security considerations, with Ofgem oversight to ensure transparency and compliance with non-discrimination principles.

5 Separate Low-carbon Baseload Pricing from Marginal Generation to Reduce Gas-driven Volatility

Restructure price formation so that cheap low-carbon electricity is no longer priced as if it were gas. This builds directly on CfD principles but extends them system-wide.

Under this approach, existing CfD would continue to operate, but their role would evolve. Rather than acting purely as a top-up or clawback mechanism against a gas-set wholesale price, CfDs for nuclear and renewables would effectively feed into the low-carbon baseload pool, with strike prices informing the averaged or regulated baseload price. This would reduce the scale and frequency of CfD payments, lower exposure of consumers and the public purse to gas-driven price spikes, and maintain investor confidence by preserving contracted revenue certainty.

Implementation would involve splitting the market into a low-carbon baseload pool (nuclear and contracted renewables) priced on averaged or regulated cost, and a marginal market for balancing and peaking assets. Gas would continue to play a critical role, but only where it is genuinely needed. For data centres, this would anchor prices to underlying generation costs while preserving incentives for flexibility, storage, and firm capacity investment.

6 Use Spatial Pricing and Direct Connections to Align AI Infrastructure with Energy Abundance

This recommendation turns grid congestion and renewable curtailment into an advantage by aligning location decisions with system reality. It supports structurally lower prices without ongoing subsidy.

Implementation should combine zonal or spatial pricing with proactive industrial strategy. Regions with persistent low-cost, low-carbon oversupply should offer lower local electricity prices, complemented by planning fast-tracks, NSIP designation, grid prioritisation and targeted tax relief. Direct and private-wire connections should be simplified, standardised and incentivised where they reduce congestion. This requires coordination across Ofgem, planning authorities, local and devolved governments to ensure reliability standards suitable for four-nines data centre operation. It should be market driven and could be further accelerated by empowering local government and combined authorities. We recognise that consumer electricity pricing is a politically sensitive issue and consumers' needs should be central to the evaluation of feasibility, to ensure a fair solution that will build trust and engage communities.

7 Establish a Carbon Trading Framework for Data Centres to Align Cost, Efficiency, and Sustainability

Introduce carbon trading as a core strategic lever, not a compliance burden, making it a prominent pillar of energy pricing reform. This would enable data centres to convert low-carbon, system-friendly behaviour into financial value.

Implementation should begin with voluntary pilot, baseline-and-credit schemes for large data centres, aligned with but separate and different from the UK ETS. Facilities would earn tradable credits by consuming low-carbon power, locating near clean generation, shifting load, or contracting long-term clean supply. The Government sets the rules and safeguards competitiveness; financial markets provide liquidity and risk management. Over time, this would lower effective energy costs, improve investment certainty, and reinforce the UK's leadership in carbon and digital infrastructure markets.

Strategic Implications: Energy Pricing as a Strategic Lever for AI

These mechanisms are intended to function collectively as a coherent framework, not as a set of disconnected or ad-hoc interventions. As we also highlighted in the previous chapter, the central challenge facing the UK is not the absence of policy tools, but the absence of coordination across energy, fiscal, planning, and digital policy. As this chapter has shown, energy pricing for AI data centres is shaped simultaneously by wholesale market design, non-commodity charges, network costs, planning decisions, and strategic designation. When treated as separate domains, these factors have combined to unintentionally penalise energy-intensive digital infrastructure, even as government policy seeks to promote AI capability, productivity, and growth.

A key implication of this chapter is that energy pricing must be treated as a strategic input to AI infrastructure, not as a passive outcome of global markets. The UK does not lack low-carbon energy resources, AI capability, or investor interest. What it does lack is a pricing framework that reflects the realities of its generation mix rather than the marginal cost of gas and provides long-term certainty for capital-intensive infrastructure. Addressing this misalignment is therefore not a peripheral energy issue, but a foundational condition for delivering the UK's AI ambitions and realising the country's potential to pioneer the development of Green AI.

The chapter also demonstrates that no single intervention is sufficient on its own. The most credible path forward is layered rather than radical. Measures that build on existing schemes can provide near-term relief and signal intent, while market and network reforms can reshape price formation and risk allocation over time. Planning and strategic designation provide the enabling context that makes differentiated treatment legitimate and durable. What matters most to investors is not which individual lever is pulled, but whether the Government can demonstrate a stable, cross-government commitment to aligning energy pricing with its long-term objectives for AI, resilience, and economic competitiveness.

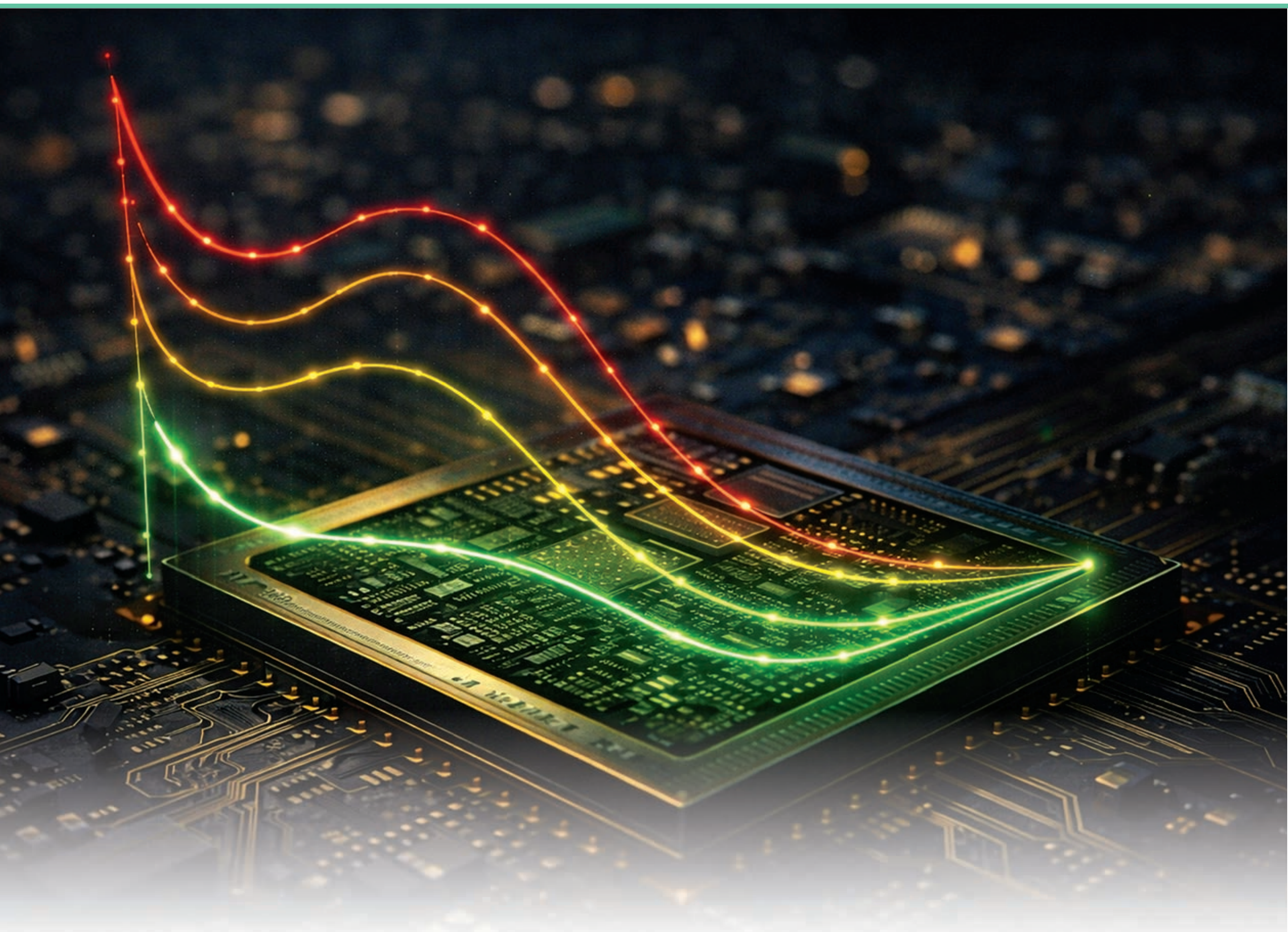
At the same time, this chapter highlights an important boundary. Lower or more stable energy prices alone do not guarantee better outcomes. Without parallel attention to how energy is used, pricing reform risks enabling higher volumes of low-value or inefficient compute. Energy policy can create the conditions for investment, but it cannot determine whether that investment delivers productivity, innovation, or public value.

This leads directly to the focus of the next chapter, which examines how constraints on energy cost, availability, and volatility can be translated into competitive advantage through targeted innovation in models, hardware, system architecture, and deployment choices. Together, these chapters frame a Green AI approach that is not defined solely by emissions, but by efficiency, resilience, and the intelligent use of scarce resources. Concurrently, the Government can use policy to lower systemic risk and encourage innovation to ensure that every unit of energy consumed by Green AI delivers maximum economic and social value.

Industry Perspective

"What makes UKAI's report powerful is that it moves beyond hype and into delivery. By putting energy resilience, scalable systems and community benefit at the heart of AI infrastructure, this report shows how the UK can grow its AI economy in a unique way that is both economically strong and socially legitimate."

Spencer Lamb, MD and Chief Commercial Officer
Kao Data



Chapter 3 – Targeted Innovation

Pioneering solutions to enable future-ready systems.

1. Introduction: the UK is a Green Technology Innovator

Debates about the environmental impact of AI often focus on scale: ever-larger models, ever-expanding data centres, and rising electricity demand. This chapter takes a different view. It shows that the future of AI does not depend solely on building more compute, but on how intelligently that compute is designed, deployed and integrated. Across hardware, software, infrastructure, and local deployment, innovation is already reshaping the energy profile of AI, delivering more capability with less energy, capital, and material use.

Chapter 3 explores this shift in detail. It examines targeted innovations in compute hardware that reduce energy lost to data movement, power delivery, and heat, including photonics, compound semiconductors, neuromorphic architectures, and next-generation cooling. It then looks beyond the chip to the full lifecycle of data centres, highlighting UK strengths in low-carbon construction, power efficiency, digital twins, and predictive maintenance. At the model and application layer, the chapter shows how smaller, adaptive, and frugal approaches to AI can dramatically reduce energy demand without sacrificing real-world performance. Finally, it considers innovation at the 'place layer': modular data centres, edge inference, shared infrastructure, and waste-heat reuse that embed AI more productively into communities.

As the trade body for AI businesses across the UK, this is the work that UKAI does. We find, engage, and champion the innovative companies, researchers, and practitioners who are already building this future, and we highlight the role they play in making the AI industry more efficient, resilient, and sustainable. The examples in this chapter are not speculative. They are real UK businesses, universities, and local initiatives demonstrating what the next phase of AI infrastructure looks like in practice.

Crucially, this research also makes clear where the UK can lead. It will not win by competing on the sheer scale of compute deployment. But there are many areas where the UK has genuine, structural advantage: efficiency-driven hardware, application-layer innovation, systems integration, and place-based deployment models that work under real-world constraints. Whilst we have attempted to be comprehensive, this chapter only touches the tip of the iceberg when it comes to cataloguing the UK innovation ecosystem. It concludes with recommendations for immediate actions and structural changes that can leverage the UK's strength in innovation as part of a national mission to become a leader in Green AI.

The next challenge is to identify where the UK should focus, and then to surface, showcase and connect this innovation, to ensure that it scales. That is why this chapter leads directly into Chapter 4, which explores how to create the systems, shared infrastructure, and coordination mechanisms to bring these innovations together and multiply their benefits. Taken together, the two chapters demonstrate how targeted innovation is central to delivering Green AI in practice.

2. Energy Efficiency in Compute Hardware

Concerns about AI's electricity use often focus on the scale of computation required by modern models. In practice, however, energy demand is shaped at least as much by how efficiently data is moved, how selectively computation is performed, how power is delivered, and how heat is removed. Innovation across the hardware layer of the AI stack is therefore increasingly focused on doing less unnecessary work, shortening distances within systems, and making energy use more proportional to real demand. The combined effect is that each new generation of AI infrastructure delivers significantly more capability per unit of electricity than the last. The innovations described below highlight how the UK is well-positioned to take a leading role in the development of energy-efficient Green AI.

Photonics and Optical Interconnects: Reducing the Cost of Moving Data

As AI systems scale across many chips, the energy required to move data between processors and memory becomes a dominant driver of electricity use. Optical and photonic technologies help address this challenge by using light for data transmission, reducing the energy losses and heat associated with electrical interconnects, particularly at high bandwidths. Silicon photonics are especially important for chip-to-chip, on-package, and rack-scale communication, enabling far higher data throughput with lower power consumption than traditional copper-based links. While electronic processing remains essential, photonic interconnects significantly reduce the energy and thermal overhead of moving data, which is increasingly the dominant constraint in advanced compute systems.

In the longer term, research into on-chip photonics could allow some data movement and signalling to occur within the chip itself, potentially delivering further step-changes in efficiency by reducing electrical interconnect bottlenecks. The UK is actively supporting pioneering research and pilot production efforts in silicon and integrated photonics at universities such as Southampton[7] and Cambridge[8], including collaborative European pilot lines that aim to accelerate commercialisation of these technologies. By reducing the energy cost of data movement, photonic approaches allow large AI systems to scale without a proportional increase in electricity demand.

Neuromorphic Computation, Modelling Human Brain Efficiency

Neuromorphic computing is a broad interdisciplinary field that studies how biological brains process information and applies those principles to artificial computing systems. Rather than seeking to replicate the brain exactly, neuromorphic computation approaches abstract key characteristics of neural processing – such as event-driven signalling, massive parallelism, local memory and computation, and adaptive behaviour – and translates them into new models of computation. The core aim is to improve efficiency, scalability, and resilience by changing how computation is organised, not by simply increasing processing speed or shrinking transistors.

Computational (software–based) neuromorphic approaches

At the computational level, neuromorphic ideas can be explored without changing physical hardware. This includes neuromorphic algorithms such as spiking neural networks, event-based learning models, and brain-inspired software frameworks that run on conventional CPUs, GPUs, or FPGAs. These approaches are valuable for research, experimentation, and early deployment, and they help develop tools and skills. However, because they still rely on traditional chip architectures, their energy efficiency gains are limited compared with purpose-built hardware.

7 https://www.cam.ac.uk/research/news/cambridge-to-trial-cutting-edge-semiconductor-technologies-for-wider-use-in-major-european-project

8 https://www.southampton.ac.uk/research/institutes-centres/silicon-photonics

Physical neuromorphic architectures and circuitry

In its purest form, neuromorphic computing involves fundamental changes to chip architecture and circuitry. Instead of clocked digital logic and a strict separation between processor and memory, neuromorphic chips use event-driven, asynchronous designs in which computation is distributed across many simple processing elements. Memory is co-located with computation in synapse-like structures, reducing data movement. Many designs employ analogue or mixed-signal circuits to model neural dynamics efficiently. Embedding neuromorphic principles directly in silicon delivers the greatest potential gains, particularly for low-power inference, perception, and adaptive workloads.

Conventional computing consumes large amounts of energy through constant clocking (the use of a global timing signal to synchronise operations) and the movement of data between memory and processors. Using the techniques outlined above, neuromorphic systems reduce the energy overhead, making them especially attractive for workloads where energy efficiency, not raw throughput, is the primary constraint.

The UK is a recognised international leader in neuromorphic research, with a strong emphasis on energy efficiency. The University of Manchester's SpiNNaker[9] programme is a flagship example, demonstrating large-scale, low-power, event-driven architectures inspired by neural processing. The University of Cambridge[10] leads globally in neuromorphic algorithms, hardware–software co-design, and emerging low-energy devices. Collaborative work at the Multidisciplinary Centre for Neuromorphic Computing[11] includes institutions such as Aston, Cambridge, Oxford, Southampton, Strathclyde, and Loughborough. Supported by UKRI and EPSRC, this ecosystem positions the UK at the forefront of developing neuromorphic technologies as a long-term pathway to more energy-efficient computing.

Specialised Accelerators and Reduced–precision Computing: Doing Only the Work That Matters

A major source of energy efficiency in AI systems comes from moving beyond general-purpose CPUs towards specialised accelerators. An accelerator is a processor designed to execute a narrow class of operations far more efficiently than a general-purpose chip. They dramatically reduce the energy spent on unnecessary instruction handling and data movement, which are major sources of power consumption in conventional computing.

Accelerators require lower-precision arithmetic, which requires fewer transistor switches, reduces memory bandwidth and allows more operations to be performed per unit of energy, significantly lowering electricity consumption across both training and inference.

A leading UK example is Graphcore, which has developed the Intelligence Processing Unit (IPU)[12], a purpose-built AI accelerator designed around fine-grained parallelism and on-chip memory. By keeping entire models close to the compute units and avoiding energy-intensive data transfers, IPUs aim to deliver substantially higher performance per watt for AI workloads compared with general-purpose processors, reducing the energy intensity of advanced AI systems.

9 https://www.scieng.manchester.ac.uk/tomorrowlabs/spinnaker/

10 https://www.enterprise.cam.ac.uk/opportunities/a-new-class-of-material-for-brain-inspired-computing/

11 https://www.uk-neuromorphic-centre.net/

12 https://www.graphcore.ai/products/ipu

Memory Proximity, Advanced Packaging, and 3D Integration: Shortening Distances Inside Computing Systems

As previously mentioned, a significant share of modern AI systems' energy consumption comes not from computation itself but from moving data between processors and memory. *Memory proximity* refers to architectural approaches that place memory physically closer to compute units, reducing the distance data must travel. By shortening these paths, systems reduce latency, lower energy lost to signal transmission, and avoid repeated transfers between separate chips or boards. Keeping data 'near' computation is one of the most effective ways to improve performance per watt in AI workloads.

Advanced packaging and 3D integration are closely related approaches that improve efficiency by stacking or connecting chips in more physically efficient ways. Instead of spreading components across a circuit board, these techniques bring compute, memory, and accelerators together within a single package or stack them vertically, dramatically shortening the distance data must travel. By increasing bandwidth and reducing energy lost in data movement, they lower power consumption and heat generation. This system-level integration has become a critical lever for energy efficiency in AI hardware, delivering gains that are increasingly difficult to achieve through transistor scaling alone.

The UK has growing strength in system-level semiconductor design, particularly through its academic and translational research base. The Compound Semiconductor Applications Catapult (CSAC)[13] is developing national capabilities in advanced packaging, heterogeneous integration, and chiplet-based systems, supporting both research and commercialisation. Universities such as Bristol[14] and Southampton[15] are internationally recognised for work on 3D integration, interconnects, and memory–compute co-design. The UK plays a leading role in energy-efficient AI hardware by focusing not just on chip design in isolation, but on how entire systems are physically integrated.

Where AI Energy is Really Lost: Advanced Semiconductors Beyond the Processor

The energy footprint of AI systems is shaped not only by how efficiently chips perform computation, but also by how electricity is converted, conditioned, and delivered to those chips. Advanced compound semiconductors, such as gallium nitride (GaN), silicon carbide (SiC), and indium phosphide (InP), play a critical role in this wider system by enabling far more efficient and high-frequency components than conventional silicon.

In power delivery, GaN and SiC devices switch electricity faster and with lower losses, allowing power supplies, voltage regulators, and inverters to operate at higher efficiency and smaller size. This reduces the energy lost as heat before electricity even reaches

13 https://csa.catapult.org.uk/expertise/advanced-packaging/

14 https://www.bristol.ac.uk/cabot/news/2024/semiconductor-centre.html

15 https://www.southampton.ac.uk/news/2025/06/new-centre-to-help-uk-keep-pace-with-future-advances-in-electronics.page

processors, lowering both overall electricity consumption and cooling demand. In parallel, compound semiconductors such as InP underpin high-performance photonic and radio-frequency components used in optical interconnects, which further reduce energy losses in data movement (as discussed above). In massive data centre racks, even small percentage improvements in power conversion efficiency translate into substantial absolute energy savings.

South Wales compound semiconductor cluster

The UK has a globally recognised strength in compound semiconductors centred in South Wales, spanning research, materials, device design, and manufacturing. Anchored by institutions such as Cardiff University[16], Swansea University[17], and the Compound Semiconductor Applications Catapult[18], this cluster supports the development of advanced GaN, SiC, and photonic devices for power electronics and high-frequency applications. These capabilities directly underpin more efficient power delivery, cooling reduction, and optical connectivity for AI infrastructure.

Next–generation Chip Cooling: Unlocking Energy–efficient AI Performance

As AI chips grow ever more powerful, the heat they generate becomes a major barrier to performance and energy efficiency. Traditional air-based fans and heat sinks are reaching their limits, so the industry is increasingly turning to liquid-based and microfluidic cooling innovations that can extract heat much closer to the point where it is generated. Techniques such as direct liquid cooling, immersion cooling, and microfluidic channels etched into or around chips dramatically shorten the distance heat must travel to be removed, enabling much higher heat transfer efficiency than air alone. In some experimental systems, routing coolant through microscopic channels or specialised cold plates has been shown to reduce peak temperatures by *up to three times* compared with conventional methods, allowing chips to run faster with lower energy spent on cooling infrastructure.[19]

Some UK businesses and research institutions are pioneering this space. Iceotope,[20] a UK-based company, has established a dedicated liquid-cooling lab in Sheffield to develop and test high-density coolant systems tailored for modern servers, helping to reduce power use and carbon emissions in data centres. Academic spin-outs such as Dew Point Systems, emerging from the University of Hull,[21] are developing next-generation cooling solutions that aim to cut electricity costs and emissions by dramatic margins. These efforts, paired with global advances in microfluidics and immersion cooling, reflect a shift toward cooling systems designed to match the thermal complexity of next-generation AI chips, unlocking higher performance with lower energy overhead.

16 https://www.cardiff.ac.uk/compound-semiconductors
17 https://www.swansea.ac.uk/engineering/research/compound-semiconductor-centre/
18 https://csa.catapult.org.uk/
19 https://datacentrenews.uk/story/corintis-raises-usd-24m-for-ai-chip-cooling-with-microsoft
20 https://www.iceotope.com/
21 https://www.hull.ac.uk/news-and-events/news/new-university-of-hull-spin-out-to-transform-data-centre-industry-with-lower-costs-and-emissions

Local Intelligence: Reducing Demand for Centralised Compute

Finally, not all efficiency gains come from making hardware better; some come from avoiding unnecessary centralised computation altogether. Low-power chips deployed at the edge can analyse and filter tasks, making decisions to send on the most complex tasks to the central systems, while routing simpler tasks to more basic, less energy-intensive compute. This reduces unnecessary data transmission, storage, and large-scale compute requirements, lowering overall electricity demand for AI-enabled services.

3. Energy Efficiency Across the Full Lifecycle of Data Centres

Innovation in AI infrastructure increasingly extends beyond chips and cooling into the physical fabric of data centres themselves. In the UK, advances in construction methods, materials science, structural engineering, power distribution, and digital maintenance are contributing to lower embodied carbon, reduced electricity demand, and more efficient long-term operation.

Low–carbon Materials and Structurally Efficient Construction

A growing focus of innovation is reducing the embodied carbon of data-centre buildings, particularly through more efficient use of concrete and steel. Data centres have historically been over-engineered to accommodate vibration, load, and future expansion, often resulting in excessive material use.

The UK is pioneering research in a number of these fields, with structural engineering consultancies often harnessing these innovations to actively embed low-carbon concrete, material optimisation, and whole-life carbon assessment into the design of large, energy-intensive buildings, practices that are directly applicable to hyperscale and colocation data centres.

Firms such as Stantec UK[22] explicitly incorporate embodied-carbon modelling and structural optimisation to reduce material volumes while maintaining performance and resilience, aligning projects with UK Net Zero building standards. This industry practice is underpinned by academic research leadership, notably at University College London,[23] where large-scale work on low-carbon concrete floor systems and structural components demonstrates how emissions can be significantly reduced without compromising load-bearing requirements. Complementary research at the University of Leeds[24] focuses on next-generation cement and concrete formulations that lower CO_2 emissions in critical structural materials.

22 https://www.stantec.com/uk/services/structural-engineering

23 https://www.ucl.ac.uk/engineering/news/2023/oct/ucls-low-carbon-concrete-research-breaks-new-ground-sustainable-building

24 https://eps.leeds.ac.uk/collaborations-impact/doc/research-spotlight-making-concrete-greener

The UK has deep expertise in engineering, materials science, and construction, which can be brought together to construct much more efficient data centre buildings. Energy-efficient data centre design and build is emerging as a new and growing sector within industrial engineering and will be a key component of attempts to deliver the promise of Green AI.

Optimising Floor Design to Minimise Vibration and Concrete Volume

Floor vibration control is a critical but often overlooked constraint in data-centre design. Traditional approaches rely on over-engineering heavier structures, which increases both embodied carbon and material use. New approaches focus on active vibration control, allowing lighter, more material-efficient structures while maintaining the stringent vibration tolerances required for sensitive equipment.

Based on engineering research from the University of Exeter, CALMFLOOR[25] is pioneering next-generation floor vibration control using patented active mass damper technology. Rather than relying on heavy structural overbuild, CALMFLOOR's system actively counteracts vibration, enabling high-performance, lightweight floor designs. This allows data centres to use less concrete in their construction.

Air Handling, Ventilation, and Building–level Efficiency

Beyond cooling technology itself, innovations in airflow management and ventilation design reduce electricity consumption by ensuring energy is not wasted moving or conditioning air unnecessarily. Advances include better hot and cold-aisle containment, optimised airflow paths, improved sealing, and dynamic control of ventilation based on real-time conditions. These approaches reduce fan energy use and support more efficient cooling strategies overall.

Building services engineering firms in the UK are increasingly deploying computational airflow modelling, digital simulation, and sensor-driven ventilation control in data centres, building on capabilities originally developed for complex commercial buildings. Firms such as Arup[26] and Buro Happold[27] routinely use computational fluid dynamics (CFD) to model airflow, temperature gradients, and pressure differentials in high-intensity buildings, integrating sensor data and real-time monitoring to dynamically manage ventilation and energy use. UK-based building services specialists, such as Hoare Lea,[28] combine environmental modelling with smart building management systems to reduce fan energy and avoid over-conditioning. Together, these approaches enable ventilation systems to dynamically respond to real operating conditions rather than conservative design assumptions, lowering electricity consumption and supporting more efficient AI-ready facilities.

25 https://calmfloor.com/market/new-build-vibration-control/
26 https://www.arup.com/services/computational-fluid-dynamics/
27 https://www.burohappold.com/expertise/building-physics/
28 https://hoarelea.com/what-we-do/building-services-engineering/

Power Distribution and Electrical Efficiency Within Buildings

Electricity losses within a data centre do not stop at the grid connection. Power conversion, backup systems, and internal distribution all affect total electricity demand. UK power-engineering firms and research groups are contributing to next-generation uninterruptible power supply (UPS) systems and resilient power architectures that improve electrical efficiency while maintaining high uptime. Concurrently, UK technology companies are developing smart power monitoring and control systems that help data centres detect and eliminate hidden inefficiencies. Each avoided conversion step reduces wasted energy and the associated cooling load.

For example, Piller UK[29], a specialist in high-performance UPS and power conditioning solutions, delivers systems designed to reduce losses and improve power quality for critical infrastructure. In the research sector, the Electrical Power Research Group at Cardiff University[30] works on advanced power electronics, control strategies, and resilient grid-connected power systems that can be applied to critical facility architectures. On the monitoring and controls side, UK firms such as Powerstar Energy[31] integrate real-time power analytics with automated control to optimise energy flows and reduce waste.

Using Data to Predict and Optimise Data Centre Maintenance

Data-centre efficiency increasingly depends on how facilities are operated and maintained over time. UK strengths in digital engineering and data analytics are enabling digital twins of physical infrastructure, combining dense sensor networks with predictive analytics and machine-learning models to monitor performance, identify emerging faults, and optimise energy use in real time. Predictive maintenance allows operators to detect failing components, airflow obstructions, or inefficient power paths before they cause excess energy consumption. This enables facilities to operate closer to optimal conditions rather than relying on conservative safety margins.

Providers such as SWG UK[32] offer IoT-enabled smart building platforms that connect real-time sensor data from mechanical, electrical, and environmental systems to advanced analytics engines to flag anomalies and support predictive maintenance across assets. Specialist systems integrators such as Voltix Services[33] deploy wireless sensor networks that feed live data into analytics dashboards, enabling early fault detection and condition-based maintenance. Building services and smart-facility technology firms like Millbeck[34] implement integrated IoT and building management system (BMS) solutions to monitor equipment health and energy use, with predictive insights used to reduce downtime and inefficiencies. On the research side, the

29 https://www.piller.com/uk-en/products/ups-systems/
30 https://www.cardiff.ac.uk/engineering/research/groups/electrical-power
31 https://powerstar.co.uk/
32 https://www.swg.com/products/iot-smart-building/
33 https://voltixservices.co.uk/energy-management/live-building-monitoring/
34 https://millbeck.co.uk/smart-buildings/

Cambridge Centre for Smart Infrastructure and Construction[35] (CSIC) develops advanced sensor technologies and data models to enable smarter whole-life asset management, combining high-resolution sensing with data analytics to support predictive maintenance practices across infrastructure lifecycles.

The UK Hosts a Diverse and Rapidly Developing Innovation Ecosystem

These examples demonstrate the rich mixture of companies and researchers across the UK who are pioneering and scaling solutions that will drive more energy efficiency in data centres and beyond. These physical innovations compound the gains made at the chip and system level, ensuring that improvements in compute efficiency are not undermined by inefficient buildings. As AI deployment accelerates, this whole-system approach will be essential to ensure growth in AI capability does not lead to unsustainable growth in electricity consumption. The innovative UK expertise outlined above shows how it is uniquely well-placed to strike that balance and thus lead in the development of Green AI.

4. Models, Applications, and Software: Reducing Energy Demand Through Smarter Use of Compute

This section focuses on the model and software layers of the AI stack, where decisions about *what* computation is performed, when, and *how often* increasingly determine the electricity required to deliver AI-enabled services.

In this context, a ‘model’ refers to the trained mathematical system that performs tasks such as prediction, classification, generation, or decision-making. An ‘application’ refers to real products or services that are built using one or more models, combined with data and processes in a scalable format. While models determine the raw computational cost of a task, applications largely determine whether that cost is incurred efficiently or wastefully.

Innovation at this layer is therefore less about maximising raw capability and more about minimising unnecessary computation. Advances in model design, training techniques, and deployment software are enabling AI systems to deliver comparable outcomes more efficiently.

Crucially, these efficiencies have the greatest impact on ‘inference’, the day-to-day running of AI systems, often the response to a user prompt, which typically dominates lifetime energy consumption. Software patterns that reduce repeated processing, reuse results, or route tasks to appropriately sized models can materially lower electricity demand without reducing service quality.

35 https://www-smartinfrastructure.eng.cam.ac.uk/

The UK is well-positioned to lead innovation in the 'application layer', with strengths in computer science, applied machine learning, software engineering, and systems research, combined with a strong ecosystem of AI-native companies. This makes it particularly well-suited to innovation focused on efficiency rather than scale alone. UK universities and research institutes are advancing techniques for smaller, more capable models, while UK businesses are translating these methods into commercial applications where cost, reliability, and energy use matter as much as headline performance.

As AI adoption expands beyond frontier research into widespread economic and public-sector use, the ability to deliver energy-proportional AI through models and software will be as important as advances in hardware.

Building Smaller, More Efficient Models

A central innovation trend is turning 'frontier-scale' models into smaller, cheaper models that deliver most of the value for specific, daily tasks. It should be noted that the 'training' involved in creating a large language model (LLM) requires immense compute, and therefore energy. However, once the models have been trained, daily requirements for 'inference' require much less compute and energy. Because of the cost, training often happens where energy is cheap (e.g. near solar farms in the desert or hydro-electricity plants in the mountains), while inference is located closer to where humans live and work.

Techniques such as 'knowledge distillation' transfer capabilities from a large model into a smaller one, often confining the heavy computation to a one-off training phase and producing a model that is cheaper and faster to run at scale. The Government Digital Service[36] has produced a useful overview explaining distillation explicitly in terms of producing "lightweight, energy-efficient" models for production deployment.

UK universities are also pushing the frontier of efficiency-oriented modelling: researchers at the University of Edinburgh reported methods that compress the memory footprint used by AI models, with the aim of improving practicality and reducing resource requirements for deployment.[37] Meanwhile, the University of Southampton is explicitly framing research projects around 'energy-efficient generative AI models', including compression and system-level optimisation to reduce model sizes and accelerate inference.[38]

Dynamic and Adaptive Computation

Another major direction in model efficiency is enabling systems to selectively activate only the components they need, rather than executing an entire network for every request. This logic underpins approaches such as sparsity, conditional computation, and mixture-of-experts architectures, where compute is spent only where it adds value.

36 https://www.gov.uk/government/publications/ai-insights/ai-insights-model-distillation-html

37 https://www.ed.ac.uk/news/shrinking-ai-memory-boosts-accuracy

38 https://www.southampton.ac.uk/study/postgraduate-research/projects/sustainable-generative-ai-models-0

When deployed effectively, these techniques can significantly reduce the electricity required per query, particularly at inference time. UK universities such as Oxford,[39] UCL,[40] and Edinburgh[41] host active research programmes in efficient and adaptive machine learning, covering sparsity, conditional execution, and model–systems co-design. UK-rooted AI hardware and systems companies, including Graphcore,[42] have also contributed to this direction through research and tooling aimed at exploiting sparsity and selective execution in practice, highlighting how model-level efficiency and hardware-aware software design increasingly reinforce one another.

In parallel, 'adaptive computation' is becoming more common in applications: systems learn when a request needs heavyweight reasoning and when a lightweight pathway will do. These software patterns matter because they shift organisations away from a default of running the biggest model on every task. This is analogous to the filtering and routing of tasks that we saw at the chip level.

New Model Families and Alternative Approaches: Efficient AI Beyond Neural Scaling

Efficiency is not only coming from optimising the dominant LLMs; it is also emerging from developing alternative model families that can be more frugal and specialised for certain problem types. A recent example is Literal Labs,[43] a spin-out from Newcastle University, which is commercialising 'logic-based' approaches (Tsetlin machines) explicitly positioned as faster and 52x less energy-intensive than conventional neural approaches for some workloads. This is important in an efficiency narrative because it widens the toolkit: for many industrial tasks, the most efficient solution is often not a bigger foundation model, but a method matched to the structure of the problem.

Application–layer Efficiency: Product Design That Reduces Compute

At the application layer, efficiency gains increasingly come from how AI products are designed and served, rather than simply from which model is selected. A common and effective pattern is the use of retrieval-augmented generation (RAG) workflows.

RAG works by keeping relevant information close to where an AI system is used, allowing it to quickly pull in precise, trusted data when needed rather than searching broadly or relying on distant models, which makes responses faster, more accurate, and far more energy-efficient.

Across sectors, companies are applying RAG in distinct ways to deliver more efficient, reliable, and scalable solutions for their clients. In retail, IgniteAI Partners[44] enables faster, data-driven decision-making by improving how commercial data is analysed,

39 https://www.cs.ox.ac.uk/research/ai_ml/

40 https://discovery.ucl.ac.uk/id/eprint/10180580/

41 https://www.research.ed.ac.uk/en/organisations/institute-for-adaptive-and-neural-computation/

42 https://www.graphcore.ai/posts/training-sparse-large-scale-language-models-on-graphcores-ipu

43 https://www.ncl.ac.uk/press/articles/latest/2025/06/literallabs/

44 https://igniteaipartners.com/

while WeUno[45] connects models directly to structured business data to support accurate, traceable outputs for business-critical use cases. In capital-intensive industries, Method Grid[46] uses RAG to combine organisational knowledge with industry standards, embedding best practice into AI-powered playbooks for the delivery and assurance of major projects. Within highly regulated environments, Drutek[47] applies RAG to centralise pharmaceutical business knowledge, supporting the development of compliant and efficient solutions. Professional services firms benefit from Serpin's[48] use of RAG to accelerate document review, client research, and proposal development, while maintaining confidentiality and regulatory compliance. Beyond text-based applications, PropTexx[49] extends RAG principles to visual data, enabling the retrieval of similar room scenes and spatial layouts for the real-estate sector. AI Tappers[50] focuses on workforce productivity, using RAG-powered assistants that allow non-technical users to update internal data sources dynamically. More advanced architectures are also emerging, with Passion Labs[51] combining RAG and cache-augmented generation to support multi-agent systems, and Intent HQ[52] deploying agentic RAG to help brands identify high-performing audiences at scale.

Measuring Energy Performance at the Application Level is the First Step

UK businesses in applied AI are beginning to communicate energy and carbon efficiency as part of their product performance data. For example, Synthesia has published metrics about the carbon footprint of generating video with its platform,[53] positioning this as materially more efficient than traditional production methods and providing a concrete way for organisations to think about the footprint of a specific AI-enabled application. This is also a great example of identifying a damaging myth about how much energy is being consumed by generative AI products, and then using data to dispel and disprove this myth. While the underlying models remain complex, this kind of measurement culture matters: it pushes the market toward optimising end-to-end workflows (generation, storage, distribution, re-use) rather than focusing solely on model size.

5. Innovation in the Local Community

The Global Context

Globally, the integration of data centres into local communities is emerging as a strategic issue rather than a purely technical one. The leading activity today is concentrated in countries that combine large-scale digital infrastructure deployment

45 www.weuno.com
46 https://methodgrid.com/
47 https://drutek.com/
48 https://serpin.ai/
49 https://www.proptexx.com/
50 https://www.aitappers.com/
51 https://www.passionlabs.ai/
52 https://intenthq.com/
53 https://www.synthesia.io/post/how-much-energy-does-it-take-to-make-a-corporate-video-with-ai

with strong coordination between energy systems, planning, and industrial policy.

The UK does not currently compete in scale of deployment or in nationally mandated integration models. However, it occupies a distinct and potentially influential position: as a country with dense urban environments, strong local governance structures, world-leading universities, and growing pressure on energy and planning systems, the UK is well-placed to lead in coordination-heavy, systems-level innovation rather than volume-led build-out. This experience and emerging capability could become a key national export, an area of UK expertise that is closely linked to the developing Green AI agenda.

The UK Can Lead in Coordination and Integration

The areas where the UK has the strongest opportunity to lead are not in building the largest data centres, but in how infrastructure is deployed, shared, and embedded locally. These include modular and incremental data centre deployment aligned with real demand; edge inference and local compute for public services and regulated industries; integration of data centres into local energy systems, including heat reuse; and governance models that bring together councils, combined authorities, universities, and infrastructure operators early in the development process. These are precisely the areas where technical capability must be matched with planning, institutional coordination, and public trust – domains in which the UK has comparative strengths. We have already explored those strengths in Chapter 1 and will examine them again through the lens of ecosystems in Chapter 4.

In addition, the UK's research base in distributed systems, communications, and smart infrastructure provides a foundation for innovation at this 'place layer' of the AI stack. UK universities and national research programmes are already contributing to the technical underpinnings of distributed cloud, edge computing, and smart energy systems, while UK companies are active in modular infrastructure, digital engineering, and local-scale deployment models. Taken together, this positions the UK to act as a testbed and exemplar for community-integrated AI infrastructure, which could eventually become a playbook (with commercial products) that could be exported to larger markets.

How can innovation at this 'place' layer be developed in practice through collaboration between industry, local government, combined authorities, and universities? This is where efficiency gains from hardware and software can be translated into locally beneficial, energy-efficient AI infrastructure that supports growth without imposing disproportionate costs on communities.

Efficiency in Integration and Implementation

As AI infrastructure expands, the next frontier of efficiency is not only what is built, but *how it is embedded into the places it serves*. At this layer of the stack, innovation focuses on integrating compute facilities into local energy systems, digital networks, and public-service ecosystems so that communities gain tangible benefits: lower energy waste, improved resilience, and shared economic value.

As we have seen in Chapter 1, modern AI infrastructure is becoming more distributed and more modular. Rather than concentrating all compute in a small number of very large sites, the system is increasingly complemented by modular data centres and edge inference located closer to demand, connected by high-capacity networks and operated through shared standards and partnerships. This approach can reduce electricity demand in three ways: by right-sizing compute to real usage (reducing over-provisioning), reducing data movement (especially for latency-sensitive or data-heavy applications), and capturing value from unavoidable heat and power flows at the local level.

Frugal by Design

Frugal AI focuses on delivering useful outcomes with minimal resources. However, this is not a constraint driven compromise. It is a design philosophy that prioritises efficiency, accessibility, and resilience. The UK is leading the way in developing frugal AI, moving from research to practical applications. Cambridge University launched the Frugal AI Hub in November 2025, bringing together multi-disciplinary researchers in the Judge Business School.[54]

Frugal AI offers the UK several advantages. It lowers operating costs and reduces dependence on scarce hardware, thereby enabling greater adoption by SMEs and public services. It also creates exportable solutions for partners facing similar constraints, particularly in the global South.

Build–as–you–grow, Modular Data Centres

Modular, prefabricated components are a key innovation for integrating infrastructure into local areas without overbuilding. For councils, combined authorities, universities, hospitals, and other anchor institutions, modularity enables staged deployment: capacity can be added when demand materialises, rather than constructing a large facility years in advance. Most importantly, modular solutions can be rapidly deployed with relatively straightforward planning, and they can be easily dialled up and down, as required.

Cannon Technologies[55] is an example of a UK company that develops modular data centres and ‘build as you grow’ modular solutions aimed at faster, scalable deployment across sectors including education and government estates. This kind of approach becomes especially relevant for local and regional strategies where the objective is to provide dependable compute close to public services and local industry without locking communities into oversized, underutilised buildings.

Edge Inference and Local Compute for Public Services and Industry

Edge inference means running AI models locally, close to where the data is created, rather than sending that data to a distant data centre or cloud service for processing. By analysing data on nearby devices or small local servers – such as on a factory floor, in a hospital, or within a city network – edge inference reduces the need to move large amounts of data, cuts response times and lowers energy use. It allows AI systems to make faster, more efficient decisions using smaller, task-specific models, while only relying on central data centres when heavier processing is genuinely needed.

---

54 https://www.jbs.cam.ac.uk/2025/frugal-ai-hub-at-cambridge-judge-supports-sustainability/

55 https://cannontech.co.uk/modular-data-centre-halls/

Across the UK, local and regional innovation programmes are already trialling these ideas. The West of England Combined Authority's 5G Logistics project[56] specifically trialled mobile edge computing (MEC) as an alternative to relying solely on cloud, demonstrating how local edge capability, combined with advanced connectivity, can support more efficient and responsive systems for complex logistics environments.

Networks and Distributed Cloud: Making Locality Practical

Integrating data centres and local AI compute into communities depends on how digital networks and computing infrastructure are designed and shared, not just on having faster Internet connections. Increasingly, computing capacity is being distributed across multiple smaller, local facilities rather than being concentrated in a few very large sites. This allows councils, universities, public services, and businesses to use shared local computing infrastructure, such as modular data centres or edge facilities, without each organisation building its own. Crucially, the infrastructure is shared, not the data: each organisation's data and applications remain separated and securely controlled, in the same way that different organisations safely share commercial cloud services today. By enabling secure, multi-user local infrastructure, this approach reduces duplication, lowers overall energy use, and makes it easier to deploy AI systems that are tailored to local needs while maintaining strong standards of data security and privacy.

UK research programmes are explicitly oriented toward this direction. The Communications Hub for Empowering Distributed Cloud Computing Applications and Research (CHEDDAR)[57] is a UK research hub focused on advancing communications and distributed computing capabilities, helping underpin the technical foundations required for reliable local compute and edge deployment.

Policy Gap: From Wireless Ambition to Local Compute

The Wireless Infrastructure Strategy (2023) established important national ambitions for advanced connectivity and future telecoms. Elements of this agenda have since been absorbed into the UK's 10-Year Infrastructure Strategy,[58] helping embed digital connectivity within wider infrastructure planning.

However, significant gaps remain. Current policy is connectivity-focused rather than compute-aware, with limited consideration of how wireless networks enable edge computing, distributed cloud, and local AI inference – all central to reducing energy use and integrating AI infrastructure into communities. There is also a disconnect between national ambition and local delivery, with councils and combined authorities lacking clear guidance on how digital connectivity, local compute, energy systems, and planning should be coordinated.

The absence of a clear implementation framework – linking funding, regulation, and institutional responsibility – risks slowing deployment, and the UK could miss the opportunity to lead in place-based, energy-efficient digital infrastructure, which would undermine efforts to achieve global leadership in Green AI.

---

56 https://www.westofengland-ca.gov.uk/what-we-do/innovation/5g-logistics

57 https://cheddarhub.org/

58 https://www.gov.uk/government/publications/uk-infrastructure-a-10-year-strategy

The Government should, therefore, focus on aligning wireless policy with distributed compute, supporting local delivery through clearer guidance and coordination, and moving decisively from ambition to implementation.

Turning Heat Waste Into a Community Asset

Even with major efficiency gains, AI infrastructure produces heat. Integrating data centres into local communities increasingly means planning for heat offtake and heat networks, so that energy that would otherwise be lost into the air can be used to displace local heat demand. This does not always reduce a data centre's electricity draw directly, but it improves overall system efficiency and can materially reduce local emissions and energy costs.

Queen Mary University in London[59] is reusing waste heat from its on-campus data centre to provide heating and hot water to nearby buildings, demonstrating how institutional campuses can integrate compute and heat demand in a single operational system.

In London, national and city-level work on integrating data centre infrastructure with local energy systems has moved from concept to active development. In March 2025, the Old Oak and Park Royal Development Corporation (OPDC) appointed Hemiko[60] as the development, funding, and delivery partner for the Old Oak & Park Royal Energy Network (OPEN), the UK's first district heat network designed to capture waste heat from local data centres and supply it to homes, businesses, and community facilities.

Heat re-use can also be harnessed at a more local level: companies such as Edge Synergies[61] are pioneering the use of micro data centres within the boiler rooms of commercial real estate to repurpose waste heat into the local community. Local councils and combined authorities are uniquely positioned to benefit from these micro data centres, which can be deployed across council-owned property, saving energy costs and enabling them to benefit from new revenue streams from the customers of these micro data centres.

As we have seen in Chapter 1, the governance model is critical: heat reuse requires early coordination between developers, heat network operators, and local authorities so that pipe routes, commercial arrangements, and planning conditions are all aligned before a site is built out.

---

59 https://www.qmul.ac.uk/media/news/2024/se/queen-mary-university-of-london-data-centre-waste-heat-to-provide-hot-water-and-heating-for-campus.html

60 https://www.london.gov.uk/opdc-announces-hemiko-development-and-funding-partner-innovative-new-heat-network

61 https://edgesynergies.com/about-us

Policy Recommendations

Quick Impact Priorities (0–24 months)

1 Make Energy-Efficient Data Centres the Default for the Public Sector

Public-sector demand can rapidly shift the market toward lower-energy AI infrastructure. The biggest efficiency gains come from whole-system optimisation covering hardware, cooling, power delivery, building design, and operations, rather than isolated upgrades. Using procurement to reward this integrated approach enables immediate progress without waiting for new build cycles.

Minimum efficiency and technology-optimisation standards should be embedded into public-sector procurement frameworks, with leadership from DSIT, DESNZ, and the Cabinet Office. Independent audits would score data centres against these criteria, and public bodies would be required to use only facilities that meet minimum thresholds. To support transition, a targeted retrofit fund would enable existing government-serving data centres to upgrade using UK-developed technologies, delivered by industry with academic support for evaluation and optimisation. The allocation of funding would be proportionate and prioritise projects with the greatest size and intensity, which would deliver the greatest post-retrofit efficiencies. Our recommendation is to start within the public sector, where it is easier to set criteria and standards, before applying these same principles across the private sector.

2 Use Transparency and Measurement to Drive Continuous Efficiency Gains

Measurement is a prerequisite for optimisation rather than an administrative burden. It demonstrates that once energy use is measured consistently across compute, cooling, power, and operations, organisations can move away from conservative over-provisioning and towards performance-led optimisation. Transparency also shifts competition towards performance-per-watt rather than raw capacity.

Common metrics for data-centre energy use and efficiency ratios should be established at national level and required for facilities serving public workloads. Results would be published through a benchmarking framework that highlights improvement over time rather than simple compliance. Operators gain clearer signals on where investment delivers value, while researchers refine metrics and methodologies as technologies and operating models evolve. UKAI would work with other trade bodies and regulators to form an independent body to review and maintain the metrics, benchmarking, and evaluation framework.

3 Win in the Application Layer and Frugal AI, Not Compute Scale

The most immediate reductions in AI energy use often come from avoiding unnecessary computation altogether. The UK's strengths are in smaller models, adaptive computation, retrieval-based systems, and alternative approaches that deliver useful outcomes with far lower energy demand. Frugal AI also enables adoption by public services, SMEs, and organisations operating in resource-constrained environments.

National AI priorities should explicitly favour efficiency-oriented models and applications, with DSIT and UKRI aligning funding toward compression, sparsity, adaptive computation, and alternative model families. Universities can advance the underlying methods, while industry embeds them into production systems and communicates energy performance as a feature. Internationally, frugal AI can form a core part of the UK's AI-for-good positioning, building international partnerships to drive impact, particularly with the global South.

4 Create a National Showcase for Energy-efficient AI Innovation

The UK hosts a rich ecosystem of energy-efficient AI innovation across hardware, software, and infrastructure. However, adoption is often slowed by limited visibility among public-sector buyers and private investors. A credible national showcase can turn technical capability into deployment momentum.

UKAI will work with DSIT and relevant departments to convene a national showcase focused on real-world case studies rather than concepts. Innovative UK businesses demonstrate deployed solutions – spanning efficient compute, data centres, applications, and local integration – to senior government leaders, public-sector decision-makers, and private investors. Academia provides independent credibility and evaluation, while industry leads demonstrations. The Government's role could be to link the showcase to procurement programmes, enabling follow-on funding or pilots. The showcase becomes a practical bridge between innovation and scaled adoption, a theme we will explore more in Chapter 4, where we examine the role of ecosystems.

Structural Priorities (2–4 years)

5 Back UK-leading Efficiency-driven Compute Innovation

The UK's comparative advantage in AI hardware lies in reducing energy demand per unit of compute rather than increasing scale. The UK has established leadership in photonics, compound semiconductors, neuromorphic computing, advanced packaging, and next-generation cooling, technologies that reduce energy consumption and minimise loss.

Strategic prioritisation through DSIT and UKRI should focus existing R&D, pilot manufacturing, and funding on these fields. Universities can lead foundational research and skills development, while industry scales solutions through testbeds and early commercial deployment. Catapults and applied research centres bridge laboratory breakthroughs to market-ready systems, ensuring efficiency gains translate into deployed infrastructure.

6 Position the UK as a Global Leader in Efficient Data–centre Design and Integration

Future efficiency gains increasingly come from buildings, power systems, cooling, maintenance, and heat reuse, not just chips. The UK's strengths in engineering, construction, digital twins, and smart infrastructure create a pathway to global leadership in energy-efficient data-centre design and integration.

Demonstrator projects should integrate UK innovations across construction methods, cooling, power electronics, and operational optimisation, subject to planning coordination and innovation. Engineering firms and technology providers can package these capabilities into exportable design standards and services. Universities can underpin this with research on whole-life carbon, system integration, and digital maintenance. UK local government is well respected for planning, building standards, and regulation, while trade and diplomatic channels promote UK expertise as a reference model for efficient AI infrastructure worldwide.

7 Embed AI Infrastructure Into Communities Through Local Coordination and Shared Solutions

The next frontier of efficiency lies in how AI infrastructure is embedded into places. Modular data centres, edge compute, shared infrastructure, and heat reuse can reduce energy demand while delivering local economic and social value. The UK's dense cities and strong local governance create an advantage in this coordination-heavy approach.

Beacon councils and combined authorities should be supported to pilot modular data centres across public estates, shared local compute for public services, and systematic waste-heat reuse. Industry supplies modular, edge, and heat-integration solutions, while universities support system design, evaluation, and skills. National policy aligns planning, energy, and digital frameworks so successful local models can be replicated and scaled across regions.

8 Position Innovation, Not Scale, as the UK's Competitive Strategy in AI Infrastructure

The UK is structurally disadvantaged in competing on sheer scale of AI compute, but it is well positioned to lead in efficiency, integration and system design. Making this an explicit national strategy reframes constraints as advantages and aligns policy with the UK's real strengths.

This positioning should be embedded across the AI Opportunities Action Plan and future policies, with leadership from DSIT and alignment with HM Treasury. International engagement and standards diplomacy reinforce the approach by promoting efficiency, performance-per-watt and system integration as global norms. Industry and academia can then innovate toward a coherent national objective, avoiding fragmented attempts to compete on scale where the UK lacks structural advantage.

Competing on scale implies very high upfront capital expenditure, long payback periods, exposure to grid and planning constraints, and sustained pressure for subsidy or state-backed de-risking. An efficiency-led approach, by contrast, improves capital productivity by delivering more economic and public value per pound invested, while reducing the risk that the Government ends up underwriting oversized assets that become under-utilised or stranded as technology and demand evolve.

This can be made concrete through a focus on modular, energy-efficient, build-as-you-grow infrastructure that lowers initial capex, accelerates deployment, and limits balance-sheet risk for both public and private investors. At the same time, efficiency-driven advances at the hardware and application layers reduce operating costs and exposure to volatile energy prices, improving long-term cost certainty for public services. From a Treasury perspective, this shifts the AI infrastructure debate away from 'more spending' and towards better value for money: lower subsidy exposure, fewer contingent liabilities, and stronger fiscal discipline across the full asset lifecycle, while still supporting productivity growth and economic resilience.

Strategic Implications: Innovation as the Route to Sustainable AI Growth

This chapter demonstrates that innovation is already reshaping the energy and resource profile of AI, not through a single breakthrough, but through cumulative gains across hardware, software, and system design. Across chips, models, physical infrastructure, and deployment patterns, the direction of travel is clear: AI is becoming more selective, more specialised, and more energy-proportional. Increased capability no longer needs to translate directly into increased electricity demand.

A central implication is that efficiency is not a constraint on AI development but a source of competitive advantage. Innovations such as specialised accelerators, reduced-precision computing, adaptive models, edge inference, and modular infrastructure all point towards an AI ecosystem that delivers more value per unit of energy, capital, and water. This aligns closely with the UK's structural conditions: high energy prices, constrained grids, dense urban environments, and strong research depth. Rather than competing on scale alone, the UK is well-positioned to lead in designing Green AI systems that work efficiently under real-world constraints.

The chapter also highlights that innovation increasingly happens at the interfaces between layers: where hardware design meets software optimisation, where models are shaped by deployment context, and where physical infrastructure is integrated into local energy and digital systems. These gains are reinforced when innovation is coordinated between universities, industry, infrastructure operators, and local institutions, rather than pursued in isolation.

In that context, the Government's role is to create conditions where efficiency-oriented innovation can translate into deployment at scale. That includes aligning incentives with performance per watt rather than headline capacity, supporting measurement and transparency at the application level, and ensuring that infrastructure, planning, and skills systems are receptive to new, more modular, and distributed approaches. Innovation policy and infrastructure policy increasingly converge at this point.

Importantly, this chapter also clarifies a limit. Technical efficiency gains alone do not guarantee system-level outcomes. Without mechanisms to share infrastructure, coordinate deployment and avoid duplication, even highly efficient technologies risk being deployed inefficiently at scale. The next challenge is therefore not only how AI systems are built, but how they are organised, shared, and repeated across places and sectors.

This leads directly to the focus of the next chapter. Chapter 4 examines how efficiency-driven innovations can be embedded into scalable systems through clusters, shared infrastructure, common standards, and coordinated local ecosystems. Together, Chapters 3 and 4 frame a Green AI approach grounded not in restraint or subsidy, but in intelligent design: using innovation to reduce resource intensity and system-level coordination to ensure those gains are realised consistently as AI adoption expands.

Industry Perspective

"This report is an important contribution because it links AI growth to environmental, economic and social sustainability. It rightly shows that the future of AI depends on energy, infrastructure and social legitimacy working together. System constraints and fragmented delivery may threaten competitiveness but also climate progress and social buy-in, making its recommendations on joined-up planning and long-term resilience essential."

Alex Smith, CEO
FuturePlus

Chapter 4 – Scalable Systems

Fostering ecosystems that enable implementation of efficient AI at scale

1. Introduction: Systems Enable Scale and Knowledge Sharing

As the previous chapter has shown, the UK does not suffer from a shortage of ideas, innovation, or infrastructure ambition. It excels in world-class research, early-stage innovation, and pilot projects across AI, energy, and infrastructure. What it lacks, and what now matters most, are the systems that connect these elements together and allow them to scale. Without coherent systems, innovation remains fragmented, infrastructure under-utilised, and economic value captured elsewhere. Systems are the mechanism through which ideas become deployable solutions, innovations attract investment, and infrastructure delivers sustained benefits to citizens.

In this context, systems should be understood as more than technical platforms or organisational structures. They are the connective tissue that links policy, institutions, markets, capital, and delivery. Effective systems align incentives across public and private actors, reduce friction between sectors and create predictable pathways from experimentation to adoption. They allow learning to compound rather than reset with each new project, enabling knowledge, standards, and best practice to be shared rather than siloed. For a country facing systemic challenges, from energy resilience to productivity and public service reform, this connective function is critical.

Crucially, systems are what make scale possible. Investors do not fund isolated pilots; they fund credible pathways to customers, demand, and repeatable deployment. Businesses do not grow on innovation alone, but on access to markets, interoperable infrastructure, and trusted governance. Well-designed systems reduce risk by clarifying standards, aggregating demand, and aligning long-term objectives, making it easier to attract capital and harder for value to leak out of the UK economy. In doing so, they turn national ambition into something investable.

For citizens, the benefits of system-led growth are tangible. Systems enable infrastructure that works together rather than in isolation, public services that adopt technology safely and effectively, and economic growth that is more evenly distributed across regions. By connecting AI capabilities with energy networks, transport systems, and public institutions, systems help ensure that innovation delivers productivity gains, resilience, and improved quality of life rather than fragmentation or exclusion.

This chapter sets out how to build strong, integrated systems, often described as ecosystems. We need these in place to move the UK from innovation leadership to delivery at scale. After this introduction, the following sections explore how shared standards, open and federated approaches, test beds, procurement, skills, and investment frameworks should work together to create a compounding advantage. The chapter also analyses whether AI Growth Zones are functioning as ecosystems, describes how UK business interests can be represented and shows how ecosystems encourage investment in AI and energy. It also discusses how the development of Green AI in the UK can both open access to international markets and build consumer trust at home.

As this chapter argues, systems are, ultimately, not an abstract concept: they are the practical means by which the UK can translate innovation into world leadership in the field of Green AI, generating sustainable economic growth that will benefit both investors and citizens alike. Therefore, the chapter concludes with clear recommendations setting out how the UK's AI ecosystems can be cultivated to enable delivery at scale. But before turning to those factors, it is necessary to explore the importance of systems thinking at this crucial moment in the development of the UK's AI ecosystem.

Why Systems Thinking Matters Now

For more than a decade, the UK has excelled at generating innovation through individual projects, pilots, and centres of excellence. Yet too often, these initiatives remain isolated, struggling to scale beyond their original context. The result is a landscape rich in experimentation but poor in system-wide impact. As we saw in Chapter 1, many of the solutions that have been proposed by the Government have remained siloed within the jurisdiction of competing departments. Ecosystems matter now because the challenges the UK faces (decarbonisation, energy resilience, infrastructure renewal, and AI adoption) are systemic rather than sectoral.

AI, energy, and infrastructure must now be built together because each increasingly depends on the others. AI systems drive efficiency and optimisation across energy grids and transport networks, but also create new demands for power, cooling, and data infrastructure. Energy systems are becoming more decentralised and data-driven, relying on AI for forecasting, balancing, and resilience. Infrastructure planning increasingly depends on digital twins, predictive maintenance, and real-time data. Treating these domains separately creates inefficiency, risk, and missed opportunity.

As this report has consistently demonstrated, fragmentation remains one of the UK's biggest constraints. Responsibilities are split across departments, regulators, local authorities, and markets, each optimising for their own objectives. The problem is not that delivery is regional, but that national strategy does not yet provide a sufficiently clear mission, metrics, and interfaces to ensure regional innovation scales coherently rather than in parallel silos.

2. Shared Standards and Interoperability

Standards function as the operating system of ecosystems. Without shared technical, commercial, and governance standards, collaboration becomes bespoke, slow, and expensive. With them, innovation becomes interoperable, repeatable, and scalable.

Interoperability by design is particularly critical across compute, energy, data, and governance. AI systems must integrate with energy management platforms, infrastructure sensors, and public sector data while meeting safety, security, and sustainability requirements. Common standards allow organisations to plug into shared systems without surrendering control or duplicating effort.

The UK is unusually well placed to lead in this space. It has globally respected standards bodies, a strong regulatory tradition, and credibility as a trusted partner. Institutions such as the British Standards Institution (BSI) and the AI Security Institute (AISI), and sector regulators, provide a foundation for developing standards that balance innovation with trust. Importantly, the UK can use standards not only to manage risk domestically but also to shape global interoperability, working with other middle powers to avoid fragmentation between competing technological blocs.

The Commonwealth and other pan-global networks provide a natural platform for collaboration in this space. By co-developing skills, standards, and sector-specific models for health, energy, agriculture, and climate, the UK can support development while strengthening its own industrial base.

By convening discussion and collaboration around shared standards, the UK can position itself as an honest broker: enabling national control where necessary, while maintaining compatibility with international markets and systems. This role is particularly valuable in AI, where trust, safety, and sustainability are increasingly decisive factors for adoption.

3. Open and Distributed Technology Enabling Ecosystems

Shared resources are the raw materials for ecosystems

Open models and shared data infrastructure are increasingly central to healthy ecosystems. Open source software, open standards, and shared reference architectures reduce duplication, accelerate learning and lower barriers to entry for new participants. In the UK context, open approaches also align with public value, transparency, and long-term resilience.

Open-source libraries play a critical role in strengthening AI ecosystems by lowering barriers to entry and accelerating the diffusion of innovation. Platforms such as Hugging Face have become central infrastructure, enabling developers to access, adapt and deploy machine-learning components at low cost and with minimal friction. This has supported the rapid development and widespread adoption of capabilities such as security, robustness, and evaluation tools that might otherwise be confined to a small number of well-resourced organisations.

At the same time, large technology firms have increasingly engaged with open-source ecosystems. For example, Meta has released its Llama models under an 'open' licence, contributing to experimentation and downstream innovation. However, it is important to note that such releases are often only partially open: while model architectures and interfaces may be shared, the core training artefacts – including the underlying learning weights and data – remain closed. This distinction has important implications for transparency, competition, and long-term ecosystem resilience.

Federated learning enables efficient and safe data sharing

Federated learning is an approach to training AI models in which data remains within the organisation where it is generated. A shared model is sent to the organisation's servers, where it is trained locally and used to update the model's parameters. Only these updates are shared back, rather than the underlying data itself. This allows multiple parties to collaboratively improve AI systems without exposing sensitive or proprietary information.

Federated learning and other privacy-preserving data techniques are particularly important for AI deployment across sensitive and distributed systems, such as health, energy and transport. These approaches allow models to be trained across multiple organisations without centralising data, respecting sovereignty, security, and trust constraints. UK organisations working at the intersection of AI and critical systems, such as Mind Foundry[62] or Faculty[63], have demonstrated how advanced AI techniques can be applied responsibly in regulated environments.

62 https://www.mindfoundry.ai

63 https://faculty.ai

Balancing openness with intellectual property

Public digital infrastructure plays a crucial role in avoiding lock-in and fragmentation. Shared data platforms, identity systems, and interoperability layers allow multiple suppliers to innovate on top of common foundations. The challenge is designing openness without undermining investability. Ecosystems must strike a balance: core infrastructure and standards can be open, while allowing companies to build proprietary services and capture value at higher layers of the stack. As we have noted in previous chapters, there is an opportunity for the UK to continue to develop innovative products and services at the 'application' layer.

2. From Test Beds to Playbooks

Many of the ideas set out in this report are already being trialled by organisations across the UK through localised testbeds and living labs. Effective ecosystems should enable these initiatives to be socialised, with results and lessons shared for the common benefit of participants. In doing so, ecosystems help successful approaches to be identified, adapted and scaled.

As ecosystems accumulate case studies, methods, and evidence from real-world trials, this collective experience becomes a valuable source of knowledge for other organisations and regions. The challenge now is to translate this learning into accessible and reusable forms, such as shared playbooks or blueprints that can support wider adoption and replication.

The UK has a unique opportunity here. Its dense population, diverse regions, integrated infrastructure networks, and strong local governance create ideal conditions for experimentation at meaningful scale. Whether at a regional or national government level, the learnings should be actively shared. At an international level, the UK can itself be seen as a test bed – from smart energy systems to AI-enabled transport and health infrastructure – for solutions that are directly relevant to other advanced economies.

Government as a convenor of experimentation

Ecosystems enable learning to be shared, codified and reused, rather than reinvented in each new pilot. The Government has a critical role in enabling these regional ecosystems by encouraging investment and experimentation, by reducing regulatory uncertainty, sharing risk, and signalling long-term commitment. Devolved governments and combined authorities are equally important, acting as place-based convenors who can align planning, procurement, and local services. Initiatives such as innovation zones, AI growth zones, and regional digital twins point toward a more systematic approach.

Crucially, the UK can develop and export the playbook for Green AI deployment, based on making the UK a living lab to test and scale the ideas set out in this report.

5. Harnessing Procurement to Accelerate Ecosystems

Procurement is one of the most powerful but underused tools for ecosystem building. When coordinated effectively, it acts as both a market signal and a scaling mechanism, creating predictable demand for integrated solutions.

Aggregating demand across public sector bodies reduces risk for suppliers and investors, enabling them to invest with confidence in capability and capacity. Buying outcomes rather than technologies – an explicit direction of UK government policy – allows suppliers to innovate while remaining accountable for system performance. This is particularly important for AI-enabled infrastructure, where value lies in long-term efficiency, resilience, and sustainability rather than upfront technology deployment. We are pleased to note that central government has signalled an approach to procurement that favours 'consortia', bringing together different partners and combining innovative but small AI businesses with established, larger businesses with the resources to scale.

Shared procurement frameworks can also embed standards, interoperability, and sustainability requirements, reinforcing other elements of the ecosystem. Over time, this creates predictable markets for integrated solutions, encouraging competition on quality and performance rather than bespoke compliance.

6. Skills and Capabilities: Ecosystems as Engines of Knowledge

Ecosystems ultimately run on people and knowledge as much as on infrastructure. Indeed, knowledge is the most important raw material in any effective ecosystem, particularly where the goal is to deliver AI systems that are not only powerful, but energy-efficient and environmentally sustainable. In the context of AI, energy, and infrastructure, the challenge is not simply addressing skills shortages in individual sectors but developing hybrid capability – bringing individuals and teams who understand AI systems, energy consumption, infrastructure constraints, and optimisation, together. This hybrid capability is essential if the UK is to lead in Green AI, where efficiency, performance, and sustainability must be designed from the outset.

A core function of ecosystems is to link the skills that businesses need to the skills that schools, colleges, and universities are creating, particularly those required to improve energy efficiency across the AI lifecycle, from model design and training to deployment and operation. Apprenticeships, further education, and vocational pathways provide a major opportunity to build this capability at scale, including skills in power management, cooling, grid integration, systems optimisation, and energy-aware software engineering. However, these pathways must be dynamic and responsive, evolving in step with rapid advances in AI models, compute architectures, and low-carbon energy systems.

Mayoral and combined authority leadership can play a decisive role in aligning skills development with energy-efficient AI deployment. We welcome the creation of the Manchester Baccalaureate, under Andy Burnham's leadership, as an example of how place-based education reform can respond to local economic and technological priorities. This approach could be extended to include explicit pathways in digital, AI, and energy-efficient infrastructure skills. Over time, there is an opportunity for mayors and combined authorities to compete to develop the most relevant Green AI skills pipelines in their regions, before sharing lessons learned and best practice with other UK regions.

Ecosystems must also recognise the emergence of new technical and vocational roles created by energy-efficient AI infrastructure. Roles associated with data centres, grid connections, energy storage, cooling systems, heat reuse, and demand flexibility will be central to delivering Green AI in practice. These are high-quality, place-based jobs that can have a direct and visible impact on local communities. Where strong ecosystems exist, these roles are not only created but multiplied, supporting resilient supply chains and long-term regional employment linked to Net Zero objectives.

Universities should be pivotal partners in developing this capability, particularly through applied research into energy-efficient AI, system optimisation, and low-carbon infrastructure. Yet universities are often siloed or positioned at the margins of delivery. In part, this reflects funding models that are faculty-specific, limiting the time and incentives for collaboration. To unlock their full contribution, funding should be reoriented towards applied challenges that support Green AI ecosystems, with explicit encouragement for collaboration between and within universities, including those in the same city.

Public sector capability is equally critical. Delivering energy-efficient AI systems at scale requires public bodies to act as informed system integrators, able to balance performance, cost, energy use, and environmental impact. Ecosystems support this by embedding learning in long-term programmes and shared infrastructure, rather than relying on short-term projects or external consultancies that fragment knowledge and dilute accountability.

By enabling continuous learning and collaboration across AI, energy, infrastructure, and skills, ecosystems create the conditions for energy efficiency to become a competitive advantage rather than a constraint. They are among the most effective ways to spread best practice in Green AI, build trust across sectors and develop the next generation of leaders capable of managing complex, low-carbon digital systems. In doing so, ecosystems do not just support skills development, they also underpin the UK's ambition to become a global leader in energy-efficient, responsible, and sustainable AI.

7. Are AI Growth Zones Functioning as Ecosystems?

The Government launched the concept of AI Growth Zones in January 2025, as part of the AI Opportunities Action Plan, with the explicit ambition of creating regional ecosystems around AI infrastructure, compute, and investment. In principle, AI Growth Zones are a strong and widely supported idea. Many businesses across the UK AI sector recognise their potential to attract investment, anchor infrastructure and catalyse regional growth.

However, progress to date has raised questions about delivery, clarity, and depth. As of January 2026, 12 months after the publication of the AI Opportunities Action Plan, only four zones have been confirmed. This has led to growing uncertainty across industry about timelines, selection criteria, and the pace at which the remaining zones will be designated. For organisations considering long-term investment in infrastructure, skills, or partnerships, this lack of visibility risks undermining confidence and momentum.

More fundamentally, there is limited evidence that AI Growth Zones are yet operating as ecosystems in the sense described in this report. Through engagement with members and partners – including businesses, universities, and research institutions located within designated zones – many report little or no involvement in shaping the local vision, governance, or activity of the zones themselves. While high-level discussions appear to be taking place with major investors and large technology providers, smaller firms, researchers, and enabling organisations often feel disconnected from the process and unclear about how to participate.

This matters because effective AI and infrastructure ecosystems depend on precisely these actors, who bring skills, applied research, experimentation, supply-chain capability, and long-term regional presence. Without mechanisms for structured engagement, shared learning, and coordinated action, AI Growth Zones risk becoming narrowly defined investment sites rather than dynamic, resilient ecosystems. This needs to change; AI Growth Zones need to be rebooted.

To realise their full potential, AI Growth Zones will need greater clarity of purpose, faster execution, and a more inclusive multi-level approach to ecosystem-building. This includes transparent governance, clearer pathways for local organisations to engage, and active support for collaboration across infrastructure, energy, skills, and innovation. These are the conditions under which AI Growth Zones can move beyond designation and become engines of sustainable growth, and this is where UKAI can play a convening and enabling role.

Devolving Power to Enable Active AI Ecosystems

Mayors and combined authorities are uniquely positioned to drive the kind of rapid, place-based change required to build effective AI ecosystems. Unlike central government programmes, which often operate through national funding rounds and sequential decision-making, combined authorities bring together strategic control over transport, planning, skills, economic development, and, increasingly, energy and infrastructure. Many already possess the physical assets, institutional capacity, geographic footprint, and devolved budgets needed to act decisively.

Across the UK, however, combined authorities express growing frustration with the pace and ambition of national AI initiatives, including AI Growth Zones. Even where Growth Zones have been designated, local leaders report a piecemeal and siloed approach, with limited scope to shape delivery, align related investments or mobilise the wider ecosystem of businesses, universities, and public-sector partners. This risks constraining innovation precisely in those places best equipped to lead it.

As devolution deepens, combined authorities should be empowered not merely to participate in centrally defined programmes, but to actively design and govern AI ecosystems tailored to local strengths and challenges. This means giving mayors the authority to coordinate infrastructure planning, attract and anchor compute and data centre investment, align skills pipelines and convene local industry and research partners around shared objectives. Crucially, this should be supported by flexible funding and outcome-based accountability, enabling experimentation, iteration, and learning at pace.

At the same time, this empowerment must be grounded in evidence and iteration. Combined authorities are well placed to act as ‘living laboratories’ for AI-enabled growth testing new approaches, measuring impact, and rapidly adapting based on real-world outcomes. This data-driven approach can help ensure that ambition is matched by responsibility, particularly in areas such as energy use, environmental impact, and public trust.

There is an opportunity here for mayoral leadership to go further. The devolved government in Scotland has demonstrated how place-based policy can be used to align innovation and sustainability through the creation of Green Freeports. A natural next step would be for a combined authority mayor to pioneer a Green AI Freeport or Green AI Zone, bringing together low-carbon energy, compute infrastructure, skills, and regulation within a clearly defined geography. Such an initiative would not only accelerate local growth but would also provide a scalable model for how the UK can lead internationally in responsible, energy-efficient AI.

8. Ensuring UK Business Interests are Represented

As the UK shapes the energy and infrastructure foundations required to support the growth of AI, it is important to consider the balance between domestic and international interests. This is not a question of choosing one over the other, but of ensuring that the UK's AI ecosystem develops in a way that is open, competitive, and fair, and that it delivers lasting value to the UK economy.

UKAI strongly supports foreign investment and the participation of overseas companies in the UK AI market. Global technology firms play an essential role as suppliers, partners, and investors, and they will be critical to delivering the Green AI future set out in this report. However, the pace and scale of growth in the AI sector mean that UK businesses, particularly small- and medium-sized companies, can struggle to gain visibility for their innovations alongside well-established, multi-billion-dollar global corporations. For policymakers, it is often easier to engage with familiar international players than to identify and support the rapidly growing cohort of innovative UK firms.

This is where ecosystems have a vital role to play. Effective, place-based ecosystems help surface emerging domestic capabilities, connect UK businesses to skills, investment, and infrastructure and create pathways into public procurement and private markets. Because ecosystems are inherently local, they provide more opportunities for UK companies to engage directly, build trusted relationships and be recognised as key partners in delivering AI-enabled energy and infrastructure solutions.

The AI sector has welcomed the establishment of the AI Energy Council and recognises that its membership must be limited. However, there is an opportunity to strengthen its impact by ensuring greater representation from UK-based businesses, including medium-sized firms that reflect the breadth and diversity of the domestic AI industry. This is not an argument against the inclusion of global technology companies, whose role in shaping the future of AI infrastructure in the UK is undeniable. Rather, it is a call for a balance, one that reflects where long-term growth, resilience, and economic benefit for the UK will increasingly be generated: by British businesses – large and small – operating within strong, well-connected ecosystems.

9. How Ecosystems Encourage Investment in AI and Energy

For investors and startups alike, ecosystems reduce uncertainty by providing a clear and credible framework within which innovation can scale. In the context of AI and energy, this is particularly important: investment decisions depend not only on individual technologies, but on confidence that compute, power, data, regulation, and skills will evolve together in a coordinated and sustainable way. Ecosystems offer this assurance by aligning standards, demand, governance, and long-term strategic direction across multiple actors.

Rather than backing isolated firms or one-off projects, investors can back systems with momentum, places where risk is shared across infrastructure, markets, and institutions, and where successful innovation can be replicated and scaled. What investors need to see are clear pathways from research to deployment, predictable markets for low-carbon compute and AI-enabled energy services, and credible public and private institutions capable of long-term stewardship. Well-designed ecosystems provide these conditions by reducing friction through coordination, common standards, and shared physical and digital infrastructure.

Ecosystems also play a critical role in surfacing innovative companies to investors, particularly outside London and the South East. Many investors lack visibility into emerging regional clusters, not because of a lack of opportunity, but because they lack local networks, familiarity with place, or efficient ways to discover talent and innovation beyond established hubs. Ecosystems provide a ready-made entry point: a curated community of startups, scale-ups, researchers, and infrastructure providers, underpinned by trusted convenors and shared priorities. This makes it easier for investors to identify credible opportunities, assess risk and build relationships, while giving high-potential companies greater access to capital.

By lowering the cost and complexity of both deployment and discovery, ecosystems create clear benefits on both sides. Startups gain pathways into investment, procurement, and infrastructure, while investors gain efficient access to diversified, regionally grounded opportunities aligned with long-term national priorities such as Green AI.

The cost of fragmentation in AI and energy is high. Misaligned incentives, slow decision-making, and bespoke local requirements deter investment, delay deployment and risk pushing innovation and capital elsewhere. By contrast, ecosystems make integration the default rather than the exception, enabling AI systems to plug into energy networks, data centres, and public infrastructure in ways that are efficient, resilient, and low-carbon.

For the UK, this shift is fundamental to becoming a global leader in Green AI. Ecosystems translate national ambition on AI, energy security, and Net Zero into investable propositions, anchored in real places, credible institutions, and long-term demand. In this sense, ecosystems are not just a delivery mechanism; they are an investment proposition in their own right, creating a clear win-win for innovators, investors, and the UK economy.

10. Building Global Markets for Green AI

The UK has a long history of creating processes, standards, communities, and institutions that are adopted by businesses and governments around the world. This reflects the UK's identity as a trading nation, grounded in principles of fairness, transparency, and trust. Over time, the UK has demonstrated that doing what is right and doing what is commercially effective are not in conflict but are mutually reinforcing. Our legal system, financial services sector, and regulatory frameworks have coevolved to embed these principles into how markets operate and how international deals are structured.

This provides a strong foundation for UK leadership in Green AI, not only because it aligns with environmental and societal values, but because it is commercially viable and internationally credible. The UK is widely trusted to strike a balanced path between innovation, regulation, and market dynamism. This trust is a strategic asset, particularly in areas such as Green AI, where questions of energy use, sustainability, and governance must be addressed alongside competitiveness and growth.

Given this international standing, the UK is well placed to leverage the networks, coalitions, and partnerships it has already established. The UK maintains one of the world's most extensive diplomatic networks and exercises significant soft power, enabling it to shape global norms and open markets, punching well beyond its weight. This soft power has been central to commercial success, as demonstrated by initiatives such as the GREAT campaign, which has helped British businesses access new and growing international markets. This model provides a clear template for building cross-border ecosystems with like-minded partners and should now be applied to Green AI.

As the UK begins to define what Green AI means in practice, and to develop the playbooks and blueprints for delivering it, there is a substantial opportunity to export not just technology, but expertise. There is a growing global market for precisely this kind of applied knowledge: how to deploy AI at scale while managing energy demand, reducing emissions, and maintaining system resilience. If the UK leads in codifying and demonstrating these approaches, Green AI could become a major export opportunity, potentially exceeding even the success of the renewables sector.

Few countries are currently active in this space. This gives the UK a first-mover advantage, reinforced by the groundwork already laid through the development of world-class renewable energy markets and carbon trading mechanisms. While some larger economies have been slower to engage with Green AI, often for political or strategic reasons, the long-term direction of travel is clear. In a resource-constrained world with rapidly growing demand for AI, energy efficiency and sustainability will become unavoidable priorities. The UK has both the capability and the credibility to lead this transition.

11. Green AI and the Path to Consumer Trust

Consumers are a central, and often overlooked, part of the AI ecosystem: not only as the end-users of products and public services, but as informed participants who understand why AI is being deployed and how it delivers value sustainably. At the same time, Green AI itself provides an opportunity to reset the relationship between consumers and AI by addressing concerns around energy use, environmental impact, and system accountability, in a uniquely British way.

Ultimately, it is consumers who buy the products and services that businesses create and who rely on the services delivered by the public sector. Therefore, their confidence directly shapes adoption, market demand, and political legitimacy. This makes consumer engagement a core requirement for a successful AI-driven economy.

Engaging consumers, however, is fundamentally a question of trust. Public perceptions of AI remain mixed, and in many cases, negative, particularly in relation to energy and water consumption, environmental impact, and a lack of transparency. Addressing this requires more than reassurance. The AI industry must actively provide clear, credible information about how AI systems work, how resources are used, and how efficiency and sustainability are being designed into systems from the outset. Transparency and understanding are the foundations on which confidence is built.

While the AI sector is fragmented, it collectively holds the data, expertise, and evidence needed to support this effort. What has been missing is coordination. UKAI is therefore working to bring together businesses, researchers, and practitioners to develop a nationwide campaign to improve public understanding of AI: what it is, how it is used, and how it can deliver tangible benefits for people and communities. Green AI must sit at the centre of this narrative, demonstrating that innovation and sustainability can progress together.

Ecosystems provide the structures through which this trust can be built and maintained, connecting businesses, universities, local government, and communities in shared, place-based initiatives. By empowering these actors to collaborate, share knowledge and engage openly with the public, ecosystems turn abstract policy goals into lived experience. It is through strong, inclusive ecosystems that Green AI can be delivered responsibly, at scale, and with public consent. Therefore, the final section of this chapter sets out a series of recommendations aimed at strengthening these ecosystems and enabling them to deliver on that promise.

Policy Recommendations

Quick Impact Priorities (0–24 months)

1 Make Green AI a National Mission

The UK should articulate a clear, unifying national mission that aligns AI deployment, the energy transition, and infrastructure renewal. Framing this ambition around making the UK a global leader in Green AI would provide a strong and coherent direction of travel, one that connects existing Net Zero commitments with the UK's wider goals for AI-led growth and productivity.

At present, AI, energy, and infrastructure policy is often developed in parallel rather than in concert, creating fragmentation and uncertainty. A single national mission would help align policy, regulation, and public investment behind a shared objective, giving public bodies, industry, investors, and local leaders greater confidence in the long-term direction of UK policy.

Crucially, this is a strategic alignment that could be driven and owned by the Cabinet Office, rather than a new spending commitment. By providing clarity and shared purpose across government, a mission-led approach can reduce friction, accelerate delivery and unlock private investment more quickly, creating the conditions for progress without waiting for new programmes or funding. This new mission builds on, enables and supports existing missions (Growth, Justice, Opportunity, Health, and Clean Energy) rather than displacing them.

2 Support a National Industry–led Campaign to Build Consumer Trust in AI, With Green British AI at its Core

What is required is a coordinated, sustained campaign led by industry, with the Government acting as an active partner rather than the primary messenger. The objective should be to establish a shared industry–government narrative that reframes AI away from abstract risk and scale alone, and towards usefulness, energy efficiency, reliability, and responsible deployment. Positioned in this way, Green British AI becomes a practical and distinctive UK proposition rather than a purely policy concept.

The UK has a strong track record of successful industry-led campaigns that have shaped public understanding and behaviour in areas such as financial services, cybersecurity, and online safety. A similar approach is now needed for AI. Crucially, the campaign should be explicitly segmented. For the general public, the focus should be on everyday benefits, safeguards, and environmental responsibility. For SMEs, it should emphasise practical adoption, affordability, and access to trusted UK-based providers. For public-sector users, the emphasis should be on value for money, energy efficiency, resilience, and alignment with public missions.

UKAI is uniquely placed to convene this effort. As a neutral, industry-wide body, it can bring together researchers, practitioners, and companies of all sizes to develop shared messages, evidence, and real-world case studies. Government support could consist of endorsement, alignment with public-sector communications, and modest enabling resources.

3 Balance the Number of British Businesses on the AI Energy Council

As the UK plans the future energy system needed to support AI, it is essential that a broader range of British AI companies have a voice in shaping those decisions. The current AI Energy Council plays an important strategic role, but its impact would be strengthened by mechanisms that reflect the diversity of the UK's AI sector and the interests of end users.

This could include the introduction of rotating seats for UK-based AI companies, ensuring representation from mid-sized firms as well as larger domestic players, and creating a formal channel through which the wider AI industry can contribute views and evidence. Consideration should also be given to ensuring that the AI consumer perspective is represented, reflecting public concerns around energy use, cost, and environmental impact.

Providing UK AI businesses with a clearer role in discussions about energy planning would help align infrastructure decisions with real-world deployment needs. In doing so, the Council would better reflect the ecosystem it serves and support more balanced, future-facing policy outcomes.

4 Lay the Foundations for Green AI Standards and Global Interoperability

The UK should prioritise the early development of shared technical, data, and governance standards to ensure interoperability across AI, energy, and infrastructure systems by default. Initial priorities could include standardised methods for measuring and reporting AI energy and carbon intensity, alongside interoperable metadata standards that link AI workloads with energy and infrastructure systems. This remains an underdeveloped area globally, and one where the UK has distinctive strengths in standards-setting, regulation, and assurance. Acting early would provide clarity to industry, reduce fragmentation and position the UK as a trusted international convenor in Green AI.

Delivery could be led through existing institutions – including DSIT's standards and assurance programmes, the British Standards Institution, and sector regulators such as Ofgem and Ofcom – working in partnership with industry and academia. The UK can leverage its established leadership in ISO, IEC, and ITU to drive early Green AI standards internationally, helping to align approaches on energy efficiency, interoperability, and governance while reducing global fragmentation. Initial focus areas should reflect both domestic impact and international relevance, such as common metrics for AI energy efficiency, transparent reporting of compute and resource use, and interoperability between AI systems, data centres, and energy networks.

Crucially, this work should be developed in collaboration with international partners. Shared Green AI standards would support interoperability across borders, lower barriers to adoption and be particularly valuable for the Global South, where energy constraints are more acute and sustainable digital infrastructure is essential for development. By working with like-minded countries and multilateral partners, the UK can help ensure Green AI standards support inclusive growth while strengthening its global leadership role.

5 Designate National AI and Energy Living Labs

To accelerate the integration of AI, energy, and infrastructure at scale, the Government should designate and support a small number of national AI and Energy Living Labs. These living labs would act as real-world testbeds, demonstrating how AI systems, energy networks, data centres, and public infrastructure can be designed, deployed and operated together under operational conditions. Their purpose would be learning by doing: gathering data and testing technologies, governance models, and market mechanisms in live environments rather than through isolated pilots.

Given the complexity involved, the initial focus should be on two flagship living labs, recognising that attempting to deliver too many sites at once risks dilution and delay. These early living labs should be supported through a combination of targeted funding, regulatory flexibility, and enabling incentives, drawing on lessons from initiatives such as Green Freeports, while remaining focused on demonstrable outcomes rather than designation alone.

Crucially, this should be seen as an iterative, phased process. The first wave of living labs would generate evidence, case studies, standards, and operational insight, which can then be disseminated to other regions. A clearly signalled second wave would allow additional places and ecosystems to follow quickly, adapting proven approaches to local contexts rather than starting from scratch. In this way, knowledge and capability would accumulate over time and spread through the system.

Living labs could either be embedded within AI Growth Zones, bringing them to life by creating the energy, coordination, and experimentation that make them active and dynamic, or operate separately, outside formal Growth Zone boundaries. While both serve complementary functions, living labs are designed with ecosystem-building as the primary objective. Whether co-located or standalone, they would strengthen the AI Growth Zone as a concept by turning it from an investment designation into a functioning system for delivery, enabling combined authorities to coordinate industry, universities, energy providers, and public bodies around integrated infrastructure and innovation.

Through strategic use of existing innovation, research, and place-based growth programmes, national AI and Energy Living Labs can act as centres of excellence that can be quickly replicated and scaled.

Structural Priorities (2–10 years)

6 Using Public Sector Procurement to Shape and Scale Effective Ecosystems

Public sector procurement is one of the most powerful and currently underused tools for shaping and scaling effective AI and energy ecosystems. When used deliberately, procurement can do far more than purchase individual technologies: it can coordinate demand, create predictable

markets and signal long-term intent to industry. By aggregating demand across departments, agencies, local authorities, and the NHS, and by buying outcomes rather than bespoke solutions, public procurement can help ecosystems coalesce around shared challenges and integrated approaches.

Challenge-led and outcome-focused procurement can provide early revenue and validation for ecosystem participants, reduce investor risk and guide innovation toward real system needs. In doing so, it supports not just individual suppliers but the development of interoperable solutions across AI, energy, infrastructure, and services. This approach allows ecosystems to scale through delivery, rather than relying solely on grants or pilots.

To achieve this, public sector capability is critical. Acting as a system integrator for complex, AI-enabled ecosystems requires technical literacy, commercial expertise, and confidence in managing long-term delivery. Without this capability, procurement risks defaulting to narrow specifications or fragmented outsourcing, limiting learning and value creation.

Investing in public sector capability across commercial, digital, and project delivery functions enables the Government to be a more intelligent customer and a more effective partner to ecosystems. While capability-building is a long-term endeavour, it is essential to sustaining reform in procurement practice, reducing reliance on external intermediaries, and ensuring that public spending actively shapes resilient, integrated ecosystems rather than reinforcing fragmentation.

The Procurement Act 2023 provides the legislative headroom for this shift, but realising its potential will depend on how departments and public bodies use new flexibilities in practice, particularly to commission outcomes, aggregate demand and support interoperable ecosystem solutions rather than bespoke contracts.

7 Support Open and Federated Models as Ecosystem Infrastructure

Open and federated approaches should be recognised as foundational infrastructure for effective AI and energy ecosystems. Open-source technologies, shared data standards, and federated learning models enable collaboration across organisations while preserving data sovereignty, security, and trust. By reducing dependency on closed, proprietary systems, these approaches help prevent lock-in, support interoperability and allow innovation to emerge from across the ecosystem rather than a small number of dominant providers.

The Government has an important role in supporting this infrastructure through strategic direction, assurance frameworks, and sustained investment in shared digital capabilities. Aligning digital strategy, national data initiatives, and AI trust frameworks around open and federated principles would give industry greater confidence to adopt and build on them at scale.

While the benefits of open and federated infrastructure accrue over time rather than immediately, they are structural in nature. Consistent support enables knowledge, capability, and value to compound across sectors, strengthening resilience, competition, and long-term ecosystem health.

8 Future Systems Skills Lab

The Government should use AI and Energy Living Labs as focal points for developing place-based skills frameworks and curricula that respond directly to the needs of local ecosystems. Building on examples such as the Manchester Baccalaureate, living labs can bring employers, educators, and public bodies into the same environment, allowing industry to articulate the skills it needs, and course providers, trainers, and teachers to rapidly design, test and adapt programmes in response.

This approach enables skills development to keep pace with fast-moving advances in AI, energy systems, and infrastructure, while grounding learning in real-world deployment. Apprenticeships, vocational pathways, further education, and higher education can all be integrated within this living lab model, creating clear progression routes and shared standards.

By positioning skills development as a core function of living labs, regions can compete to become exemplars of best practice, raising quality, relevance, and ambition across the system. Over time, shared learning between living labs can help spread successful models nationally, aligning skills provision with the delivery of the Green AI mission across all UK regions.

9 Establish a National Green AI Delivery Partnership With Industry to Build and Scale Green AI Ecosystems

A dedicated industry-led body with responsibility for building, coordinating, and scaling the Green AI mission set by the Government. This body would bring together learning from living labs, AI Growth Zones, and regional initiatives, ensuring that insights are captured, shared and translated into practical guidance. Its role would include codifying case studies, developing and maintaining shared roadmaps and blueprints, and convening stakeholders across industry, academia, and government.

This body would be a delivery partnership (non-statutory, place-based delivery partnership), bringing together key stakeholders from industry, academia, and the public sector. It would also play a central role in communication and promotion, building national and international awareness of UK leadership in Green AI. UKAI would act as a founding partner and secretariat, helping to coordinate activity, share learning and support regions in translating ambition into action. Its position as an industry-led organisation, not a policy unit or government agency, enables it to contribute practical insight and convening capability, while working alongside government and other partners to build and scale Green AI ecosystems across the UK. Building on the innovations that are funded by existing innovation networks (UKRI, Catapults), the partnership would convene industry and local government to codify what works into case studies and playbooks, sharing knowledge and best practice and helping regions move from pilot activity to scaled deployment while minimising duplication.

Strategic Implications: From Connected Systems to a Green AI Advantage

This chapter has argued that ecosystems are the missing delivery layer between ambition and outcomes. Where previous chapters examined the components of sustainable AI (connected infrastructure, aligned energy pricing and efficiency-driven innovation), this chapter has shown how those elements can be brought together into systems that are repeatable, investable, and scalable over time and across borders. Without ecosystems, progress remains fragmented and fragile. With them, learning compounds, risk is shared, and delivery accelerates.

Ecosystems also provide the bridge between national ambition and local legitimacy. By embedding AI infrastructure into places through modular deployment, shared infrastructure, heat reuse, edge compute, and coordinated governance, ecosystems make efficiency visible and benefits tangible. This is critical to sustaining public trust and political consent as AI deployment expands.

The overarching implication is therefore one of confidence. The UK does not need to win a race to scale at any cost. It can lead by demonstrating how AI can be deployed intelligently and responsibly in a resource-constrained world: using less energy per outcome, integrating infrastructure into communities, and aligning innovation with public value in the UK and globally. By treating AI, energy, and infrastructure as a single mission delivered through ecosystems, the UK can make its ambition to lead in Green AI both credible and achievable.

This is not the end of the agenda, but the point at which strategy becomes executable. The task now is sustained commitment: to coordination over fragmentation, to efficiency over volume, and to systems that endure beyond individual projects or political cycles. If that commitment is maintained, Green AI can move from concept to comparative advantage and from aspiration to reality.

How Ecosystems Drive Efficiency

In the context of Green AI, energy, and infrastructure, an ecosystem is best understood as a coordinated system of actors, assets, rules, and incentives that enables innovation to move reliably from research and pilots into deployment at scale. It is more than a 'cluster' in the traditional sense, and it is not a single programme or platform. Instead, it is a dynamic system that connects strategy to delivery, aligns public and private interests and enables learning to compound over time.

Ecosystems can operate at many different levels, from communities and cities to combined authorities, regions, nations, and international partnerships. While the boundaries of an ecosystem may be somewhat arbitrary, once defined, they provide a useful mechanism for quickly understanding what is included, what is excluded, and how different actors relate to one another. The UK has seen a number of previous attempts to support innovation ecosystems, but these have often been characterised by top-down designation rather than genuine empowerment of local businesses and stakeholders. The result has too often been superficial ecosystems in name rather than function. To deliver impact, ecosystems must be deeper and more inclusive, designed around real collaboration, shared capability, and active participation by organisations of all sizes.

An effective ecosystem brings together central government; devolved administrations and local authorities; public sector institutions such as the NHS, transport bodies, and utilities; universities and research organisations; private companies ranging from startups to infrastructure operators; and the investors who finance growth and deployment. Each plays a distinct role, but the value of the ecosystem lies in how these roles are coordinated.

The purpose of an ecosystem is not simply to accelerate innovation, but to reduce friction, share risk and enable learning and repeatability at scale. In the context of Green AI, ecosystems are essential because progress depends on the coordinated development of algorithms, data, infrastructure, regulation, and skills, none of which can be delivered by a single organisation in isolation. Well-functioning ecosystems create common standards, shared infrastructure, predictable demand, and trusted relationships, allowing energy and compute-intensive AI systems to be designed, tested and deployed in ways that are efficient, secure, and environmentally sustainable.

This ecosystem approach provides a practical route to leadership in Green AI: The UK can export this expertise by shaping international partnerships, standards, and coalitions.

While it is relatively easy to talk about building ecosystems, it is harder to turn this into a reality. Each ecosystem requires partners and leaders to set out the objectives and to deliver. This is where UKAI has an important role to play, to bring these different parties together. This chapter aims to present concrete proposals and practical ideas.

Final Reflections and Next Steps

This report has argued that the future success of the UK's AI sector will be determined less by the pace of technical innovation alone, and more by the interconnected systems approach that enables AI to be deployed at scale. Energy, infrastructure, markets, regulation, and coordination are no longer peripheral considerations; they are now decisive. For the UK, this creates both a challenge and an opportunity. The challenge is that many of the constraints holding back AI growth are structural and cross-cutting. The opportunity is that these are precisely the areas where the UK can differentiate itself.

Taken as a whole, the four chapters of this report set out a coherent logic. Chapter 1 established that AI infrastructure only delivers value when energy, planning, connectivity, skills, and demand are treated as a single system. Chapter 2 showed that energy pricing and market design are not peripheral constraints but strategic levers that shape where and how AI can be deployed. Chapter 3 demonstrated that innovation across hardware, software, and deployment models is already reducing the resource intensity of AI and can turn constraint into competitive advantage. Chapter 4 completed the picture by showing how those gains can be embedded into ecosystems that enable scale without waste, growth without backlash, and innovation without fragmentation.

A central implication of this report is that Green AI is not a niche agenda or a trade-off against competitiveness. It is a systems strategy rooted in efficiency, coordination, and long-term value creation. In a world of constrained grids, high energy costs, and rising public scrutiny, the ability to deploy AI that is energy-efficient, locally integrated, and economically productive will increasingly define and determine leadership. The UK's conditions of dense infrastructure, strong institutions, world-class research, and credible regulation make this approach not only desirable but realistic.

For the Government, the role that emerges across the chapters is one of system steward rather than project sponsor. Leadership does not require directing every investment or picking technologies, but it does require setting a clear national mission, aligning incentives across policy domains, and creating the institutional conditions for ecosystems to form and persist. Standards, procurement, shared demand, skills pipelines, and test beds are not secondary tools; they are the mechanisms through which strategy becomes delivery.

Across all four chapters, several key takeaways stand out. First, AI is now inseparable from energy and infrastructure. Without smarter, more connected foundations, the UK risks slow delivery, rising costs, and declining competitiveness. Second, energy pricing matters as much as energy supply. High and volatile prices undermine investment and favour short-term optimisation over long-term value. Third, innovation must be targeted toward real-world performance, particularly energy efficiency and system integration, rather than abstract measures of scale. And finally, none of these changes will deliver impact unless they are embedded in ecosystems that allow implementation to scale.

The biggest problems identified in this report are not a lack of talent, ambition, or capital. They are fragmentation, misaligned incentives, and unclear signals. Infrastructure is planned without energy certainty; innovation is funded without a clear route to deployment; markets reward volume rather than efficiency; and responsibility for system outcomes is dispersed across institutions with limited coordination. These failures are cumulative. Left unaddressed, they will continue to push AI investment and economic value away from the UK, even as domestic capability grows.

What is different about the approach set out in this report is that it treats these issues as parts of a single system rather than isolated policy challenges. Smarter infrastructure creates the conditions for affordable, predictable energy. Fairer pricing rewards efficient use of that infrastructure. Targeted innovation reduces demand on the system while increasing economic value. Scalable systems ensure that what works in one place can be repeated, financed and trusted elsewhere. Each chapter reinforces the others; none is sufficient on its own.

The timelines for action are therefore both urgent and realistic. In the next 12 to 24 months, the Government can deliver meaningful progress through clearer national priorities, improved coordination, targeted market reforms, and better use of existing policy levers. Over the medium term, three to five years, these changes should translate into faster infrastructure delivery, more stable investment conditions, and measurable reductions in the energy intensity of AI deployment. Over the longer term, the UK can embed a durable advantage by making Green AI the default approach rather than an exception.

The Government has a central role to play. It must set a clear mission for Green AI, align policy across departments, reform markets where they are no longer fit for purpose and act as a steward of long-term system performance. This does not require the Government to do everything itself, but it does require clarity, consistency, and leadership.

Industry also has a responsibility. AI businesses, infrastructure providers, and investors must engage with efficiency, transparency, and collaboration as sources of competitive advantage. They can help shape standards, participate in testbeds, share evidence on what works and invest in solutions that deliver long-term value rather than short-term gain. Trade bodies such as UKAI have a crucial role in convening this dialogue and translating industry insight into practical policy.

The UK has the assets to lead in AI, but leadership will not happen by accident. It requires a clear mission, aligned systems, and coordinated delivery across government, industry, and academia. This is an area where the UK can lead; this report sets out how to do that in practice. The choice now is whether to act or to fall behind as others move faster. With the right decisions and the determination to drive coordinated action, the UK can lead the world as a Green AI superpower.

Summary of Recommendations

	Integrated Infrastructure	Fairer Pricing	Targeted Innovation	Scalable Systems
1	Treat AI Infrastructure as a Connected National System	Anchor Electricity Prices for Strategic Infrastructure through Long-term, Low-carbon Contracting	Make Energy-Efficient Data Centres the Default for the Public Sector	Make Green AI a National Mission
2	Establish a Cross-government AI Infrastructure Delivery Unit	Reduce Delivered Electricity Costs for Data Centres Using Established Fiscal and Levy Frameworks	Use Transparency and Measurement to Drive Continuous Efficiency Gains	Support a National Industry-led Campaign to Build Consumer Trust in AI, with Green British AI at its Core
3	Give National Strategic Weight to AI and Compute Infrastructure	Treat Data Centres as Active System Participants, not Passive Energy Consumers	Win in the Application Layer and Frugal AI, Not Compute Scale	Support a National Industry-led Campaign to Build Consumer Trust in AI, with Green British AI at its Core
4	Require Early Energy Integration in AI Infrastructure Proposals	Reframe Electricity for Strategic Digital Infrastructure as National Infrastructure, Not a Traded Commodity	Create a National Showcase for Energy-efficient AI Innovation	Lay the Foundations for Green AI Standards and Global Interoperability
5	Introduce Standard Transparency Metrics for AI Infrastructure	Separate Low-carbon Baseload Pricing from Marginal Generation to Reduce Gas-driven Volatility	Back UK-leading Efficiency-driven Compute Innovation	Designate National 'AI and Energy Living Labs'
6	Anchor AI Infrastructure Growth to Evidence-based Demand	Use Spatial Pricing and Direct Connections to Align AI Infrastructure with Energy Abundance	Position the UK as a Global Leader in Efficient Data-centre Design and Integration	Using Public Sector Procurement to Shape and Scale Effective Ecosystems
7	Align Planning Reform With Grid, Energy, and Skills Constraints	Establish a Carbon Trading Framework for Data Centres to Align Cost, Efficiency, and Sustainability	Embed AI Infrastructure Into Communities Through Local Coordination and Shared Solutions	Support Open and Federated Models as Ecosystem Infrastructure
8	Shift Energy Policy Toward Shared System Upgrades		Position Innovation, Not Scale, as the UK's Competitive Strategy in AI Infrastructure	Future Systems Skills Lab
9	Treat Digital Connectivity as Critical AI Infrastructure			Establish a National Green AI Delivery Partnership with Industry to Build and Scale Green AI Ecosystems
10	Embed Community Benefit and Local Legitimacy Into Infrastructure Design			
11	Treat Infrastructure Skills as a Delivery-critical Asset			

ukai

Key Policy Takeaways

- **AI is now a systems challenge, not just a technology one**
  The UK's AI success depends on coordinated energy, infrastructure, markets, and regulation, not on technical innovation alone.
- **Energy pricing and infrastructure integration will determine competitiveness**
  High, volatile electricity costs and fragmented planning are now core barriers to AI deployment and investment.
- **The UK should compete on efficiency and integration, not compute scale**
  Energy-efficient infrastructure, frugal AI, and application-layer innovation offer a more sustainable and defensible advantage than trying to outbuild global hyperscalers.
- **Fragmentation is the UK's biggest constraint, not talent or capital**
  Misaligned incentives and weak coordination across government and markets are pushing AI investment and value away from the UK.
- **Green AI can be a national advantage if the Government sets a clear mission**
  With aligned policy, smarter markets, and targeted use of public procurement, the UK can not only lead in Green AI but also export it globally.

About UKAI

UKAI is the UK's only trade association representing the entire AI sector, providing a unified voice for tech and non-tech businesses who are harnessing AI to drive economic growth. UKAI brings together a thriving eco-system of businesses, investors and government, driving collaboration to secure the UK´sposition as a global AI leader.

Golden Cross House
8 Duncannon Street
London WC2N 4JF

www.ukai.co
contact@ukai.co