Executive Summary: Managing the Unprecedented Energy Demand of AI Data Centres

The International Workshop on Managing Global Energy Demand of AI Data Centres, held on September 9-10 at **Imperial College London,** brought together a diverse group of experts from academia and industry to address one of the most pressing challenges of the energy transition. The consensus was that data centres, if still viewed as passive loads, may cause major challenges for the grid's control, operation and planning. However, they have significant potential to evolve into flexible, "grid-interactive" assets that benefit both the data centres and the grid. The workshop highlighted the urgent need for collaborative, interdisciplinary solutions to ensure grid stability and security, and facilitate a sustainable transition.

The Looming Crisis: A Mismatch of Speed and Scale

The workshop's opening session, "Challenges of Surging Energy Demand in Data Centres," set a stark tone. Speakers from the **UK National Energy System Operator (NESO)**, **UK Power Networks (UKPN)**, and Professor Sara Walker from the **UK Energy Demand Research Centre** presented a clear picture of an impending crisis. Data centre electricity demand is no longer a marginal issue; globally, it reached 415 TWh in 2024—the equivalent of 132% of the UK's total annual electricity consumption. The situation is particularly critical in the UK. NESO forecasts that electricity demand from data centres could reach 5–7 GW by 2030, with some of its 2025 Future Energy Scenarios (FES) models projecting that connections could approach 15 GW by 2050.

One core problem, that highlighted by multiple speakers, is a severe mismatch in development speeds. The construction of new data centres is proceeding at a pace far exceeding the grid's capacity to expand. This staggering growth underscores the immediate pressure on the grid. This challenge is not unique to the UK; speakers cited real-world examples in Ireland and the ERCOT grid in Texas, where data centre-related instability and fault-triggered large-scale load drops have already occurred. Discussions centred on the "black box" nature of data centres, where a lack of granular data on their operational and load profiles makes it difficult for grid operators to manage grid stability. Attendees agreed that this situation resembles the early days of wind power integration, where limited data and a lack of standards created significant challenges. For the UK, this new demand layer is superimposed on an already strained grid, which is grappling with the large-scale integration of intermittent renewables, the electrification of heat and transport, and frequent curtailment of wind power in Scotland.

Technological Frontiers: Towards Energy-Efficient and Flexible Operations

A key focus of the workshop was on innovative solutions to mitigate demand and integrate data centres more effectively. In the session on "Energy-Efficient Data Centres," speakers from the **National University of Singapore** and **Lenovo** detailed the limitations of conventional air cooling, which is struggling to manage the dense power loads of modern AI servers (often exceeding 150 kW per rack). The discussion converged on liquid cooling technologies—specifically direct-to-chip and single-phase immersion cooling—as a critical solution. A case study presented from a Microsoft project showed that liquid cooling could achieve a 20% energy and GHG saving over traditional methods.

Beyond efficiency, a central theme across sessions was the untapped potential of data centres for grid flexibility. The workshop moved the conversation beyond simple energy reduction to viewing data centres as intelligent, grid-responsive assets. Speakers from **Cardiff University** and **Black & White Engineering**, detailed how data centres could provide valuable demand response services, utilising their Uninterruptible Power Supplies (UPS), on-site battery storage, and workload shifting. From a grid stability perspective, Professor Le Xie of **Harvard University** highlighted the risks of large load trips, citing an instance of a 400 MW trip in West Texas, while Professor Christopher Knittel of **MIT** discussed the complex trade-offs between the cost and emissions benefits of temporal flexibility. Professor Zhaohao Ding from **North China Electric Power University** emphasised that advanced scheduling methods achieve cost savings close to optimal solutions with improved scalability. The real-world pilots with Alibaba in China demonstrated a significant potential for temporal load shifting by deferring batch workloads in response to grid regulation signals. However, how to motivate data centre owners to adopt flexible energy management strategies, and how to raise awareness of their power-intensive nature and sustainability issues from the end-user perspective, remain open research questions.

Furthermore, the session on "Grid Planning and Interconnection" also highlighted significant risks, such as the danger of a mass backup switchover during grid disturbances, where hundreds of megawatts of load could suddenly disconnect, causing widespread system instability. Experts from **Siemens Energy, the Energy Systems Integration Group (ESIG)**, and **EPRI Europe** emphasised the urgent need for harmonised Fault Ride Through (FRT) standards to prevent such systemic failures.

A UK-Specific Call for a Holistic Approach

The workshop concluded with a session chaired by Professor Mark O'Malley, alongside Mr Loren Long and Mr Simon Allen, titled "Opportunities of synergising Future Data Centres with low carbon Energy Systems. " This session offered a powerful synthesis of the preceding discussions, with a clear and urgent message for the UK. While the challenges are global, their confluence with the UK's specific energy goals—such as accelerating AI adoption, scaling up offshore renewables, and the electrification of heat and transport—makes the need for action particularly pressing. The consensus was that a fragmented approach, focusing on isolated issues like cooling or grid codes, would be insufficient.

The workshop's final conclusions highlight several critical areas for future work and research:

1. **Systemic Integration**: The problem cannot be solved by data centres or the grid alone. Drawing from international examples, experts stressed that the UK must develop **systemic solutions** that address both planning and operational challenges

simultaneously. This is essential to manage the immense new load requests and prevent the grid instability seen elsewhere.

2. **Bridging the Data Gap**: A fundamental obstacle in the UK remains a lack of good-quality, publicly available data on data center energy consumption patterns and load dynamics. Overcoming this **"black box" problem** is crucial for UK grid operators to perform effective modelling and planning for future demand.
3. **Harnessing Flexibility**: The untapped flexibility of data centre workloads—such as shifting non-latency-sensitive tasks (e.g., training large language models) to align with UK renewable energy generation—presents a major opportunity for decarbonization. Considering the UK's immense offshore renewable energy potential, the workshop raised the crucial point of resolving the geographic mismatch between generation and demand. This implies the need for innovative solutions that could involve **co-locating** computing infrastructure with energy resources to enable **direct energy integration**, bypassing costly and congested terrestrial grid infrastructure.
4. **Policy and Market Innovation**: Creating new market mechanisms, tariffs, and regulatory standards is essential to **incentivise UK data centres** to participate in grid services and to ensure their growth aligns with the nation's broader sustainability goals.

A proactive, collaborative, and system-level approach is not only desirable but essential to prevent the AI revolution from overwhelming the UK’s energy transition. The findings of this workshop provide a clear and compelling case for immediate action and cross-sector cooperation.