March 14, 2025

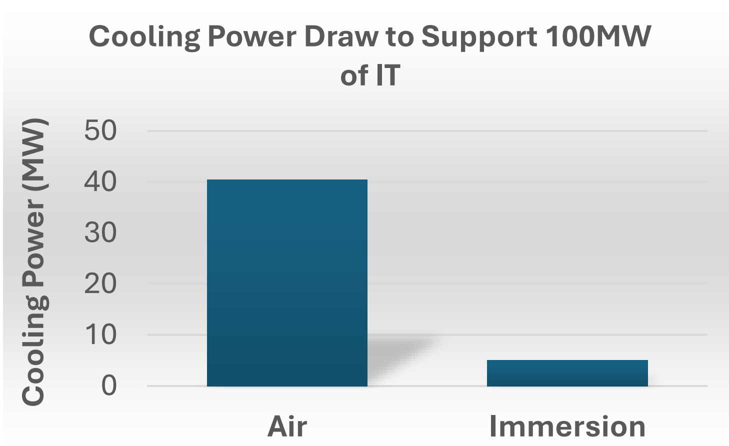
Office of Science and Technology Policy

Executive Office of the President, The White House

RE: Request for Information on the Development of an Artificial Intelligence Action Plan

Artificial Intelligence Policy Topic: Energy Consumption and Efficiency. To ensure that the U.S. holds the competitive edge in developing and deploying Artificial Intelligence (AI) and maximizes AI output, the White House policy on AI should prioritize state-of-the-art energy efficiency in the design of data centers. As AI continues to transform industries and stimulate economic growth, it is imperative that energy constraints do not impede its progress.

Today **6350 MW**[1] of data center capacity is under construction in U.S. in primary markets. If traditional data center air-cooled infrastructure is implemented, **40% of total facility power**[2] will be spent on cooling the servers alone. This inefficiency wastes energy that could be used to improve server speed and performance. Better thermal management technology exists and should be encouraged to maximize the performance of these mission critical facilities. Immersion cooling, for example, significantly reduces the energy needed for temperature control, allowing the redeployment of power to accelerate the core functions of AI technology.



[3]

[1] North America Data Center Trends H2 2024 | CBRE
[2] https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy
[3] Open Compute Project, TCO model: Cooling Environments/Immersion - OpenCompute
https://docs.google.com/document/d/1OU9RiL-vOIWCGN8k746Q4cW9FHPh7b79Hrd7T-3soeE/edit?usp=sharing

Given that cooling systems currently represent the second-largest energy consumption function in data centers after IT, optimizing the energy needed for thermal management is a critical necessity. As the United States embarks on large-scale data center infrastructure expansion that will be with us for decades to come— White House policy that prioritizes energy efficiency in thermal management can radically shift the dynamic, creating a competitive advantage for the US to advance AI, by reallocating large fractions of existing and future grid capacity away from cooling, and instead to productive AI performance.

The Unprecedented Growth of Data Centers

The digital revolution has fundamentally transformed the landscape of data storage and processing, leading to an unprecedented growth in data centers worldwide. This growth is driven by a multitude of factors, including the increasing adoption of cloud computing, the proliferation of digital services, and the surge in data-intensive applications.

Over the past decade, the number and capacity of data centers have expanded dramatically. According to the International Energy Agency, data center workloads increased by approximately 550% between 2010 and 2020, while the energy use associated with these centers remained relatively flat thanks to efficiency gains in hardware and cooling. During this period, data centers accounted for roughly 1–2% of total U.S. electricity use.[4]

Today, a second wave of data center growth is being ushered in by AI, one of the most transformative technologies of the 21st century, with applications ranging from machine learning and natural language processing to autonomous systems and predictive analytics. AI workloads are particularly resource-intensive, requiring substantial computational power and storage capacity. As organizations increasingly deploy AI-driven solutions, the demand for robust data center infrastructure has intensified. The convergence of standard business workloads and AI applications is driving an increase in the average size of U.S. data centers from 40 megawatts (MW) today to 60 MW by 2028, with about one-third of campuses exceeding 200 MW.[5]

The rapid growth of data centers, particularly those supporting AI workloads, has significant implications for energy consumption. Goldman Sachs Research forecasts a 50% increase in global data center power demand by 2027 and up to a 165% increase by 2030, driven by AI advancements.[6] This surge presents both a challenge and an opportunity—to implement cutting-edge energy efficiency measures that can sustain growth while minimizing strain on the power grid.

---

[4] Global trends in internet traffic, data centres workloads and data centre energy use, 2010-2020 – Charts – Data & Statistics - IEA
[5] AI data center growth: Meeting the demand | McKinsey
[6] AI to drive 165% increase in data center power demand by 2030 | Goldman Sachs

Energy Efficiency in AI Policy Will Drive Competitive Advantage

The United States is in a race for dominance in the AI space. The nation with the most competitively advantageous public policy will win this race. Policy that empowers US business to stay ahead of global competitors is paramount. AI is not only a cornerstone of economic advancement but also a critical component in military and strategic capabilities. Energy efficiency as a core pillar in AI policy delivers a strategic advantage by ensuring the energy dedicated to AI facilities is used to build, support and deploy this game changing technology more effectively. By minimizing the energy spent on cooling, more resources can be allocated to running complex algorithms and computations. This energy efficiency translates to lower operational costs and higher performance outputs, giving the United States a competitive edge in developing and deploying cutting-edge AI technologies. It also supports the scalability of AI infrastructure, enabling rapid expansion and innovation without the limitations imposed by traditional cooling methods.

Cooling Alone Consumes 30-40% of Energy Used by Data Centers

Energy consumption in data centers can be broadly categorized into three principal areas: IT equipment, cooling, and other infrastructure.

1. **IT Equipment**: This includes servers, storage devices, and networking hardware. IT equipment is the primary consumer of energy in data centers, accounting for 50-60% of total energy use.[7]
2. **Cooling**: Cooling systems are essential for maintaining optimal operating temperatures and preventing overheating. These systems typically consume about 40% of the energy used in data centers.[8]
3. **Other Infrastructure**: This category includes power distribution units, lighting, and security systems, which collectively account for around 10-20% of energy consumption.[7]

Servers generate heat as a byproduct of their operation. This heat is produced by the electrical resistance within the server components, such as the central processing units (CPUs), graphics processing units (GPUs), memory modules, and power supplies. The more computationally intensive the tasks, the more heat is generated.

Figure 1[9] highlights an inflection point in IT cooling demand led by the emergence GPU-based compute. Each data point represents a chip generation from a major chip OEM with an inflection starting in 2018 based on the GPU-based AI boom. 'Inverse Thermal Resistance' is a measure that incorporates the increasing power (Watts) of newer chips and the lower air temperature limits (°C) data center operators must balance with overall

---

[7] Data center sustainability | Deloitte insights
[8] Why invest in the data center economy | McKinsey

energy efficiency. High power GPUs require a greater amount of cooling which is unlikely to be addressed by simply lowering the cooling air temperature. Liquid cooling solves this crisis as it enables a higher cooling efficiency (i.e., Inverse Thermal Resistance) than traditional air cooling and without the need of inefficient cold operating temperatures.

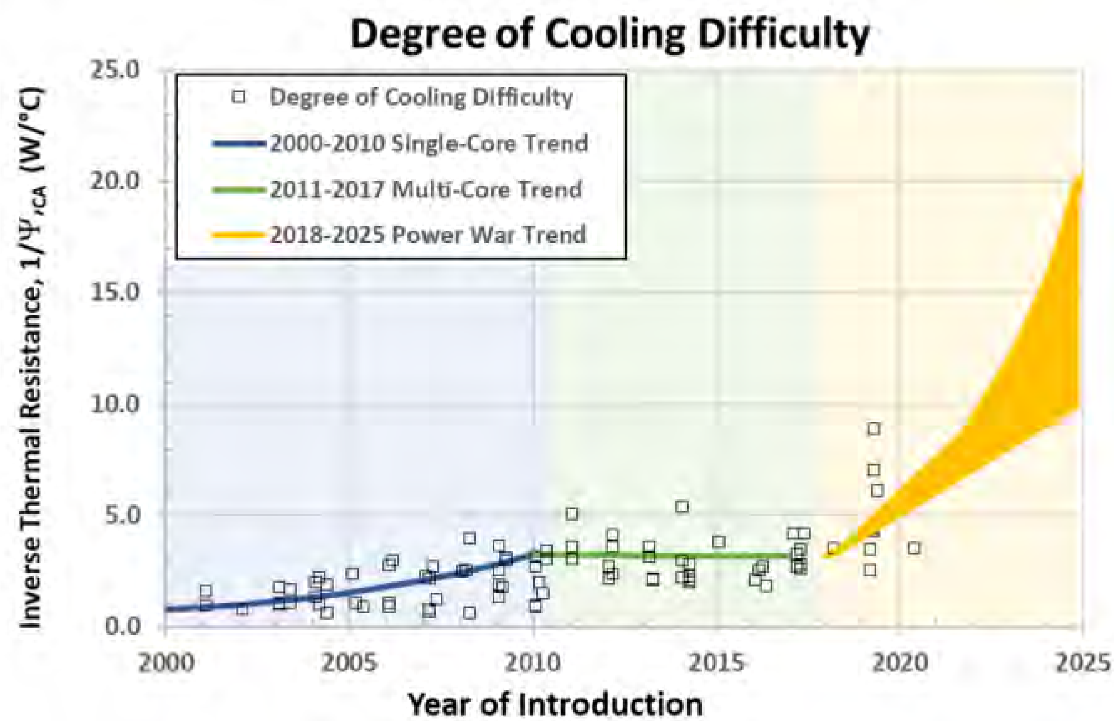


Figure 1. Degree of cooling difficulty for socket cooling, 1/Ψca, ASHRAE

Legacy Cooling Technology is Energy Intensive

Air cooling remains the most common method for cooling data centers. Traditional systems involve chillers, pumps, control room air conditioning and air-handling units (CRAC/CRAH), and fans to keep temperatures optimal as air passes over IT via fans attached to the server. Advanced methods create a “cool aisle” for undiluted cold air and push the warm air to a “hot aisle”, but both methods are energy intensive.

Chillers, which supply cold water to CRAH units, use significant energy, particularly when operating at full capacity and not linked to actual demand. Pumps circulating chilled water also waste energy if they lack variable-speed controls, forcing them to operate inefficiently under varying loads. CRAC and CRAH units, essential for delivering cold air and removing hot air, often operate at constant speeds, leading to overcooling and unnecessary energy use. Even with hot and cold aisle containment, poorly designed air distribution systems can create bypass airflow, causing fans to work harder to compensate for air imbalances. Despite advancements in airflow management, these supporting systems remain major contributors to energy waste, making data center cooling one of the largest operational costs.

Server manufacturers are pushing air cooling to its limits, impacting power use. The industry has reduced fan power as a % of server power significantly, from 20% to as low as 2%. However, as can be seen in Figure 1, increasing cooling demand is forcing fan power to rise again – with denser configurations utilizing fan power again at 10-20%.[9]

Cold plate cooling is being implemented in newly built data centers to address AI heatloads. The coexistence of cold plates with traditional air-cooled infrastructure represents a significant new evolution in cooling technology. Cold plates, which involve direct liquid cooling of electronic components, are often integrated with air-cooled systems to form a hybrid cooling strategy that leverages the strengths of both methods.

While this method is more efficient at dissipating heat due to the superior thermal conductivity of liquids compared to air, it still requires fans to remove heat from other non-plated heat generating components. The absorbed heat is transferred to a heat exchanger, where it is released into the ambient environment, often with the assistance of air-cooled systems.

Integrating cold plates with air-cooled infrastructure offers data centers a step towards enabling AI without requiring extensive redesigns of traditional data centers. Despite this benefit, this is a “band-aid” approach that has significant downsides. Cold-plate systems still rely on air-based cooling infrastructure, including CRAH/CRAC units, chillers, and air distribution fans, to dissipate heat. This two-step cooling process introduces inefficiencies as it necessitates pumps for coolant circulation and subsequent heat transfer to air, which has a lower thermal conductivity than liquid. Additionally, heat exchanger losses between the water-glycol system and the air-cooled infrastructure further diminish efficiency.

Studies indicate that cold plate liquid cooling systems can cut cooling energy use by 20-50% compared to air cooling. With an average reduction of 30% for cold plates[9], a data center's cooling energy would drop from 40% to 28%.

Despite this reduction, a considerable amount of energy continues to be wasted. Immersion cooling, which is detailed below, can lower cooling energy consumption to less than 6%, achieving an 87% reduction in cooling energy.[3]

Energy Efficient Alternatives, like Immersion Exist and Can Significantly Reduce Energy Consumption for Cooling

Immersion cooling, a method in which IT hardware is submerged in non-conductive dielectric fluid, has a history that dates back several decades. Initially explored in the 1960s and 1970s for military and aerospace applications, immersion cooling was recognized for its potential to manage high heat loads in compact electronic systems. Over

---

[9] Cold Weather Shipping Acclimation and Best Practices

the past 50 years, immersion cooling has evolved significantly, with advancements in fluid technology and cooling system design.

In the 1980s and 1990s, it found applications in supercomputing and high-performance computing environments, where traditional air-cooling methods were insufficient. Today, AI has significantly renewed interest in immersion cooling, driven by the unprecedented computational demands and heat generation of AI hardware.

Firstly, immersion cooling provides superior thermal management by directly submerging electronic components in thermally conductive, electrically non-conductive dielectric fluids. These fluids allow the heat generated to be transferred away more efficiently than air. This direct contact with the cooling medium enables higher power densities without the risk of overheating, which is critical for maintaining the performance and reliability of AI systems. The efficient heat dissipation offered by immersion cooling ensures that AI workloads can be processed without thermal throttling, thereby enabling faster AI model training and inference.

Secondly, immersion cooling enhances energy efficiency by eliminating the need for traditional air conditioning and high-speed fans, which are typically required to maintain optimal temperatures in air-cooled data centers. As described above, air conditioning systems and fans consume a significant amount of energy, contributing to higher operational costs. Immersion cooling eliminates the need for traditional air-cooling components, cutting cooling energy consumption by up to 87%. Based on the Open Compute Project's Total Cost of Ownership model, Figure 2 provides a comparison of an air-cooled data center to an immersion cooled data center cooling power consumption. Both data centers support 100MW of IT, but the air-cooled data center requires 40MW of power for cooling while the immersion cooled data center only requires 5MW. This means 35MW of savings in power.

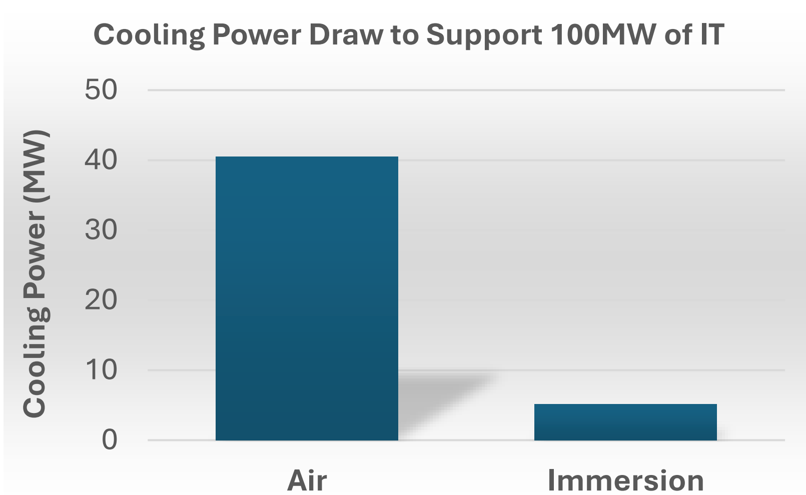


Figure 2. Total cost of ownership model

In a 400MW data center, this same scenario could scale to 140MW of energy savings. 140 MW of energy can power approximately 116,000 average U.S. households[10] or could be redeployed to scale AI infrastructure by an additional 140MW.

Immersion Cooling Enables Additional Benefits

Increased Computational Density and Space Reduction: Traditional air-cooling systems require ample space for airflow management, including hot and cold aisles, raised floors, and overhead ducts. Immersion cooling systems, however, can be more compact, as they do not rely on airflow for heat dissipation. This reduction in space requirements allows data centers to increase their computational density, fitting more servers into the same physical footprint. This is especially advantageous because it could enable data centers to be built in areas where it is not currently economically viable such as in high-cost real estate areas where maximizing space efficiency is a priority.

Improved Hardware Durability: Immersion cooling supports the reliability and longevity of AI hardware by maintaining consistent operating temperatures and minimizing thermal stress. Fluctuations in temperature can cause expansion and contraction of electronic components, leading to mechanical stress and potential failure over time. By providing a stable thermal environment, immersion cooling reduces the risk of such failures, thereby extending the lifespan of expensive AI hardware and reducing maintenance costs. This reliability is particularly important for AI applications that require continuous, high-intensity processing, such as real-time data analytics, machine learning model training, and complex simulations.

Water Use Reduction: Water is the second-largest resource consumed by data centers, after electricity. A hyperscale data center can use up to 200 million gallons of water per year[11]. While data centers use a significant amount of water for cooling, the majority of their water footprint actually comes from electricity generation. On average, **7.6 liters of water** are required to produce **1 kWh of electricity** in the U.S., whereas a typical data center directly consumes **1.8 liters of water per kWh** for cooling[12]. This means that for every unit of power a data center uses, nearly **four times more water** is consumed indirectly through power generation than directly for cooling. Immersion cooling helps mitigate both sources of water consumption. By significantly improving energy efficiency, it reduces a data center's overall power demand, thereby decreasing the amount of water used for electricity production. Additionally, immersion cooling enables the use of **non-evaporative chillers**, such as dry coolers, which eliminate or drastically reduce direct

---

[10] U.S. Energy Information Administration (EIA). "How much electricity does an American home use?" 2023. EIA Report

[11] EU moves toward regulating data center energy and water use | CIO

[12] United States Data Center Energy Usage Report, LBNL, 2016

water consumption, making it a key solution for addressing data centers' growing water footprint.

Noise Reduction: Another benefit of immersion cooling is the quieter operation it affords. Traditional air-cooled data centers rely on thousands of fans and air conditioning units, which generate significant noise. Immersion cooling eliminates the need for these noisy components, resulting in a much quieter data center environment. This can be particularly beneficial to workers and in settings where noise levels need to be minimized, such as in urban data centers or facilities located near residential areas.

Hundreds of servers (or more, when it comes to hyperscale data centers) operating in compact spaces can together create excessive noise. By some estimates, the average noise level around server areas reach up to 92 A-weighted decibels (dBA).[13]

Research has shown that the constant humming of heating, ventilation, and air conditioning (HVAC) systems can create noise sometimes in excess of 80 dBA, nearing the maximum acceptable threshold of 85 dBA during an 8-hour exposure period set by the National Institute for Occupational Safety and Health (NIOSH).

In contrast, **immersion cooling is quiet**—and it has the potential to eliminate the excessive whirring and humming of HVAC systems while bringing a range of other benefits. At a time when labor is hard to come by, creating a safe environment where employees enjoy working can be a critical differentiator.

Community Benefits – Easing Strain on Shared Resources

Adopting immersion cooling offers significant community benefits. Traditional cooling systems place a heavy burden on local power grids, often leading to increased competition for limited energy resources. By reducing the energy footprint of data centers, immersion cooling helps alleviate this strain, ensuring that more power is available for other essential services and needs within the community. This is particularly important in areas where energy resources are scarce or where the power infrastructure is already under significant pressure.

About Lubrizol

Lubrizol, is a science-based company headquartered in Ohio whose specialty chemistry delivers solutions to advance mobility, improve well-being and enhance modern life. Every day, the innovators of Lubrizol strive to create extraordinary value for customers at the intersection of science, market needs and business success, driving discovery and creating breakthrough solutions that enhance life and make the world work better. Founded in 1928, Lubrizol has global reach and local presence, with more than 100

[13] Data Centers Aren't Loud, Right? | Sensear

manufacturing facilities, sales and technical offices and more than 7,000 employees around the world. For more information, visit www.Lubrizol.com.

Disclaimer

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the AI Action Plan and associated documents without attribution.

Contact Information

Patrick Finnegan, Senior Director Government Affairs

[redacted]