International Research Journal of Multidisciplinary Scope (IRJMS), 2026; 7(2): 1351-1364

Review Article | ISSN (O): 2582-631X **DOI: 10.47857/irjms.2026.v07i02.08460**

Cooling Strategies for Energy-efficient Data Centers: Narrative Review and Future Perspective

Luca Fiorillo[1,2]*, Sharon Ridolfo[3], Attilio Caravelli[4], Dario Milone[5]

[1]Department of Medicine and Surgery, University of Enna Kore, Cittadella Universitaria, Enna, 94100, Italy, [2]Department of Dental Research Cell, Dr. D. Y. Patil Dental College & Hospital, Dr. D. Y. Patil Vidyapeeth (Deemed to be University), Pimpri, Pune 411018, India, [3]Institute of Environmental Assessment and Water Research, Spanish Research Council (IDAEA-CSIC), c/Jordi Girona 18-26 08034, Barcelona, Spain, [4]Reggio Calabria, 89100, Italy, [5]Department of Engineering, University of Messina, 98100 Messina, Italy.

*Corresponding Author's Email: lfiorillo@unime.it, lucafiorillo@live.it

Abstract

The exponential growth of data centers, driven by the proliferation of cloud computing, Artificial Intelligence (AI), and big data analytics, has significantly intensified global energy demand, with cooling systems accounting for a substantial proportion of overall consumption. This narrative review comprehensively examines both conventional and advanced cooling strategies designed to enhance energy efficiency and sustainability within data center infrastructure. The limitations of traditional air-based and chiller-centric methods are discussed, highlighting inefficiencies in high-density computing environments. The manuscript then details alternative cooling technologies, including free cooling, Direct-to-Chip liquid cooling, Immersion Cooling (IC), and emerging modalities such as underwater and space-based data centers. Real-world case studies, regulatory frameworks, and sustainability initiatives from Europe, North America, Asia, and emerging regions are analyzed to illustrate the operational and environmental benefits of each approach. Furthermore, the comparative environmental impact of AI-assisted versus traditional human manuscript production is briefly assessed. The review concludes by advocating for a multi-faceted integration of cooling strategies, regulatory compliance, and technological innovation to achieve low Power Usage Effectiveness (PUE), minimize water consumption, and foster a sustainable digital infrastructure. The review synthesizes evidence from peer-reviewed literature, institutional reports, and industry case studies published between 2005 and 2025, evaluating key performance metrics including Power Usage Effectiveness and Water Usage Effectiveness. Regional regulatory frameworks from Europe, North America, and Asia are comparatively analyzed. Future directions, including underwater and space-based data center concepts, are also discussed.

Keywords: Cooling Strategies, Data Centers, Energy Efficiency, Immersion Cooling, Power Usage Effectiveness, Sustainability.

Introduction

Data centers have become essential infrastructure for the digital economy, supporting a wide range of services, including cloud computing, Artificial Intelligence (AI), and other advanced technologies. However, they consume enormous amounts of electricity. A pivotal study found that while data center compute output rose sharply, global data center energy consumption only increased approximately 6% from 2010 to 2018, reaching about 205 TWh (roughly 1% of global electricity), thanks to aggressive efficiency gains in Information Technology (IT) and cooling equipment (1). Nonetheless, current trends such as AI and edge computing are accelerating energy demand. Recent projections indicate that data centers could account for about 2% of worldwide electricity use by 2025 (over 500 TWh) and potentially double that by 2030 if efficiency improvements lag behind demand (1, 2). This massive energy consumption not only incurs high operating costs but also translates to a significant carbon footprint when fossil-based power is used. Moreover, data centers require large volumes of water and land, and their cooling needs pose major sustainability challenges (3).

A substantial portion of a data center's energy intake is expended not on computing, but on removing the heat generated by servers. Traditional air conditioning and chiller systems can consume 30-40% or more of a facility's total energy (3, 4). The metric Power Usage Effectiveness (PUE), defined as total facility power divided by IT equipment power, encapsulates this overhead. A perfect PUE of 1.0 means all energy is directed toward computing. In practice, early 2000s data centers often had PUE values of 2.0 or

(Received 30th September 2025; Accepted 26th March 2026; Published 23rd April 2026)

higher. Thanks to industry efforts, the global average PUE has improved to around 1.5-1.8 in recent years (5, 6). However, many legacy facilities still operate above 1.7 (3, 7). State-of-the-art hyperscale data centers can achieve a PUE of approximately 1.1 or even lower, approaching the thermodynamic and practical limits of efficiency. Achieving such efficiency across the board is imperative as the data center industry strives to minimize its environmental impact amid growing demand.

Cooling is a critical focus for efficiency gains, not only because of its significant share of energy use, but also due to its implications for water consumption and hardware reliability. Conventional cooling methods, including Computer Room Air Conditioner (CRAC) units, chillers, and cooling towers, are inherently energy-intensive and often rely on water evaporation for heat rejection. An average data center can use approximately 1.1 million liters of water per day for cooling purposes, raising serious concerns in water-stressed regions (5). The increasing scale and density of modern computing hardware, particularly AI training clusters and GPU-intensive workloads, have exacerbated this challenge by generating significantly higher heat loads per unit area than traditional enterprise servers. As public awareness of these impacts grows, there is increasing pressure on data center operators to adopt more sustainable practices. In response, both industry and regulators have introduced targets and standards for energy and cooling efficiency, and researchers have developed a range of advanced cooling technologies.

The environmental footprint of data centers extends beyond electricity. Water consumption for cooling purposes has emerged as a critical sustainability concern, particularly in water-stressed regions where large-scale evaporative cooling competes with municipal and agricultural water needs. Additionally, the carbon emissions associated with data center electricity use are substantial when fossil-fuel-based power generation is involved. The growing recognition of these interconnected challenges of energy, water, and carbon has prompted a global response from policymakers, industry consortia, and researchers, leading to the development of advanced cooling technologies and regulatory frameworks designed to drive efficiency improvements across the sector (1, 3, 8). The urgency of this transition is underscored by the rapid expansion of AI workloads, which demand denser and hotter computing hardware that pushes conventional cooling approaches to their physical limits (5, 7, 9). This narrative review synthesizes the current state of data center cooling technologies and emerging alternatives that aim to reduce energy and water usage. The review outlines the limitations of traditional cooling approaches and the thermal management challenges posed by modern high-density IT hardware. Alternative cooling solutions, from air-side and water-side free cooling to liquid and Immersion Cooling (IC), are surveyed, and novel concepts such as underwater and space-based data centers are discussed. Both the direct energy consumption by IT equipment and the ancillary energy used for cooling are taken into account. Sustainability initiatives and regulations worldwide are examined, and future directions, including the concept of off-planet data centers, are proposed as potential pathways toward truly sustainable computing infrastructure. By compiling over 25 authoritative sources, this review aims to provide a comprehensive and up-to-date overview of how the data center industry can maintain critical IT services while minimizing environmental impact (1, 3, 5, 9).

Methodology

This study was conducted as a narrative review of the existing literature on data center cooling technologies, energy efficiency, and sustainability. A narrative review approach was chosen because the topic spans multiple disciplines, including engineering, environmental science, information technology, and policy, and the aim was to provide a broad, integrative synthesis rather than a narrowly focused quantitative meta-analysis. While narrative reviews do not follow the structured protocol of systematic reviews (such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, PRISMA), the present work adopted a rigorous and transparent literature search strategy to ensure comprehensiveness and reliability.

Literature searches were performed across multiple databases, including PubMed, Scopus, Web of Science, and Google Scholar, using combinations of keywords such as "data center cooling," "Power Usage Effectiveness," "liquid cooling data center," "Immersion Cooling," "free cooling data center," "data center energy efficiency," "underwater data center," and "space-based data center." Searches were conducted from January 2005 to March 2025 to capture both foundational and recent literature. No language restrictions were applied, although the majority of retrieved sources were in English.

The selection of sources prioritized peer-reviewed research articles, review papers, books, and book chapters. In addition, authoritative institutional reports from organizations such as the International Energy Agency (IEA), the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), the European Commission, and government agencies were included to capture regulatory and policy dimensions. Industry white papers and technical reports were included only when peer-reviewed equivalents were unavailable and the information was critical for completeness.

The quality of included studies was appraised using the Critical Appraisal Skills Programme (CASP) checklist for qualitative and review studies, adapted as appropriate for technical and policy documents. Studies were assessed for clarity of objectives, appropriateness of methodology, relevance of findings, and applicability to the review's scope. Sources that did not meet minimum quality thresholds or that presented unsubstantiated claims were excluded.

Data extraction focused on key metrics such as PUE values, energy savings percentages, water usage figures, cooling capacity, and regulatory targets. The extracted data were organized thematically and synthesized narratively under the headings presented in this review. No statistical meta-analysis was performed, consistent with the narrative review design.

Review

Energy Demand and Cooling Challenges in Modern Data Centers

Global data center energy usage has grown significantly over the past two decades, driven by the rapid expansion of internet services, cloud computing, and, more recently, the increasing demand for AI workloads. However, efficiency improvements in servers, power distribution, and cooling have helped temper the growth in total energy use (1). A pivotal analysis found that global data center energy consumption only rose approximately 6% from 2010 to 2018, reaching about 205 TWh in 2018, roughly 1% of global electricity, thanks to aggressive efficiency gains (1). Nonetheless, current trends such as AI and edge computing could accelerate energy demand, making continued efficiency advancements crucial. By 2030, in a pessimistic scenario without new efficiency breakthroughs, data centers worldwide could draw over 1,000-1,300 TWh (2-3% of projected global electricity) underscoring the urgency for improved energy management (1, 2).

PUE has become the de facto standard for measuring the energy efficiency of data centers. Globally, the average PUE reported by operators from 2021 to 2024 hovers around 1.5-1.6, meaning approximately 60% of energy is delivered to IT equipment and 40% is consumed by cooling and other facility overhead. There is considerable variability: older and smaller data centers often have PUE above 2.0, whereas hyperscale cloud facilities built with cutting-edge cooling can achieve PUE below 1.1 (10, 11). For example, one major technology company reports a fleet-wide PUE of 1.10 across its large data centers, and a 2013 facility in Lulea, Sweden, which leverages free cooling and streamlined design, operates at a PUE of approximately 1.07 (12). These low PUE values approach the practical minimum (1.0), yet a gap remains between best-in-class and typical facilities.

Cooling systems are often the single most significant factor in PUE above 1.0. Traditional data center cooling infrastructure includes CRAC or Computer Room Air Handler (CRAH) units, refrigerant-based chillers, cooling towers, pumps, and fans to circulate cold air. In a typical legacy cooling architecture, chillers cool water which is pumped into air handlers that blow cold air under a raised floor or through contained cold aisles; the hot air leaving servers is captured, re-cooled via heat exchangers, and the cycle repeats (4). Such systems are reliable and have long been standard, but they suffer from several inefficiencies. The cooling process often involves multiple energy conversions, such as compressors, pumps, and fans, that cumulatively draw a large amount of power. It is not uncommon for cooling equipment to account for approximately 40% of a data center's total energy use in a conventional setup (4,13). One breakdown showed that in a data center where cooling accounts for approximately

40% of the facility's energy, the chiller alone consumes approximately 16%, with pumps and cooling tower fans using another 12% each (4).

Additionally, traditional air-based cooling struggles to handle the rising heat densities of modern IT hardware (13). Today's high-performance servers and AI accelerators can consume 300-500 W per Central Processing Unit (CPU) or Graphics Processing Unit (GPU), and rack power densities in cutting-edge deployments have jumped from approximately 5-10 kW per rack a decade ago to 30-80 kW per rack for AI training clusters (7). Air cooling becomes less effective at these heat flux levels, as large volumes of air must be moved at high speed, leading to disproportionately higher fan energy and difficulty in maintaining uniform temperatures. Hot spots and recirculation of hot air can occur if airflow is not perfectly managed (4, 13). Even with techniques like hot-aisle/cold-aisle containment, conventional Heating, Ventilation, and Air Conditioning (HVAC) approaches are nearing their practical limits in such high-density scenarios.

Another challenge is climate dependence. Traditional systems using vapor-compression refrigeration operate year-round, even in cold climates where outdoor air can provide free cooling. Many older facilities were not designed to utilize outside cool air or water, and continue to operate power-hungry chillers even on cool days. Where economizers are present to utilize outside air, other issues arise: bringing outside air into server rooms risks introducing dust, corrosive pollutants, or humidity that can damage equipment. Thermal environmental requirements have also historically necessitated low temperatures. Earlier guidelines often recommended server inlet temperatures of 20-25 degrees Celsius with narrow humidity bands, prompting heavy cooling efforts. In recent years, ASHRAE relaxed its recommendations, with typical allowable inlet temperature now ranging up to 27 or even 32 degrees Celsius for certain classes of IT equipment. This broader envelope means facilities can be run a bit warmer to save energy, although not all legacy data centers have adopted these settings (4).

The consequence of these factors is evident in power overhead statistics: many data centers still require on the order of 0.5 kW of cooling per 1 kW of IT load (PUE of approximately 1.5), and in inefficient cases up to 1-1.5 kW of cooling per 1 kW IT (PUE 2.0-2.5) (4). This excess energy not only incurs cost and carbon emissions but also places strain on electrical grids. Moreover, water-cooled chillers and cooling towers consume large amounts of water through evaporation, raising sustainability issues. In regions facing drought, the water footprint of data center cooling has sparked community concerns and even led to project permit delays (5, 12).

Traditional Cooling Methods and Limitations

In a conventional data center design, air conditioning units chill the air, which is then circulated through the server room, absorbing heat from the equipment, and subsequently re-cooled in a continuous cycle (4). The components of a standard cooling loop include CRAC or CRAH units, which contain coils that cool air using chilled water or refrigerant and blow it into the raised floor or cold aisle, with multiple units distributed across the data hall for redundancy and coverage (13). Chillers, which are large refrigeration machines (centrifugal or compressor-based), remove heat from the water circulating in the CRAH coils, rejecting that heat to an external condenser loop. Chillers are energy-intensive, and their compressors draw significant power, especially in warm weather. In many legacy facilities, chillers run continuously at partial load even when not strictly necessary. Cooling towers or dry coolers evaporate water to dissipate heat from the condenser water loop into the outside air, leading to water loss requiring makeup water and necessitating chemical treatment. Pumps circulate chilled water to CRAH units and condenser water to cooling towers, while large fans drive airflow through CRAHs and across cooling coils (4). These ancillary devices consume a non-trivial amount of electricity on a continuous basis.

One major limitation of this approach is the mixing of hot and cold air. Unless physical separation is enforced via containment systems, a portion of the hot exhaust air from servers inevitably recirculates and mixes with the cold supply air, raising its temperature and requiring the cooling system to work harder (4). Older data centers without aisle containment suffer significant inefficiencies as a result. Even with effective airflow management, the temperature difference that air cooling can sustain is limited, and very high-density racks may overwhelm the cooling airflow, resulting in local hotspots.

Another inherent inefficiency is that air has a very low heat capacity and thermal conductivity compared to liquids. Large volumes of air must be moved to carry away heat, requiring powerful fans. Those fans not only consume electricity but also add heat to the air, which the cooling system must then remove. As server power increases, the volume and velocity of air needed rise superlinearly, driving up fan energy usage. In contrast, liquid cooling can carry much more heat per volume flow, highlighting why a paradigm shift is underway for high-performance systems (5, 9, 14).

Traditional cooling also tends to over-provision capacity for worst-case conditions (peak IT load on the hottest day of the year). Most of the time, neither the IT load nor the weather hits that extreme, meaning chillers and CRACs operate at part load, where they are often less efficient. Without intelligent control, this results in a poor Coefficient of Performance (COP) during off-peak times. The deployment of AI-driven adaptive cooling controls has emerged as a prominent example of addressing this issue. By analyzing historical data and real-time sensor inputs, an AI system was reported to achieve a 40% reduction in cooling energy at a major technology company's data centers (9). This demonstrates that even within traditional cooling setups, smarter operation can trim waste. Nevertheless, there are physical limits to air cooling efficiency at scale, and more radical changes in cooling approach are being pursued.

Alternative and Advanced Cooling Technologies

Air-Side and Water-Side Free Cooling: One of the simplest and most effective ways to reduce cooling energy is to utilize the environment's natural cooling capacity, a technique commonly referred to as free cooling. Free cooling means reducing reliance on compressors by using naturally cool air or water from outside the data center to dissipate heat. There are two primary modes: air-side free cooling, which brings outside air into the facility to cool the IT equipment, and water-side free cooling, which uses an external water loop or a nearby water source to absorb heat via a heat exchanger (15, 16).

In cool climates or seasons, free cooling can drastically reduce or even eliminate the need for traditional air conditioning. For instance, a major data center in Lulea, Sweden, takes advantage of the sub-Arctic air, achieving a PUE averaging approximately 1.07, with essentially only the fans and minimal additional cooling overhead beyond the IT load (12). Similarly, many Nordic data centers and others in temperate climates have reported PUE in the 1.1-1.3 range, primarily due to the use of free cooling, which leverages low ambient temperatures and inexpensive hydroelectric power.

Direct air-side free cooling involves using outside air as the supply air to the servers. Large dampers or louvers bring in cool outside air, which passes through filters and humidity conditioning systems. The primary energy cost is the operation of fans; no chiller is required as long as the outside air temperature and humidity are within acceptable ranges. The approach is highly energy-efficient when conditions are right, but it requires clean air, as contaminants in outside air can damage sensitive electronics if not filtered (17). Furthermore, outside air humidity must be controlled, as excessively dry air can cause static electricity issues, while excessive moisture can lead to condensation or corrosion. ASHRAE's recommended allowable humidity ranges must be respected (4,12). Due to these concerns, indirect air-side free cooling is often preferred, where heat is transferred via a heat exchanger without direct mixing of outdoor and indoor air (4, 16).

Water-side free cooling uses water as the heat transfer medium. A prominent example is a data center in Hamina, Finland, which pumps cold water from the Baltic Sea to cool its servers, taking advantage of the sea's thermal mass to absorb heat without chillers (12). The limitation is geography: only facilities near a large cold water body can utilize this method, and environmental regulations may restrict the amount of heat discharged. Combining air-side economization with evaporative cooling can reduce overall data center energy use by 20-50% in favorable climates (15). A case study with indirect free cooling plus Thermal Energy Storage (TES) showed energy savings of approximately 33% and achieved a PUE of approximately 1.2 (4). However, in tropical or hot-humid climates, free cooling opportunities are limited, and alternative techniques such as liquid cooling become more attractive.

Liquid Cooling (Direct-to-Chip and Cold Plate Cooling): As server power densities have risen, the industry has increasingly turned to liquid cooling as a superior medium for heat removal. Liquids such as water or special coolants have far higher thermal conductivity and heat capacity than air, enabling more effective cooling of hot electronics (5, 9). Direct-to-Chip cooling, also known as cold plate cooling, is one widely adopted method. In a Direct-to-Chip system, each major heat-generating component, such as the CPU or GPU, is outfitted with a water-cooled cold plate, a metal block with internal microchannels through which coolant flows. Heat is conducted from the chip into the cold plate and then into the circulating liquid, which exits the server chassis and connects to a heat exchange system that transfers the heat to facility water or a secondary cooling loop (13).

The advantages of Direct-to-Chip liquid cooling are significant. First, it allows much higher power densities per rack because water can carry away heat far more efficiently. One deployment of cold plate liquid cooling in a Chinese mega-data center enabled racks to support 65 kW of IT load each while maintaining a very low PUE around 1.05 (4). Air cooling would be hard-pressed to achieve such density without overheating. Second, liquid cooling can often use warmer coolant temperatures than air cooling requires. Systems using water at approximately 45 degrees Celsius can still effectively remove heat from CPUs, and crucially, the water loop can be cooled by dry coolers without the need for a chiller. Such warm-water cooling enables a reported 40% reduction in total power consumption for cooling and a 3.5 times improvement in cooling efficiency compared to legacy air systems (5). Third, Direct-to-Chip liquid cooling reduces or eliminates the need for server internal fans, which consume power (often 5-10% of server power in air-cooled, high-density servers), improving reliability and avoiding thermal throttling.

Challenges include concerns about leaks, as water dripping onto electronics could be catastrophic. Modern solutions utilize dripless quick-disconnect fittings and employ dielectric liquids or treated water. There is also added complexity of plumbing, with racks requiring supply and return manifolds, pumps, and heat exchangers, which adds capital cost and maintenance overhead compared to simple air-cooled racks (5, 9). Cold plate cooling is seeing rapid adoption for High-Performance Computing (HPC) and AI, and it has been estimated that by 2026, up to 40% of data centers may incorporate some form of liquid cooling (7). An alternative approach is Rear-Door Heat Exchangers (RDHx), where the back door of a server rack is replaced with a radiator panel. As hot air from the servers exits, it passes through this water-cooled radiator, removing 50-80% of the heat by liquid and easing the burden on room air conditioning. While not as efficient as Direct-to-Chip, RDHx is easier to implement on existing racks (13).

Immersion Cooling (IC): It involves submerging the entire server or components in a bath of dielectric fluid. In IC, servers are placed in sealed tanks filled with a special non-conductive liquid that exhibits good heat absorption properties. The fluid directly contacts all heat-generating components, cooling them very effectively. There are two main types: single-phase immersion, where the fluid never changes phase and simply convects to carry heat away from components to the tank's heat exchanger, and two-phase immersion, which utilizes a fluid that boils at a low temperature, carrying away latent heat as vapor that condenses on a cooled coil at the top of the tank (3, 18).

IC offers several compelling benefits. Regarding energy savings, immersion systems often eliminate server fans, reducing IT power consumption. Cooling energy is primarily provided by small pumps for the water loop and fluid circulation. A well-designed immersion-cooled data center can achieve a cooling PUE of close to 1.02-1.05, meaning virtually all energy is directed to IT, with only approximately 2-5% overhead for cooling. Industry reports indicate that IC can reduce electricity use for cooling by up to 90-95% compared to traditional air cooling. For example, one vendor reported cases where immersion cut total data center power by 10-20%, equating to 50-70% reduction in cooling energy, and nearly eliminated water usage (3).

Additionally, IC provides high thermal uniformity and capacity. Components submerged in fluid experience very uniform temperatures with minimal hotspots, as the fluid reaches all around circuit boards, cooling not just CPUs but also GPUs, memory, and voltage regulators that air cooling may struggle with. This uniform cooling can increase hardware reliability by reducing thermal stress and allow higher clock speeds before reaching thermal limits (18). IC also achieves zero water evaporation through closed water loops and dry coolers, dramatically improving Water Usage Effectiveness (WUE) and potentially achieving zero onsite water use. Space savings and noise reduction are additional benefits (3, 18).

Challenges associated with IC include the expense and maintenance requirements of the dielectric fluids, the messiness of mineral oil if used, and the need for specially designed chassis or enclosures. Two-phase fluids raise concerns about fluid loss and environmental impact, some perfluorocarbon coolants have a very high Global Warming Potential (GWP) if released. Despite these challenges, IC has been gaining traction. A major cloud provider announced in 2021 that it was running some production cloud servers in a two-phase immersion tank to support high-density workloads, observing no hardware failures attributable to immersion. In China, one facility achieved a PUE of 1.09 using advanced cooling techniques including immersion (4). Several specialist vendors now offer turnkey IC solutions, and the Open Compute Project (OCP) has a working group on advanced cooling solutions, indicating broad industry interest (3, 18).

Other Novel Cooling Approaches: Beyond the mainstream alternatives described above, a number of innovative cooling techniques are being explored or have niche implementations. Two-phase cold plates or refrigerant cooling systems use refrigerant in closed loops that boil at the chip and condense in a radiator, achieving even higher heat transfer coefficients than water while keeping components cool. Heat pipes and vapor chambers are increasingly used on motherboards to passively spread and move heat from hot spots to cooler areas (9).

Thermal Energy Storage (TES) can augment cooling systems to reduce peak energy use by storing cooling potential during off-peak times, for example by making ice or chilling water at night when temperatures are lower or electricity is cheaper. A data center can then utilize the stored cold during the day, running compressors less frequently. Studies have shown that coupling free cooling with TES extends the hours of free cooling, with one such system achieving 33% energy savings (4). TES contributes to sustainability by load-shifting and enabling better use of Renewable Energy Sources (RES) (16).

Radiative cooling is a cutting-edge research area involving the use of the coldness of outer space as a heat sink through special materials, such as nano-engineered photonic surfaces, that can efficiently radiate infrared heat to the sky. A recent study designed radiative cooler panels for data centers and suggested they could help reduce cooling load when integrated with traditional systems (19). However, radiative cooling is strongly weather-dependent and currently provides only a supplemental effect. Geothermal and underground cooling, involving the placement of data centers underground or using geothermal resources, offers stable, cooler temperatures and can ease cooling needs (15). Waste heat reuse, though not a cooling technology per se, improves sustainability by utilizing the heat that cooling systems remove for productive purposes such as heating buildings or industrial processes. In Scandinavian countries, it is becoming increasingly common for large data centers to feed waste heat into District Heating (DH) networks. The European Union and local governments have started incentivizing heat reuse, with Germany initially considering requiring new data centers to reuse a portion of their waste heat (20). Warm water coming out of cold plates at approximately 60 degrees Celsius is quite suitable for feeding into a DH loop, creating a synergy with liquid cooling that could become a standard element of sustainable data center design in urban areas (13).

Sustainability Initiatives and Regulatory Perspectives

Governments and industry organizations around the world have introduced measures to improve data center sustainability. These measures range from voluntary efficiency standards and reporting frameworks to binding regulations that mandate performance targets. Table 1 provides a summary of selected data center energy efficiency policies and targets in various jurisdictions. Table 1 presents regulatory frameworks from eight distinct regions, illustrating the global convergence toward stricter energy efficiency standards for data centers. Key metrics include PUE thresholds, renewable energy requirements, and heat reuse mandates, with implementation timelines ranging from 2022 to 2030. Notable variations exist between regions, with European jurisdictions generally adopting more stringent and binding targets compared to voluntary frameworks elsewhere.

In Europe, the European Union (EU) has been proactive in addressing data center sustainability. Under the EU's revised Energy Efficiency Directive (EED) of 2023, data centers above 500 kW IT load are required to measure and report their energy performance, including PUE, Water Usage Effectiveness (WUE), and use of waste heat. Initially, the focus is on transparency, with the EU poised to set efficiency targets in the coming years based on collected benchmarks. A coalition of industry players in Europe launched the Climate Neutral Data Centre Pact (CNDCP) in 2021, a voluntary agreement committing to ambitious goals by 2030, including achieving PUE of 1.3 or lower in cool climates, 100% carbon-free power, and exploring heat reuse. Germany enacted the Energy Efficiency Act (Energieeffizienzgesetz) in 2023, which sets concrete PUE limits: existing data centers must reach a PUE of 1.5 or lower by 2027 and 1.3 or lower by 2030, while any new data center operational from 2026 onwards must achieve a PUE of 1.2 or lower (20). Germany initially considered mandating at least 30% waste heat reuse by 2025, later relaxed to 20% by 2028 following industry consultation. The Scandinavian countries have generally incentivized green data centers through energy tax breaks and support for heat reuse.

The United States at the federal level enacted the Data Center Optimization Initiative (DCOI), requiring federal government data centers to maintain PUE of 1.5 or lower for existing facilities and 1.4 or lower for new data centers, with mandated metering and regular reporting (21). At the state level, proposals have been introduced in Virginia, Oregon, and Washington regarding energy efficiency and environmental rules for data centers. In China, the Ministry of Industry and Information Technology (MIIT) set a target of PUE of 1.3 or lower for large cloud data centers by end of 2023, with some provinces such as Guangdong mandating PUE of 1.25 or lower for new projects by 2025. China is also exploring the integration of Liquefied Natural Gas (LNG) infrastructure for cooling, with a pilot project in Guangdong utilizing the cold energy from LNG regasification to cool a data center, reportedly reducing cooling electricity consumption by over 50% (4).

Singapore, after freezing new data center construction from 2019 to 2022, now allows new builds only under strict criteria: PUE of 1.3 or lower, use of innovative cooling such as IC, and plans to offset carbon. Only proposals meeting these sustainability criteria receive approval (22).

Emerging regions have also begun to address data center sustainability. In India, the Bureau of Energy Efficiency (BEE) has started developing energy performance benchmarks for data centers, and the rapid growth of digital infrastructure in Southeast Asia and Africa is prompting the development of region-specific efficiency guidelines, although formalized regulatory frameworks remain less mature compared to those in Europe or North America. In Latin America, countries such as Brazil and Chile are beginning to incorporate data center energy efficiency into broader energy policy frameworks, driven by the growth of cloud computing hubs in Sao Paulo and Santiago.

From these trends, a clear pattern emerges: governments in both the East and West are setting PUE thresholds that essentially force the adoption of advanced cooling technologies. A PUE of 1.3 or 1.2 for large facilities cannot be effectively met with old-style CRAC and chiller designs alone, and likely requires extensive economization, liquid cooling, or other innovative measures. The push for heat reuse means data centers must plan for their waste heat streams, which incentivizes liquid cooling since it is easier to capture heat in water than from hot air exhaust. These policies reinforce the business case for the alternative cooling strategies detailed above.

Table 1: Selected Data Center Efficiency Targets and Regulations in Different Regions

Region	Policy/Initiative	Key Requirements	Timeline
European Union	EU Energy Efficiency Directive (2023)	Mandatory energy reporting for data centers >500 kW; targets for PUE, renewable energy, and heat reuse (20).	Reporting from 2025; targets 2027-2030
Germany	Energy Efficiency Act (2023)	Existing DCs: PUE <=1.5 by 2027, <=1.3 by 2030. New DCs (from 2026): PUE <=1.2. Heat reuse: 10% by 2024, rising to 20% by 2028 (20).	2024-2030 (phased)
Netherlands/Ireland	Policy moratoriums	Temporary moratoriums on new mega-data centers (2019-2022); now require sustainability plans.	Ongoing

United States (Federal)	DCOI (2016, updated 2019)	Federal DCs: existing PUE <=1.5; new PUE <=1.4. Mandated metering and reporting (21).	FY2018 (ongoing)
China	MIIT Guidelines	Large cloud DCs: PUE <=1.3 by 2023. Some provinces mandate PUE <=1.25 for new projects by 2025 (4).	2023-2025
Singapore	Sustainable DC Pilot (2022)	New builds: PUE <=1.3, innovative cooling (e.g. IC), renewable energy (22).	Pilot from 2022; ongoing
India	BEE benchmarking (developing)	Energy performance benchmarks under development for data centers.	In progress
Global Industry	CNDCP (2021)	Voluntary targets: PUE <=1.3 cool climate, <=1.4 temperate, <=1.5 warm climate by 2030; 100% renewable energy (20).	2030 goals

Underwater and Space-based Data Centers

Two of the most unconventional concepts in data center cooling are underwater and orbital (space-based) data centers. In 2018, a major technology company deployed a self-contained data center module on the seafloor off the coast of Orkney, Scotland. The cylindrical pod housed 864 servers and related cooling infrastructure, all sealed from the water. The ocean provided a vast cold sink: seawater-circulated heat exchangers on the vessel's exterior dissipated the server heat into the sea, eliminating the need for chillers. Over a two-year test, the underwater data center performed remarkably well, with the servers experiencing one-eighth the failure rate of an identical batch on land (23). The reasons include the stable cold environment and the absence of human interference, including no dust, no oxygen to corrode components, and no accidental disturbance of equipment. The pod was also powered entirely by locally generated renewable energy.

Regarding space-based data centers, the proposition involves placing computing infrastructure in orbit around Earth, leveraging the 3 Kelvin background of space as the ultimate heat reservoir and abundant solar energy that is available nearly continuously without atmospheric loss. Recent efforts by startups have included testing prototype computing payloads on the International Space Station and developing satellites with high-density compute modules and large radiators to dissipate heat. One company claims its design offers substantially more GPU compute than any prior spacecraft and can run workloads in orbit with latency of approximately 20 milliseconds if in Low Earth Orbit (LEO) (24). The practical viability, scalability, and economic feasibility of these concepts remain subjects of ongoing investigation.

Discussion

The findings of this review indicate that the data center industry is undergoing a fundamental transition in cooling paradigms, driven by the twin pressures of escalating heat densities and intensifying sustainability requirements. The evidence strongly supports that no single cooling strategy can address all operational contexts; rather, a portfolio approach tailored to local climate, facility scale, workload characteristics, and regulatory environment is required (3, 9, 15).

Traditional air-based cooling, while still dominant in the installed base, is reaching its practical limits for high-density computing. The inefficiencies identified, including hot-cold air mixing, low heat capacity of air, and over-provisioning for worst-case scenarios, collectively contribute to PUE values that are increasingly unacceptable under modern regulatory standards. The evidence from AI-driven cooling optimization, which demonstrated a 40% reduction in cooling energy, suggests that significant gains remain possible even within traditional setups. However, these gains are incremental and cannot overcome the fundamental thermodynamic limitations of air as a cooling medium for heat loads exceeding approximately 20-30 kW per rack (9).

Liquid cooling and IC represent the most promising pathways for high-density deployments. The PUE values of 1.05-1.1 achieved with Direct-to-Chip cooling and the 90-95% reduction in cooling energy reported for IC are compelling (3, 4). However, several trade-offs and implementation barriers warrant critical examination. The capital expenditure for liquid cooling infrastructure, including manifolds, pumps, heat exchangers, and specialized plumbing, is substantially higher than for traditional air-based systems. Lifecycle Environmental Impact (LEI) assessments remain incomplete: while operational energy savings are well documented, the embodied energy and environmental costs of manufacturing dielectric fluids, particularly fluorocarbons with high GWP, have not been comprehensively evaluated. The end-of-life disposal or recycling of these fluids also poses challenges that the industry has yet to fully address (3, 18).

The regulatory landscape analysis reveals a convergence of expectations across jurisdictions, with PUE targets in the range of 1.1-1.3 becoming the norm for new large-scale facilities. However, the impact of these regulations on technology adoption varies significantly. In Europe, where binding targets are most advanced, operators have a clear incentive to deploy liquid cooling and heat reuse systems. In the United States, where the approach has been more voluntary outside federal agencies, market forces and corporate sustainability commitments, rather than regulation, have driven adoption among hyperscale operators. In emerging regions such as India, Southeast Asia, and Africa, the absence of mature regulatory frameworks creates both a risk and an opportunity: without efficiency mandates, operators may default to cheaper traditional cooling, but the greenfield nature of many installations in these regions offers a chance to leapfrog directly to advanced technologies (4, 22).

The novel concepts of underwater and space-based data centers merit a critical perspective. The underwater data center concept, validated by a major two-year experiment off the coast of Scotland demonstrated remarkable reliability (one-eighth the failure rate of comparable terrestrial servers) and eliminated the need for mechanical cooling (23). However, the scalability and economic viability of this approach remain uncertain. Deployment and retrieval of underwater modules is complex and expensive, maintenance is highly constrained, and the environmental impacts of continuous heat discharge into marine ecosystems, while assessed as negligible for single modules, require further study for large-scale deployments. Similarly, space-based data centers, while theoretically offering unlimited cold sink and solar power, face formidable challenges: launch costs remain high despite recent reductions, radiation damage to electronics, latency constraints for distant orbits, and the absence of on-orbit maintenance capabilities all represent significant barriers. Claims of operational costs 97% lower than terrestrial facilities rely on highly optimistic assumptions about launch costs and hardware longevity that have not yet been empirically validated (24). These concepts are best characterized as experimental pathways that may find niche applications rather than near-term replacements for terrestrial infrastructure.

The findings of this review are consistent with several recent analyses that underscore the urgency of transitioning to advanced cooling paradigms. A study demonstrated that data center water consumption is a growing concern that has been systematically underestimated in sustainability assessments, particularly in regions where evaporative cooling remains the dominant heat rejection method (8). This aligns with the present review's identification of water usage as a critical sustainability dimension alongside energy efficiency. As found in past research, a comprehensive evaluation of liquid cooling technologies concluded that direct-to-chip and immersion cooling systems offer the most promising pathways for managing the thermal loads of next-generation AI accelerators, a finding that corroborates the evidence presented in this review regarding PUE values of 1.05–1.1 for liquid-cooled deployments (14). Furthermore, a past research provided statistical evidence that location-specific factors, including climate, grid carbon intensity, and local regulations, play a decisive role in determining achievable PUE values, reinforcing the portfolio approach advocated in this review (6). The convergence of regulatory pressures, particularly in Europe and Asia, with technological maturation of liquid cooling solutions suggests that the industry is approaching an inflection point where advanced cooling will transition from a competitive advantage to a baseline requirement. However, the economic barriers to retrofitting existing facilities remain substantial, and the disparity between hyperscale operators and smaller enterprise data centers in terms of cooling efficiency is likely to persist without targeted policy interventions or financial incentives (4, 7, 20).

The AI versus human manuscript production comparison presented in this review offers a provocative illustration of differential environmental footprints of different knowledge production methods. As summarized in Table 2, AI-assisted manuscript generation consumes approximately 200-400 times less energy than traditional human research workflows. Table 2 compares five key environmental and operational parameters between AI-driven and traditional human manuscript production, including active writing time, energy consumption, carbon dioxide emissions, operational continuity, and ancillary resource usage. The data highlight a significant disparity in energy intensity, with AI models completing equivalent tasks in a fraction of the time and with substantially lower carbon output. However, these figures should be interpreted with caution, as they do not account for the cumulative energy cost of AI model training. While this comparison is necessarily simplified, omitting factors such as the training energy of AI models and the irreplaceable value of human critical thinking, it highlights the potential for AI tools to reduce the environmental costs of scholarly activities when used as adjuncts to human oversight (25). This is thematically relevant to the review's focus on energy efficiency, as it demonstrates that AI's environmental impact extends beyond data center operations to the broader ecosystem of knowledge production.

Table 2: Comparative Environmental Impacts of Manuscript Generation by AI and Human Researchers

Factor	AI Model	Human Researchers (4 individuals)
Active Writing Time	Approximately 1.5 hours	Approximately 3 weeks (60-90 hours)
Energy Consumption	40-60 Wh (0.04-0.06 kWh)	12,000-18,000 Wh (12-18 kWh)
CO2 Emissions	Approximately 20-30 g CO2	Approximately 6,000-9,000 g CO2 (6-9 kg)
Operational Continuity	Continuous	Intermittent with pauses
Ancillary Resource Usage	Cloud GPUs	Laptop computers, internet, software

Conclusion

The challenge of efficiently cooling data centers in an era of unprecedented compute demand is being met with a wave of technological innovation. This narrative review has explored how the industry is moving beyond energy-intensive air cooling and embracing a spectrum of alternative cooling strategies, from harnessing cool ambient air and water, to adopting liquid-based cooling at the server level, to immersing entire systems in dielectric fluids. These approaches are not merely theoretical: real-world implementations have demonstrated substantial gains. Air-side and water-side economization already enable PUE improvements that save significant percentages of energy; liquid cooling installations are achieving PUE values around 1.1 or below, even in extensive facilities; and immersion-cooled systems have shown cooling energy reductions on the order of 90%, while potentially eliminating water usage. The benefits extend beyond energy, encompassing improved reliability, higher compute density, and opportunities to reuse waste heat.

Equally important, the optimal cooling strategy depends on local climate, facility scale, workload characteristics, and regulatory environment. A data center in Finland may rely on free cooling from cold air and seawater, whereas one in a hot climate might deploy IC. The confluence of industry initiatives and regulatory pressure is accelerating the adoption of these technologies. Efficiency targets, such as a PUE of 1.3 or lower by 2030 in the EU or 1.2 for new builds in Germany, send a strong signal that future data centers must incorporate advanced cooling and holistic energy management from the outset. Achieving low-impact, high-efficiency cooling is a cornerstone of making digital infrastructure compatible with a sustainable future.

Limitations

This review has several limitations that should be acknowledged. First, as a narrative review, it does not follow the structured systematic review methodology of PRISMA, which limits the reproducibility of the literature search and selection process. Although a rigorous search strategy was employed and the CASP checklist was used for quality appraisal, the possibility of selection bias in the included sources cannot be entirely excluded. Second, the rapidly evolving nature of data center technologies and regulations means that some information presented here may become outdated shortly after publication, particularly regarding PUE targets and emerging regulatory frameworks in regions such as India and Africa. Third, the quantitative efficiency claims reported in this review, such as PUE values and energy savings percentages, are largely derived from manufacturer reports, industry white papers, and case studies rather than independently audited data. The generalizability of these figures across different facility sizes, climates, and workloads may be limited. Fourth, the lifecycle environmental impacts of advanced cooling technologies, including the embodied energy of manufacturing dielectric fluids and their end-of-life disposal, are not comprehensively covered due to a paucity of published lifecycle assessment studies in this domain. Fifth, the AI versus human manuscript comparison is necessarily simplified and does not account for the substantial energy invested in training AI models, which when amortized across uses would increase the per-task energy estimate. Finally, the discussion of space-based and underwater data centers relies heavily on early-stage experimental data and corporate projections, which may not reflect the actual performance and economics of these concepts at scale.

Future Prospects

The future of data center cooling is poised to be dynamic and innovative. Several key trends are likely to shape the field in the coming decade. First, wider adoption of liquid and IC in mainstream deployments is expected, moving beyond niche HPC applications to become the default for new high-density installations. Hybrid facilities that use a mix of free cooling, liquid cooling, and heat recovery, dynamically managed by AI for optimal efficiency under varying conditions, will become increasingly common.

Second, integration with external systems will deepen. Data centers are increasingly viewed as heat suppliers, grid balancers, and entities co-located with other industries to symbiotically use resources, as seen with the LNG-cooled data center pilot in China, which cuts cooling power by half by using otherwise wasted cold energy. Third, geographical diversification will continue, with data centers placed in locations optimized for efficiency, including cold climates, underground facilities, and sites near RES, a trend already visible in the concentration of cloud infrastructure in Nordic and Arctic regions.

Fourth, materials science advances, including the development of better thermally conductive materials, nano-coatings for corrosion prevention, and high-temperature electronics, may reduce cooling requirements. Emerging semiconductor technologies such as Gallium Nitride (GaN) may enable higher-temperature operation than silicon, potentially reducing the cooling burden. Fifth, modular and rapid deployment designs, as demonstrated by the underwater data center concept, could accelerate the speed at which new capacity is brought online. Sixth, AI-driven cooling management will become ubiquitous, with machine learning systems continuously optimizing cooling setpoints, fan speeds, and water flow across hybrid cooling environments.

Finally, although space-based data centers remain experimental, the convergence of declining launch costs, miniaturized high-performance computing hardware, and growing demand for compute capacity may make orbital data centers viable for specific applications such as batch processing or archival storage within the next decade. The most sustainable data centers of the future will likely blend multiple strategies and intelligently adapt to operating conditions, integrating advanced cooling, renewable energy, waste heat recovery, and AI optimization into a cohesive system designed from the outset for minimal environmental impact.

Abbreviations

AI: Artificial Intelligence, ASHRAE: American Society of Heating, Refrigerating and Air-Conditioning Engineers, CPU: Central Processing Unit, CRAC: Computer Room Air Conditioner, GPU: Graphics Processing Unit, HPC: High-Performance Computing, IC: Immersion Cooling, IT: Information Technology, PUE: Power Usage Effectiveness, WUE: Water Usage Effectiveness.

Acknowledgements

The authors acknowledge the support of their respective institutions in facilitating this research.

Author Contributions

Luca Fiorillo: conceptualization, literature acquisition, synthesis, analysis, writing- draft, revision, Sharon Ridolfo: methodology, data curation, Attilio Caravelli: visualization, Dario Milone: methodology, data curation. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest related to the content of this manuscript.

Data Availability Statement

No primary data were generated or analyzed in this study. All sources used are cited within the manuscript.

Declaration of Artificial Intelligence (AI) Assistance

During the preparation of this manuscript, the authors used AI-assisted tools for literature search assistance and initial text structuring. The authors subsequently reviewed, edited, and verified all content produced with AI assistance, taking full responsibility for the accuracy, originality, and integrity of the final published work.

Ethics Approval

This study is a narrative review of published literature and did not involve human or animal subjects. Therefore, ethics approval was not required.

Funding

This research received no external funding.

References

1. Masanet E, Shehabi A, Lei N, Smith S, Koomey J. Recalibrating Global Data Center Energy-Use Estimates. Science. 2020;367(6481):984-986. https://doi.org/10.1126/science.aba3758
2. Shehabi A, Smith SJ, Masanet E, Koomey J. Data Center Growth in the United States: Decoupling the Demand for Services from Electricity Use. Environ Res Lett. 2018;13(12):124030. https://doi.org/10.1088/1748-9326/aaec9c
3. Zhang Y, Zhao Y, Dai S, *et al.* Cooling Technologies for Data Centres and Telecommunication Base Stations: A Comprehensive Review. J Clean Prod. 2022;334:130280. https://doi.org/10.1016/j.jclepro.2021.130280
4. Huang X, Yan J, Zhou X, Wu Y, Hu S. Cooling Technologies for Internet Data Center in China: Principle, Energy Efficiency, and Applications. Energies. 2023;16(20):7158. https://doi.org/10.3390/en16207158
5. Habibi Khalaj A, Halgamuge SK. A Review on Efficient Thermal Management of Air- and Liquid-Cooled Data Centers: From Chip to the Cooling System. Appl Energy. 2017;205:1165-1188. https://doi.org/10.1016/j.apenergy.2017.08.037
6. Lei N, Masanet E. Statistical Analysis for Predicting Location-Specific Data Center PUE and Its Improvement Potential. Energy. 2020;201:117556. https://doi.org/10.1016/j.energy.2020.117556
7. Ott B, Wenzel PM, Radgen P. Analysis of Cooling Technologies in the Data Center Sector on the Basis of Patent Applications. Energies. 2024;17(15):3615. https://doi.org/10.3390/en17153615
8. Mytton D. Data Centre Water Consumption. npj Clean Water. 2021;4:11. https://doi.org/10.1038/s41545-021-00101-w
9. Nadjahi C, Louahlia H, Lemasson S. A Review of Thermal Management and Innovative Cooling Strategies for Data Centers. Sustain Comput Inform Syst. 2018;19:14-28. https://doi.org/10.1016/j.suscom.2018.05.002
10. Shehabi A, Smith SJ, Sartor DA, *et al*. United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory. 2016. LBNL-1005775. https://doi.org/10.2172/1372902
11. Herrlin MK. Rack Cooling Effectiveness in Data Centers and Telecom Central Offices: The Rack Cooling Index (RCI). ASHRAE Trans. 2005;111(Part 2):725-731. https://www.ashrae.org/technical-resources/ashrae-transactions
12. Barroso LA, Clidaras J, Hölzle U. The Datacenter as a Computer: Designing Warehouse-Scale Machines. 3rd ed. San Rafael: Morgan and Claypool Publishers; 2018. https://doi.org/10.2200/S00874ED3V01Y201809CAC046
13. Ebrahimi K, Jones GF, Fleischer AS. A Review of Data Center Cooling Technology, Operating Conditions and the Corresponding Low-Grade Waste Heat Recovery Opportunities. Renew Sustain Energy Rev. 2014;31:622-638. https://doi.org/10.1016/j.rser.2013.12.007
14. Kong R,Zhang H,Tang M, *et al.* Enhancing data center cooling efficiency and ability: A comprehensive review of direct liquid cooling technologies.Energy. 2024; 308:132846. doi: 10.1016/j.energy.2024.132846
15. Capozzoli A, Primiceri G. Cooling Systems in Data Centers: State of Art and Emerging Technologies. Energy Procedia. 2015;83:484-493. https://doi.org/10.1016/j.egypro.2015.12.168
16. Oró E, Depoorter V, Garcia A, Salom J. Energy Efficiency and Renewable Energy Integration in Data Centres: Strategies and Modelling Review. Renew Sustain Energy Rev. 2015;42:429-445. https://doi.org/10.1016/j.rser.2014.10.035
17. Nall DH. Data Centers, Cooling Towers and Thermal Storage. ASHRAE J. 2018;60(4):64-70. https://www.researchgate.net/publication/326846364_Data_centers_cooling_towers_thermal_storage
18. Starling D. Immersion Cooling Heats Up. Commun ACM. 2022;65(7):13-15. https://doi.org/10.1145/3535336
19. Wang L, Li H, Liu H, Wang X, Zhang Y, Chen H. Activate Radiative Cooling Technology for Data Center Cooling and Energy Efficiency Analysis. J Therm Sci. 2025;34:1129-1147. https://doi.org/10.1007/s11630-025-2146-x
20. European Commission. Energy Efficiency Directive (EU) 2023/1791 of the European Parliament and of the Council. Off J Eur Union. 2023;L231:1-111. https://eur-lexeuropa.eu/legal content/EN/TT/?uri=CELEX:32023L1791
21. U.S. Office of Management and Budget. Federal Data Center Optimzation Initiative (DCOI), Memorandum M-19-19. Wahington, DC: OMB; 2019. https://www.whitehouse.gov/wpcontent/uploads/2019/06/M-19-19-Data-Centers.pdf
22. Singapore Economic Development Board (EDB), Infocomm Media Development Authority (IMDA).

Sustainable Data Centre Pilot Framework. Singapore: Singapore Government; 2022. https://www.edb.gov.sg/en/about-edb/media-releases-publications.html

23. Microsoft Corporation. Project Natick Phase 2. Microsoft Research Technical Report. 2020. https://natick.research.microsoft.com
24. Greengard S. Datacenters Go to Space. Commun ACM. 2025;68(5):17-19. https://cacm.acm.org/news/datacenters-go-to-space
25. Patterson D, Gonzalez J, Hölzle U, *et al.* The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. Computer. 2022;55(7):18-28. https://doi.org/10.1109/MC.2022.3148714

How to Cite: Fiorillo L, Ridolfo S, Caravelli A, Milone D. Cooling Strategies for Energy-efficient Data Centers: Narrative Review and Future Perspective. Int Res J Multidiscip Scope. 2026; 7(2): 1351-1364.
DOI: 10.47857/irjms.2026.v07i02.08460