# AI Water Use: What We Know, What We Don't, and What It Could Mean for Tampa Bay

*An independent research report examining how artificial intelligence uses water, how much it uses, where its impacts can occur, and how future data-center development could affect the Tampa Bay region.*

Thank Your Water — October 2026

## Table of Contents


**1. Executive Summary** ..... 5

**Purpose, Scope and Review Approach** ..... 7

**2. What Does It Mean When We Say AI "Uses Water"?** ..... 8

**2.1 AI Does Not Literally Drink Water** ..... 8

**2.2 Three Parts of the Water Footprint** ..... 9

**2.3 Withdrawal vs. Consumption** ..... 10

**2.4 Water Use vs. Water Waste** ..... 11

**3. Why Data Centers Need Water** ..... 13

**3.1 Servers Produce Heat** ..... 13

**3.2 How Data Centers Are Cooled** ..... 14

**3.3 The Water-Energy Tradeoff** ..... 16

**3.4 Wastewater and Water Quality** ..... 17

**4. How Much Water Does AI Actually Use?** ..... 18

**4.1 Why There Is No Single "Water per Prompt" Number** ..... 19

**4.2 The Famous "Bottle of Water" Claim** ..... 20

**4.3 Modern Text Inference** ..... 21

**4.4 Reasoning and Compute-Intensive AI** ..... 23

**4.5 Image, Audio and Video Generation** ..... 24

**4.6 Training vs. Inference** ..... 25

**4.7 From One Prompt to Millions of Users** ..... 27

**5. How Big Is AI's Total Water Footprint?** ..... 28

**5.1 U.S. Data Centers Overall**....29
**5.2 What Portion Is Actually AI?** ....30
**5.3 Current AI Estimates** ....32
**5.4 Future AI Water Projections** ....34
**5.5 Why Scale Matters More Than the Viral Prompt**....36
**6. AI vs. the Rest of the Internet**....37
**6.1 What Conventional Digital Activity Already Uses** ....38
**6.2 What Can Actually Be Compared** ....39
**6.3 What Cannot Honestly Be Compared**....41
**6.4 Is AI Unusually Resource Intensive?**....43
**6.5 AI Is Becoming Part of "Normal Internet Use"** ....44
**6.6 What We Still Cannot Measure**....46
**7. Where AI and Data Centers Can Actually Cause Harm** ....47
**7.1 Water-Stressed Communities**....48
**7.2 Competition for Potable Water**....50
**7.3 Drought and Peak Demand**....52
**7.4 Groundwater and Surface-Water Pressure**....53
**7.5 Wastewater and Water Quality**....55
**7.6 Electricity-Related Water Use** ....57
**7.7 Semiconductor Manufacturing** ....57
**7.8 Heat** ....59
**7.9 Cumulative Development** ....60
**7.10 Environmental Justice / Community Distribution**....62
**8. Case Studies: Where Data Center Water Became a Real Issue** ....64
**8.1 The Dalles, Oregon: Water Demand and Public Transparency**....65
**8.2 Northern Virginia: When Data Centers Become a Regional Industry** ....67
**8.3 Arizona: Data Centers in a Water-Stressed Climate** ....70
**8.4 Quincy, Washington: Can Data Center Water Be Reused?** ....73
**8.5 What the Case Studies Show** ....77

**9. Are AI Companies Making It Better?** ..... 77
**9.1 Computing Efficiency** ..... 78
**9.2 Cooling Innovation** ..... 79
**9.3 Reclaimed and Non-Potable Water** ..... 81
**9.4 Better Siting** ..... 83
**9.5 Cleaner / Lower-Water Electricity** ..... 85
**9.6 Corporate Water Commitments** ..... 88
**9.7 What "Water Positive" Actually Means** ..... 91
**9.8 Limits of Water Replenishment** ..... 94
**9.9 The Transparency Problem** ..... 97
**10. Will AI's Water Problem Get Better or Worse?** ..... 100
**10.1 AI Is Becoming More Efficient** ..... 101
**10.2 Workloads Are Also Becoming More Demanding** ..... 103
**10.3 The Rebound Effect** ..... 105
**10.4 Growth Scenarios** ..... 107
**10.5 Siting Scenarios** ..... 109
**10.6 Mitigation Scenarios** ..... 111
**10.7 The Most Defensible Forecast** ..... 112
**11. AI, Data Centers and Tampa Bay** ..... 114
**11.1 Tampa Bay's Existing Data-Center Landscape** ..... 114
**11.2 What We Know About Their Water Use** ..... 116
**11.3 What We Don't Know** ..... 117
**11.4 Tampa Bay's Water System** ..... 118
**11.5 Current Water Conditions** ..... 120
**11.6 What Hyperscale Development Could Change** ..... 122
**11.7 Florida's New Data-Center Water Requirements** ..... 124
**11.8 Hillsborough County's Response** ..... 127
**11.9 TBRPC's Regional Planning Work** ..... 128
**11.10 Florida's Forthcoming Independent Study** ..... 131

**12. What Could Tampa Bay Do?** ..... 132
**12.1 Prioritize Reclaimed Water** ..... 133
**12.2 Reduce Routine Potable-Water Cooling** ..... 134
**12.3 Encourage Low-Water Cooling Technologies** ..... 136
**12.4 Require Facility-Level Water Reporting** ..... 137
**12.5 Drought Response Plans** ..... 139
**12.6 Strategic Siting** ..... 140
**12.7 Data Center & Wastewater Infrastructure** ..... 142
**12.8 Developer-Funded Infrastructure** ..... 144
**12.9 Community Benefit Agreements** ..... 146
**12.10 Regional Coordination** ..... 147
**12.11 Transparency and NDAs** ..... 149
**13. What We Still Don't Know** ..... 151
**13.1 How Much Water Individual Data Centers Actually Use** ..... 154
**13.2 How Much of the Data-Center Footprint Actually Belongs to AI** ..... 155
**13.3 How Different AI Services and Digital Activities Really Compare** ..... 157
**13.4 The Full Supply-Chain Footprint** ..... 159
**13.5 What Future AI Infrastructure Will Actually Look Like** ..... 161
**13.6 What Will Actually Be Built in Tampa Bay** ..... 162
**13.7 What Uncertainty Does and Does Not Mean** ..... 164
**14. Beyond Consumption: How Data Centers Can Affect Water and the Environment** ..... 165
**14.1 Water Quantity vs. Water Quality** ..... 165
**14.2 Cooling Water, Blowdown and Wastewater** ..... 166
**14.3 Construction, Groundwater and Stormwater** ..... 167
**14.4 Heat, Electricity and Air** ..... 168
**14.5 Cumulative and Community Effects** ..... 170
**14.6 What Is Actually AI-Specific?** ..... 171
**15. Key Takeaway** ..... 172
**16. Sources & Further Reading** ..... 173

## 1. Executive Summary

Artificial intelligence uses water through the physical infrastructure that makes it possible: data-center cooling, electricity generation and the manufacturing of semiconductors and other equipment. These pathways are not interchangeable. Their environmental significance depends on how much water is withdrawn or consumed, where it comes from, when it is used, what happens to it afterward and the condition of the water systems involved.

This report examines the evidence behind increasingly common claims about AI and water. Its central finding is that there is no single defensible number for how much water "AI" uses. Estimates differ because they measure different models, workloads, facilities, cooling systems, electricity sources, locations and portions of the computing lifecycle. Training a model, answering a short text query, performing extended reasoning and generating images or video are different computational activities. Treating them as equivalent "prompts" can obscure their resource requirements.

Published measurements and models show substantial variation in the resource requirements of individual AI interactions. Google reported that its median Gemini Apps text prompt in May 2025 consumed approximately 0.24 watt-hours of electricity and 0.26 milliliters of water. Independent peer-reviewed research published in Joule modeled a standard frontier-scale text query at a median 0.31 watt-hours, while a long reasoning workload required approximately 3.91 watt-hours under the study's modeled conditions. These findings demonstrate substantial variation between workloads; they do not establish a universal water-per-prompt figure.

Small per-interaction footprints can coexist with substantial industry-wide demand. Lawrence Berkeley National Laboratory's 2024 report estimated that U.S. data centers consumed approximately 66 billion liters of water directly in 2023, equivalent to about 17.4 billion gallons. It separately estimated nearly 800 billion liters of indirect water consumption associated with their electricity use, based on regional electricity-grid mixes. These estimates cover data centers generally, including both AI and non-AI workloads; they do not establish AI's share of that consumption.[15]

Efficiency and total demand must also be distinguished. Improvements in hardware, models, cooling and serving infrastructure can reduce resource requirements for equivalent computing, while growing adoption, reasoning, agents and multimodal generation can increase the amount of computing performed. AI can therefore become more water-efficient while using more water overall. Future demand depends on how these changes interact.

Location and infrastructure choices can substantially change the outcome. Research reviewed in this report shows that siting, cooling technology and electricity supply can produce markedly different water footprints for similar amounts of computing. A facility's annual water total also cannot fully describe its local significance: demand during drought or peak-use periods may matter more than an annual average, particularly where multiple facilities depend on the same constrained supply.

Water quantity is only one dimension of environmental impact. Cooling wastewater creates treatment and discharge requirements; construction can affect stormwater, wetlands and groundwater; and electricity supply and heat rejection can affect air quality, nearby temperatures, noise and utility costs. The cases reviewed include measured local effects, documented service disruptions, residents' reports and disputed allegations. Each must be evaluated according to its evidence and local conditions. Reducing direct water consumption does not necessarily reduce other impacts, and reclaimed-water use still requires adequate treatment and supporting infrastructure.

Most of these effects are associated with data centers generally rather than being unique to AI. AI matters because it can increase the scale, intensity and pace of computing infrastructure development. Public reporting often does not separate AI from other workloads, so measured data-center impacts cannot automatically be attributed entirely to AI. That attribution gap does not remove the need to evaluate the infrastructure's actual demands on communities and shared resources.

For the broader Tampa Bay region, the report identifies an opportunity to plan before hyperscale development becomes established. The Tampa Bay Regional Planning Council's assessment reviewed here identifies no existing large or hyperscale data centers. A future hyperscale campus could nevertheless differ substantially from the region's current facilities in electrical capacity, water requirements and supporting infrastructure. Its effects would depend on its design, operating load, location, water source and seasonal conditions.

Regional evaluation should therefore extend beyond whether enough water is available to supply an individual project. It should address withdrawal and consumption, drought resilience, reclaimed-water availability, wastewater quantity and quality, land and stormwater effects, electricity supply, infrastructure costs and cumulative demand. Reclaimed water can reduce pressure on potable supplies, but its feasibility depends on treatment capacity, distribution, storage, existing commitments and conditions at the proposed site.

Florida's 2026 legislation adds water-permitting requirements for large-scale data centers and directs an independent interdisciplinary study of their wider effects, due July 1, 2027. These measures provide an additional framework for evaluating development, while project-specific disclosure and regional coordination remain essential.

The evidence supports evaluating AI infrastructure according to its actual design and local context. Technological improvement can reduce impacts, but it does not guarantee that total demand will decline. For Tampa Bay, the practical opportunity is to establish clear expectations for siting, resource use, environmental management, transparency and infrastructure responsibility while many of the decisions shaping future development remain open.

## Purpose, Scope and Review Approach

This report examines how AI-related computing uses water, what published estimates can establish, where environmental and community impacts have been reported, and what these findings could mean for Tampa Bay.

It is a narrative evidence review drawing on peer-reviewed research, government and utility publications, corporate disclosures, public records and reputable news reporting available through September 2026. It synthesizes existing evidence rather than presenting original facility measurements or an exhaustive systematic review.

Sources are evaluated according to their methods, reporting boundaries and relevance to the claim being examined. The report distinguishes measured results from modeled estimates, company-reported performance from independent findings, and reported community impacts from established or disputed causes. Comparisons account for differences in geography, time period, workload and the meaning of the water-use metric. Illustrative calculations are identified separately from estimates of actual operations.

The Tampa Bay review examines publicly available information about existing facilities and regional planning conditions. Gaps in facility-level disclosure limit conclusions about current water consumption, seasonal demand and the share attributable to AI. Findings from other communities inform the issues that local proposals may require planners to investigate; they do not establish that the same outcomes will occur in Tampa Bay.

## 2. What Does It Mean When We Say AI "Uses Water"?

Claims that artificial intelligence "uses water" can be confusing because AI software does not interact with water directly. The water footprint exists because AI depends on physical computing infrastructure, electricity generation and the manufacturing supply chain that produces the hardware on which AI runs.

Water can enter that system at several points. Data centers may use water directly for cooling and other facility operations. Producing the electricity consumed by those facilities can require water elsewhere. Manufacturing semiconductors, servers and other equipment also carries an embedded water footprint. These are different forms of water use; combining them without identifying the boundary being measured can produce misleading comparisons.

The terminology matters as well. Water withdrawal and water consumption are not the same measurement, and neither automatically tells us whether water is being wasted or whether its use is environmentally harmful. Water that is withdrawn and returned to its source creates a different impact from water lost through evaporation, while the significance of either can depend on local scarcity, timing, water quality and competing uses.

Before asking how much water AI uses, then, it is necessary to establish what is being counted, where in the system the water is used and what happens to it afterward. Those distinctions provide the foundation for interpreting every water estimate used later in this report.

### 2.1 AI Does Not Literally Drink Water

Artificial intelligence is software, but the computations that make it work take place on physical machines. When someone asks a chatbot a question, generates an image or uses another AI-powered service, that request is processed by servers housed in data centers. Those servers consume electricity and produce heat as they perform calculations, store information and move data through networks. The heat must then be removed to keep the equipment operating safely and reliably.[1]

This physical infrastructure is where AI's connection to water begins. Depending on the facility, water may be consumed directly through cooling and humidity-control systems. Additional water can be consumed indirectly in generating the electricity that powers the servers and the rest of the data center. The amount associated with a particular computing workload therefore depends not only on the AI model itself, but also on factors such as the

servers performing the work, how efficiently they are being used, the cooling technology, local climate and the electricity sources supplying the facility.[1]

Phrases such as “AI drinks water” or “a prompt uses water” are convenient shorthand, but they can obscure what is physically happening. A prompt does not cause water to pass through an AI model. Rather, digital activity creates demand for computation, electricity and cooling within a larger physical infrastructure, and parts of that system can have a water footprint.

Understanding that infrastructure is the starting point for answering the more difficult question explored later in this report: how much water should actually be attributed to AI use?

## 2.2 Three Parts of the Water Footprint

The water footprint associated with AI and data centers extends beyond the water used inside the facility itself. A useful way to understand the full system is to separate it into three categories: direct operational water use, indirect electricity-related water use, and embedded or supply-chain water use. These are sometimes described as Scope 1, Scope 2 and Scope 3 water use, applying terminology commonly used in greenhouse-gas accounting.[2]

### *Direct Water Use*

Direct water use occurs at the data center itself. Most direct water consumption is associated with cooling, particularly in systems that use evaporation to remove heat generated by servers and other equipment. Water may also be used for humidity control and other facility operations. The amount consumed can vary substantially depending on the cooling technology, climate, operating conditions and whether water is recirculated or recovered.[1, 2]

Direct water use is the most visible part of a data center's water footprint, but it is not necessarily the largest.

### *Indirect Water Use*

Data centers also have an indirect water footprint through the electricity they consume. Many thermal power plants, including nuclear, coal and some natural-gas facilities, use water as part of their cooling processes. When electricity from those sources powers a

data center, a portion of that off-site water consumption can be attributed to the electricity required for computing.[2]

The size of this indirect footprint depends heavily on the local electricity mix. A data center supplied primarily by electricity sources with low operational water requirements can have a substantially different indirect water footprint from an otherwise similar facility supplied by water-intensive thermal generation.[2]

Reducing water consumption inside a data center therefore does not necessarily eliminate the water associated with operating it.

#### *Embedded and Supply-Chain Water*

A third component exists before the servers are even turned on. Building data centers and manufacturing the equipment inside them require water throughout their supply chains. This can include water associated with producing semiconductors, servers and other electronic hardware, as well as construction materials such as steel and concrete.[2]

Semiconductor manufacturing is particularly water-intensive because advanced chip fabrication requires large quantities of extremely clean ultrapure water for processes such as rinsing wafers, cooling manufacturing equipment and diluting chemicals. Producing ultrapure water itself also requires additional source water.[2]

This embedded water footprint is different from operational water use because much of it occurs at manufacturing and material-production facilities that may be located far from the data center itself. It nevertheless forms part of the broader physical footprint of building and continually replacing the infrastructure on which AI depends.[2]

These three pathways illustrate why asking how much water “AI uses” can produce very different answers. A calculation that includes only cooling water measures something fundamentally different from one that also includes electricity generation and hardware manufacturing. Before comparing water-use estimates, it is therefore necessary to determine which parts of the system each estimate actually includes.

## 2.3 Withdrawal vs. Consumption

One of the most important distinctions in discussions about AI and water is the difference between water withdrawal and water consumption. The terms are sometimes used interchangeably in public discussion, but they describe different things.

Water withdrawal is the amount of water removed from a groundwater or surface-water source for use. Not all of that water is necessarily lost. Depending on how it is used, some may later be discharged or returned to the environment. The U.S. Geological Survey defines return flow as water that reaches a surface-water or groundwater source after use and becomes available for further use.[3]

Water consumption, also called consumptive use, refers to the portion of withdrawn water that is no longer immediately available to the local water system. This includes water that evaporates, is incorporated into a product or is otherwise removed from the immediate water environment. In a data center using evaporative cooling, for example, some of the water withdrawn for cooling is ultimately released into the atmosphere as water vapor rather than returned as liquid water.[3]

The distinction can produce dramatically different numbers. A facility or power plant may withdraw a large quantity of water but return much of it after use, resulting in substantially lower consumption. Another system may withdraw less water overall but consume a greater percentage of what it takes in. For that reason, withdrawal alone does not describe how much water is actually removed from immediate local availability.

Discharge and return flow complete the basic water balance. Water that is not consumed may leave a facility as wastewater, be treated, or eventually return to surface water or groundwater. Returning water does not necessarily mean that its use had no environmental effect; the timing, location, temperature and quality of the returned water can also matter. But it should not automatically be counted as though the entire withdrawn volume were consumed.[3]

This distinction is particularly important when evaluating large numbers reported for AI, data centers and electricity generation. A study estimating billions of gallons of water withdrawal is not necessarily saying that billions of gallons were consumed. Likewise, two studies using different definitions cannot be compared directly simply because both describe their results as “water use.”

Throughout this report, withdrawal and consumption are therefore treated as separate measurements. When a source reports only one, that measurement should not be presented as though it represents the other.

## 2.4 Water Use vs. Water Waste

The fact that a data center uses water does not by itself establish that the water is being wasted. Water is used throughout homes, agriculture, energy production, manufacturing

and other economic activity. Evaluating whether a particular use is environmentally responsible requires more information than the volume used alone.

One important consideration is the source of the water. A facility drawing potable freshwater from a constrained local supply presents a different set of concerns from one using reclaimed wastewater. Similarly, the same volume of freshwater can have very different consequences depending on whether it is withdrawn from a water-abundant area or from a watershed already experiencing scarcity. Water impacts are therefore highly dependent on local conditions.[4]

Timing also matters. Water demand that can be accommodated during wetter periods may create greater pressure during droughts, periods of low streamflow or seasons of high community demand. Annual water totals can obscure these shorter periods of stress, particularly when a facility's demand remains relatively constant while local water availability changes throughout the year.[4]

How efficiently water is used is another consideration. Data centers can differ in cooling technology, climate, operating conditions and opportunities to recirculate water.[1] A facility may also have alternatives to potable freshwater, including reclaimed or other non-potable sources, depending on local infrastructure and water-quality requirements. Whether those alternatives are technically and economically feasible is therefore relevant when evaluating whether freshwater consumption could reasonably have been avoided.

Evaporation, reuse and discharge also affect the outcome. Water that is repeatedly circulated through a cooling system before being discharged is different from water consumed through evaporation after a single use. Water that leaves a facility as discharge may remain available within the broader water system, although its temperature, quality, treatment requirements and point of return can still create environmental consequences.[3]

For these reasons, this report does not use the terms water use and water waste interchangeably. Instead, potential impacts are evaluated in context: how much water is withdrawn and consumed, where it comes from, whether the local watershed is under stress, when the demand occurs, how efficiently the water is used, whether practical alternatives exist, how much is reused or returned, and what happens to the water after use.

This distinction does not assume that data-center water consumption is either harmless or wasteful. It provides a framework for determining when water demand represents a manageable use of a resource and when it may create avoidable or disproportionate pressure on a local water system.

## 3. Why Data Centers Need Water

Data centers convert enormous amounts of electrical energy into computation, and nearly all of that energy ultimately becomes heat. Servers, processors, memory, networking equipment and power systems therefore require continuous thermal management to remain within safe operating temperatures. Water enters the picture primarily because it can be an effective medium for moving and rejecting that heat.

There is no single way that data centers are cooled. Some cooling systems consume significant quantities of water through evaporation, while others can operate with little or no routine on-site water consumption. The cooling system chosen depends on factors including hardware density, climate, electricity use, water availability and facility design.

This creates an important environmental tradeoff. Technologies that reduce direct water consumption can sometimes require more electricity, while highly energy-efficient evaporative systems may consume more water. Evaluating a data center solely by its water use or its energy use can therefore conceal impacts shifted elsewhere in the system.

Cooling can also create water-quality considerations. Water circulating through cooling systems can become concentrated with salts and other dissolved materials, while treatment chemicals may be necessary to control corrosion, scaling and biological growth. How that water is treated, reused or discharged is therefore part of the facility's water footprint alongside the amount initially withdrawn or consumed.

Understanding why the heat exists, how different cooling systems remove it, and the tradeoffs those systems create is necessary before evaluating whether a particular data center's water use is excessive, efficient or avoidable.

### 3.1 Servers Produce Heat

Data centers convert electricity into computing work, but ultimately nearly all of the electricity consumed by servers and other information-technology equipment becomes heat. Processors such as CPUs and GPUs contain billions of transistors that rapidly switch electrical signals while performing calculations. As computing power increases, large amounts of heat can be concentrated within relatively small pieces of equipment.[2]

That heat must be continuously removed. Electronic components are designed to operate within specific temperature ranges, and excessive heat can reduce performance, shorten equipment life or cause systems to shut down. Data centers therefore require cooling systems that transfer heat away from servers and ultimately release it outside the facility.[2]

Water enters the process because it is an effective medium for transporting and rejecting heat. In some facilities, water circulates through cooling equipment without being consumed directly. In others, particularly systems that rely on evaporative cooling, some water is intentionally evaporated because the phase change from liquid water to water vapor removes substantial amounts of heat.[2]

Not every data center therefore requires the same amount of water, and some cooling designs can operate with little or no routine water consumption for heat rejection. The need common to every facility is not water itself, but heat removal. Water use depends on the engineering system chosen to accomplish that task.[1, 2]

This becomes increasingly relevant for AI infrastructure. AI workloads can require high-density computing systems containing power-intensive accelerators such as GPUs, concentrating more heat into individual server racks than many conventional computing workloads. As rack power density increases, cooling architecture becomes an increasingly important part of data-center design.[2]

For that reason, the question is not simply “Why does AI need water?” A more precise question is: How does a data center remove the heat generated by computing, and what role does water play in the cooling system it chooses?

## 3.2 How Data Centers Are Cooled

There is no single cooling system used by all data centers. Facilities can use air, water, refrigerants or combinations of these technologies to move heat away from computing equipment and release it to the surrounding environment. The cooling method matters because different designs can have very different water and energy requirements.[1, 2]

Cooling towers and evaporative cooling remove heat by allowing a portion of water to evaporate into the atmosphere. In a common configuration, heat from computing equipment is transferred to a chilled-water loop, then through a chiller to a separate condenser-water loop. That heated water travels to a cooling tower, where evaporation releases the heat outdoors. Because some water is lost through evaporation and additional water must be discharged as blowdown to control the buildup of dissolved minerals, fresh makeup water must continually be added to the system.[5]

Chilled-water systems circulate cooled water through a facility to collect and transport heat. Importantly, the presence of a chilled-water loop does not by itself reveal how much water a data center consumes. The chilled water can circulate repeatedly in a closed loop. What determines much of the facility's direct water consumption is how the collected heat

is ultimately rejected. A chilled-water system connected to an evaporative cooling tower can consume substantial water, while one rejecting heat through dry equipment can operate with much lower direct water consumption.[5]

Air cooling and dry cooling transfer heat to outdoor air without relying on evaporation as the primary means of heat rejection. Data centers can also use cool outside air directly when weather conditions permit, a practice known as air-side economizing. These approaches can substantially reduce or eliminate routine cooling-water consumption, although their effectiveness and energy requirements depend on climate, temperature, humidity and facility design.[2, 5]

Closed-loop cooling describes systems in which a cooling fluid is recirculated rather than continuously discarded. However, “closed loop” should not automatically be interpreted as “water-free.” A closed internal loop can still transfer its heat to an evaporative cooling tower elsewhere in the system. Understanding water consumption therefore requires following the entire path of heat from the servers to its final release outside the facility.[5]

Direct-to-chip liquid cooling moves liquid directly to cold plates attached to heat-producing components such as CPUs and GPUs. Instead of first transferring most of the heat from the chips into room air, the liquid absorbs heat close to where it is generated and carries it away through a recirculating loop. Liquid can transport heat more effectively than air, making this approach particularly useful for increasingly dense computing equipment. The system still requires a method for ultimately rejecting that heat, which can be air-cooled, evaporative or a combination of technologies.[5]

Immersion cooling takes liquid cooling a step further by placing computing equipment in a nonconductive fluid that directly absorbs heat from the components. The warmed fluid then transfers that heat through another part of the cooling system before being recirculated. Immersion can support high-density computing while reducing dependence on conventional server-room air cooling, although its overall water footprint still depends on how the captured heat is ultimately rejected.[2]

Finally, hybrid systems combine cooling methods and can switch between them as conditions change. A facility might rely on dry or air-based heat rejection during cooler weather and use evaporative cooling when outdoor conditions or computing loads make additional cooling capacity necessary. This can reduce annual water consumption while preserving cooling performance during periods of high demand.[5]

These differences matter when comparing data centers. Descriptions such as “liquid cooled,” “closed loop” or “chilled water” identify only part of the cooling system. To

understand its water footprint, it is necessary to determine both how heat is removed from the computing equipment and how that heat is ultimately released from the facility.

## 3.3 The Water-Energy Tradeoff

Reducing a data center's direct water consumption does not always reduce its overall environmental footprint. Cooling systems operate within a water-energy relationship in which conserving one resource can sometimes increase demand for another. This makes it important to evaluate water and electricity together rather than treating either as an isolated measure of sustainability.[6]

Evaporative cooling demonstrates this tradeoff clearly. Evaporating water is an efficient way to remove heat and can reduce the amount of electricity required for mechanical cooling. Dry or air-cooled systems can greatly reduce direct water consumption because they reject heat without routinely evaporating water, but they may require more electricity to provide the same amount of cooling, particularly under hot outdoor conditions.[6]

A study of two data centers in the Phoenix area illustrates the difference. Researchers compared an air-cooled facility with another using water-cooled chillers and evaporative cooling. The air-cooled facility had a substantially smaller water footprint, but its average Power Usage Effectiveness, a measure of the additional energy required to operate the facility relative to its computing equipment, was nearly 13 percent higher. The researchers concluded that cooling-system selection involves a measurable tradeoff between water and energy use and should account for local climate and resource availability.[6]

The relationship becomes more complicated when indirect water use is included. Additional electricity required by a low-water cooling system can itself have a water footprint if the electricity is generated by power plants that consume water. In that situation, reducing water use at the data center can shift some of the water demand from the facility to the electrical grid rather than eliminating it entirely. The magnitude of that shift depends on the sources of electricity supplying the facility.[6]

Carbon emissions introduce another dimension. A cooling system that minimizes direct water consumption may require more electricity and therefore produce greater emissions when supplied by a carbon-intensive grid. Conversely, a highly energy-efficient evaporative system may have relatively low electricity demand while consuming more freshwater locally. A low-carbon electricity source may also have either high or low operational water requirements depending on the generation technology and cooling system involved.[2, 6]

Neither low water use nor low carbon emissions should therefore automatically be treated as evidence of low overall environmental impact. The preferable balance can differ by location. In a water-stressed region, reducing freshwater consumption may deserve greater priority even if doing so carries some energy penalty. In another location, abundant reclaimed water and a carbon-intensive electrical grid could change that calculation.

The goal is therefore not simply to minimize one metric. Evaluating data-center cooling requires considering direct water consumption, indirect water use, electricity demand, carbon emissions and local resource conditions together. A solution that improves one measurement by shifting its environmental burden somewhere else may represent a tradeoff rather than a true reduction in impact.

## 3.4 Wastewater and Water Quality

Water use is only one part of the environmental impact of a cooling system. In facilities that use cooling towers, some water leaves the system through evaporation, while another portion is intentionally discharged as cooling tower blowdown. Blowdown is necessary because evaporation removes relatively pure water while leaving dissolved minerals behind, causing salts and other substances to become increasingly concentrated in the recirculating water.[5, 7]

Cooling-water chemistry must also be managed to protect equipment. Depending on the system and source water, treatment can include corrosion inhibitors, scale-control chemicals and biocides used to limit microbial growth. As water repeatedly circulates and evaporates, dissolved salts and treatment chemicals can become concentrated along with naturally occurring constituents in the source water. Blowdown removes a portion of this concentrated water before scaling, corrosion or biological growth interferes with cooling-system performance.[7]

The resulting wastewater therefore does not necessarily have the same composition as the water originally supplied to the facility. Cooling tower blowdown can contain elevated concentrations of dissolved salts and minerals as well as residual water-treatment chemicals. Its exact composition varies with the source water, cooling system, treatment program and number of times the water has been recirculated.[7]

How that water is discharged or treated matters. Depending on its composition and local requirements, blowdown may be sent to a wastewater treatment system, treated for reuse or discharged under applicable water-quality standards. Research has also demonstrated that cooling tower blowdown can be treated and reused as cooling-system makeup water, reducing freshwater demand and the amount of wastewater requiring disposal.[7]

Newer liquid-cooling technologies introduce a different set of water-quality questions. Direct-to-chip and immersion systems can reduce dependence on conventional evaporative cooling, but some designs use specialized dielectric or other cooling fluids. Research into these fluids evaluates not only thermal performance but also properties such as toxicity, biodegradability, chemical stability and broader environmental impacts. The environmental profile can differ substantially depending on the coolant selected.[8]

Questions have also emerged about fluorinated cooling fluids and PFAS. However, the evidence does not support treating all liquid or immersion cooling fluids as PFAS, nor does the presence of a fluorinated coolant by itself establish that a data center is contaminating local water. Recent research is beginning to evaluate individual fluids for persistence, degradation products, toxicity and environmental behavior. These questions are better assessed by the specific chemistry used and its potential pathways into the environment rather than by assuming that all advanced cooling systems present the same chemical risk.[8]

For this report, wastewater and coolant impacts are therefore evaluated separately from water quantity. A cooling technology that reduces freshwater consumption may still require appropriate chemical management, wastewater treatment and monitoring. Conversely, the existence of cooling-water discharge does not by itself demonstrate pollution. The relevant questions are what the discharge contains, how much is released, how it is treated, where it goes and whether it meets applicable environmental standards.

## 4. How Much Water Does AI Actually Use?

There is no scientifically defensible single number for how much water “AI” uses. The footprint of an AI interaction can vary dramatically depending on the model, hardware, workload, output length, server utilization, cooling system, location, weather, electricity source and the boundaries included in the calculation. A short text response and a long reasoning task are not equivalent workloads, and neither can automatically be assigned the footprint of image or video generation.

This complexity has often been lost in public discussion. Estimates developed for specific models under specific conditions have sometimes been transformed into universal claims about how much water every AI prompt consumes. The widely repeated comparison between ChatGPT and a bottle of water is one example: the underlying research modeled a particular generation of technology and estimated approximately 500 milliliters across multiple responses under specified conditions. It did not estimate one bottle consumed by every question.

More recent research shows just how large the variation can be. Even within text inference, workload characteristics and infrastructure choices can change energy and water requirements substantially, while reasoning and other compute-intensive applications can require far more computation than ordinary text generation. Training and inference also need to be distinguished: training represents large but episodic computational events, while inference occurs continuously as users interact with deployed models.

The appropriate question is therefore not simply "How much water does one AI prompt use?" A more useful question is: "How much water is associated with this particular AI workload, operating on this infrastructure, in this location, using this electricity and cooling system, under this accounting boundary?"

At the individual level, many AI interactions can have very small footprints. At infrastructure scale, however, small per-interaction requirements can become significant when multiplied across millions or billions of requests. This section therefore moves from the uncertainty surrounding a single prompt to the much more consequential question of what happens when AI is deployed at enormous scale.

## 4.1 Why There Is No Single "Water per Prompt" Number

A 2025 peer-reviewed analysis of data-center workloads found that water consumption per unit of computing workload can vary by more than 10,000-fold depending on the combination of infrastructure and operating conditions.[1]

Much of that variation occurs before differences between individual AI prompts are even considered. The study found more than a 1,000-fold range in water consumed per kilowatt-hour of server electricity and approximately a tenfold range in server workload efficiency. Among the most important factors were server efficiency, the water intensity of the electrical grid, server utilization, cooling technology, overall data-center efficiency and climate.[1]

The workload itself adds another layer of variability. Different AI models require different amounts of computation, and the amount required by the same model can change with the task being performed. Prompt and output length, the type of workload and the hardware processing it can all affect the amount of electricity required to complete a request. Water attributed to that computation therefore cannot be separated from the efficiency of the servers performing it.[1]

Location matters as well. Identical computing workloads performed in two data centers can have different water footprints because the facilities may use different cooling

systems, operate in different climates or receive electricity from grids with very different water requirements. Even the same facility can experience changing conditions over time as weather, cooling requirements and the mix of electricity supplying the grid change.[1]

Server utilization further complicates attempts to assign water use to an individual request. Servers consume electricity even when they are not operating at full capacity, meaning the amount of infrastructure energy allocated to each unit of useful computation depends partly on how effectively the available equipment is being used. Older or less efficient servers and larger shares of inactive equipment can increase water use per workload even when the digital task itself remains unchanged.[1]

For these reasons, a statement such as “one AI prompt uses X amount of water” is incomplete unless the underlying assumptions are known. At minimum, the estimate depends on what computation was performed, what hardware performed it, how efficiently the infrastructure was being used, how the facility was cooled, where and when the computation occurred, and how the electricity supplying it was generated.

Researchers can model or measure AI water consumption under defined conditions. Those estimates are useful when their assumptions and boundaries are clearly stated, but they cannot establish a universal amount of water consumed by every AI prompt.

These limitations are especially relevant to one of the most widely repeated claims about AI water use: that a conversation with an AI chatbot consumes a bottle of water.

## 4.2 The Famous “Bottle of Water” Claim

One of the most widely repeated claims about AI's environmental footprint is that asking an AI chatbot a question consumes a bottle of water. The claim has a legitimate research origin, but the simplified version often repeated online does not accurately describe what the researchers calculated.

The estimate comes from research by Pengfei Li, Jianyi Yang, Mohammad A. Islam and Shaolei Ren, first released as a preprint in 2023 and later published in Communications of the ACM in 2025 as Making AI Less “Thirsty: Uncovering and Addressing the Secret Water Footprint of AI Models.” The researchers developed a methodology for estimating both direct data-center cooling water and indirect water consumption associated with electricity generation.[9]

As a case study, the researchers modeled the operational water footprint of GPT-3, the 175-billion-parameter language model introduced by OpenAI in 2020. For a medium-length conversational request, defined in the study as approximately 800 words or fewer of input

and 150–300 words of output, they assumed server energy consumption of approximately 0.004 kilowatt-hours per request. They then combined that estimate with data-center efficiency, cooling-water and electricity-related water factors for different locations.[9]

The resulting estimate was not one 500-milliliter bottle for every question. The researchers estimated that GPT-3 would consume approximately 500 milliliters of water across roughly 10 to 50 medium-length responses, depending on where and when the model was deployed.[9] That corresponds to roughly 10 to 50 milliliters per response under the modeled conditions.

The estimate was also a model calculation, not a direct measurement from a water meter attached to ChatGPT. It combined assumptions about inference energy with published information about Microsoft's data centers and regional electricity-generation water intensity. The researchers themselves emphasize that water efficiency varies geographically and over time.[9]

There is another important limitation: the calculation concerns GPT-3 under the infrastructure assumptions used in the study. It should not simply be assigned to current versions of ChatGPT or other AI services. AI models, accelerator hardware, inference software and data-center cooling systems have continued to change, while newer applications can also perform substantially more computationally demanding tasks. The authors note that subsequent models may consume more energy and water for some requests, particularly more compute-intensive reasoning workloads, while continuing improvements in computational and water efficiency could reduce consumption per request.[9]

The original research therefore supports a more nuanced conclusion than the viral claim. It demonstrated that operating a large language model can have a measurable water footprint and provided a method for estimating it. It did not establish that every AI prompt, every chatbot or every current ChatGPT interaction consumes one bottle of water.

The “bottle of water” estimate is best understood as an illustration of one modeled GPT-3 deployment scenario, not a universal conversion factor between prompts and water.

## 4.3 Modern Text Inference

More recent evidence suggests that ordinary text interactions with modern AI systems can require substantially less energy, and therefore potentially less operational water per request, than many early public estimates assumed. However, the available

measurements remain limited, and results from one provider or model should not be treated as universal.

One of the strongest real-world measurements comes from Google, which published a 2025 study examining the environmental footprint of Gemini Apps in its production infrastructure. The researchers included not only the AI accelerators actively processing requests, but also host CPUs and memory, provisioned idle capacity and data-center energy overhead. For the median Gemini Apps text prompt measured in May 2025, they reported 0.24 watt-hours of electricity and 0.26 milliliters of water consumption, approximately five drops of water.[10]

That water figure is dramatically lower than the older GPT-3 estimate discussed in the previous section, but the two numbers should not be interpreted as a simple measurement of how much AI improved between GPT-3 and Gemini. They were produced using different models, infrastructure, methodologies, locations and operating assumptions. The comparison demonstrates how much estimates can differ, but it does not isolate the cause of that difference.[9, 10]

Independent peer-reviewed research published in Joule in 2026 provides another useful benchmark for modern text inference. Researchers modeled frontier-scale language models operating under realistic large-scale serving conditions and estimated a median energy requirement of 0.31 watt-hours per standard query, with an interquartile range of 0.16 to 0.60 watt-hours. The researchers found that several prominent earlier estimates overstated production-scale inference energy by approximately four to twenty times.[11]

The similarity between Google's measured 0.24 watt-hours and the independent study's modeled median of 0.31 watt-hours provides useful evidence that ordinary, optimized text inference can operate in the sub-watt-hour range. It does not establish 0.24 or 0.31 watt-hours as a universal value for AI text prompts. Different models, hardware, serving systems, prompt lengths and output lengths can still produce substantially different results.[1, 11]

There is also an important difference in the strength of the available water and energy evidence. Google's study directly provides both an energy and an operational water estimate for Gemini. The Joule study provides a broader independent estimate of energy use across frontier-scale inference, but it does not establish a universal amount of water associated with that electricity. Converting its energy figures into water would require additional assumptions about cooling, location and electricity generation, recreating many of the variables discussed earlier in this report.[1, 11]

Current evidence therefore supports a narrower conclusion than either extreme of the public debate. Ordinary modern text inference can be relatively small on a per-request

basis, including well below some early estimates, but there is still no provider-independent universal water-per-prompt figure. Per-request efficiency also does not answer the separate question of total impact when AI services process enormous numbers of requests.

And ordinary text is only one type of AI workload. When a model is asked to reason for much longer, process large contexts, use tools or perform multiple computational steps before answering, the energy requirement can increase substantially.[11] That distinction is the subject of the next section.

## 4.4 Reasoning and Compute-Intensive AI

Not all text interactions require the same amount of computation. Some AI systems can generate a response after a relatively short inference process, while reasoning models may perform substantially more computation before producing their final answer. This additional computation, often described as test-time or inference-time scaling, can significantly increase the energy required for a single request.[11]

The 2026 Joule study provides a useful demonstration of the difference. Under the researchers' modeled production-scale conditions, a standard frontier-scale text query required a median of approximately 0.31 watt-hours of electricity. When the researchers modeled long reasoning workloads using test-time scaling, median energy use increased to approximately 3.91 watt-hours per query, or about 13 times the standard-query estimate.[11]

The difference results from the amount of computation performed during inference. Reasoning systems can generate large numbers of intermediate tokens or perform extended computational steps before arriving at an answer. Those intermediate calculations may not all appear in the final response, but they still require processing by the underlying hardware. As test-time computation increases, so does the electricity required to complete the request.[11]

This is another reason that prompt counts alone are a poor measure of AI resource use. Two requests submitted through similar-looking chat interfaces can represent very different computational workloads. A short factual question answered directly and a complex problem requiring extended reasoning may each appear to the user as one "prompt," while requiring substantially different amounts of computation.

Greater electricity demand can also translate into a larger water footprint, but the relationship is not fixed. More computation produces more heat and requires more

electricity, potentially increasing both cooling demand and electricity-related water consumption. However, the amount of additional water depends on the same infrastructure variables discussed earlier in this report, including cooling technology, climate, data-center efficiency and electricity source.[1, 11]

The approximately 13-fold difference identified by the Joule study should therefore not be interpreted as meaning that every reasoning prompt uses 13 times as much water as every ordinary prompt. The study measured a difference in modeled energy demand under specific workload assumptions. Converting that difference into water consumption requires additional information about where and how the computation is performed.

The broader finding is more important than the exact multiplier: the type and amount of computation performed during inference can matter considerably more than simply counting the number of prompts. As AI systems increasingly perform extended reasoning and other compute-intensive tasks, improvements in efficiency will need to be considered alongside changes in what users are asking those systems to do.

## 4.5 Image, Audio and Video Generation

Text is only one form of generative AI. Modern systems can also create images, speech, music, sound effects and video, and these workloads can require substantially different amounts of computation. Available research indicates that the type of content being generated can matter greatly when estimating the energy, and therefore potential water, associated with an AI request.[12]

A 2024 study presented at the ACM Conference on Fairness, Accountability, and Transparency systematically measured the inference energy of more than 80 models across a range of machine-learning tasks. Among the task-specific models tested, text generation averaged approximately 0.047 kilowatt-hours per 1,000 inferences, while image generation averaged approximately 2.9 kilowatt-hours per 1,000 inferences. In that experimental setting, image generation therefore required more than 60 times as much energy per inference as text generation on average.[12]

There was also substantial variation within image generation itself. The researchers found that different image-generation models could require dramatically different amounts of energy depending on the model and characteristics of the generated image. This reinforces the same limitation encountered with text: an average measured across a particular set of models should not be converted into a universal energy or water cost for every generated image.[12]

Evidence for current AI video generation is less mature. Video generation requires producing sequences of visual information across time and can be considerably more computationally intensive than generating a single image. The International Energy Agency identifies video generation, along with reasoning and agentic applications, as examples of emerging AI workloads that can consume hundreds or thousands of times more energy per query than simple text generation. However, differences among models, video length, resolution, frame rate and generation methods make a single “energy per AI video” figure particularly unreliable.[13]

A 2025 benchmark by Julien Delavande and colleagues tested seven open text-to-video models on an NVIDIA H100 GPU. Reported GPU, CPU and RAM energy ranged from approximately 0.14 to 109 watt-hours per generated clip, a roughly 780-fold difference. However, the models used different recommended settings, including clip lengths, resolutions and sampling steps, so the results do not compare the energy required to produce equivalent videos. The benchmark illustrates variation among specific model configurations; it does not measure water consumption or establish the energy requirements of commercial services such as Sora or Veo.[55]

Audio generation is less well characterized. Researchers have begun measuring the inference energy of text-to-audio diffusion models and have found that model architecture and generation parameters can significantly affect energy consumption.[14] However, the available evidence is not yet strong or standardized enough to assign a representative energy or water figure to an ordinary AI-generated song, voice clip or sound effect.

Most importantly for this report, energy measurements should not be automatically converted into water measurements. A generated image or video that requires more computation will generally require more electricity and produce more heat, but its actual water footprint still depends on the data center, cooling system, location, electricity source and time of operation.[1] The evidence supports saying that some multimedia AI workloads can be considerably more energy-intensive than ordinary text inference; it does not support claiming that every AI image, song or video consumes a particular quantity of water.

This distinction will become important later when evaluating AI's total footprint. The future water demand of AI depends not only on how many people use it, but also on what kinds of computation increasingly make up that use.

## 4.6 Training vs. Inference

AI's operational footprint can be divided into two major stages: training, when a model learns from large datasets, and inference, when a trained model is used to generate

responses, images or other outputs. Both require computation, electricity and cooling, but they occur in very different ways.[9, 13]

Training a large AI model can be extremely computationally intensive. During training, large numbers of processors may operate continuously for days, weeks or longer while model parameters are repeatedly adjusted. This creates a concentrated period of electricity demand and associated cooling and water use. The GPT-3 case study discussed earlier, for example, estimated approximately 700,000 liters of direct on-site water consumption for training the model in Microsoft's U.S. data centers, before including additional water associated with electricity generation.[9]

That figure should not be treated as the training footprint of every large language model. Training requirements vary with model size, hardware, data, training methods, location and cooling infrastructure. Newer models may require substantially more total computation, while improvements in chips and training techniques can reduce the energy required for a given amount of computation. Public disclosure of the resources used to train major proprietary models also remains limited.[9, 13]

Inference works differently. Once a model has been trained, inference occurs each time people or software systems use it. A single ordinary text request can require only a small fraction of the energy consumed during an entire training run, but inference is repeated continuously across potentially enormous numbers of interactions. Search features, chatbots, image generators, coding assistants, recommendation systems and other AI applications can serve millions or billions of requests over time.[11, 13]

This distinction changes how the environmental footprint should be understood. Training creates a relatively concentrated computational event, while inference creates an ongoing operational demand that grows with the number and complexity of requests. As AI services become more widely used, cumulative inference can become a major, and potentially dominant, portion of a model's lifetime energy demand.[13]

The boundary between the two is also less simple than “train once, use forever.” Models can be retrained, fine-tuned, updated or replaced, while deployed systems may run multiple models and supporting processes to complete a single user request. At the same time, techniques such as caching, batching and improved hardware utilization can reduce the resources required to serve repeated inference workloads.[11]

For water accounting, training and inference should therefore not be combined without identifying the time period and activity being measured. A training-water estimate describes the resources associated with creating or updating a model. A per-query inference estimate describes part of the cost of operating it. A complete lifecycle

assessment may include both, along with the embedded water associated with manufacturing the hardware on which they run.

This distinction also explains why focusing exclusively on the dramatic resource requirements of training can give an incomplete picture of AI's long-term footprint. Training may be large and episodic; inference is smaller per interaction but continuous and scalable. Determining which contributes more ultimately depends on how the model is built, how long it operates and how extensively it is used.

## 4.7 From One Prompt to Millions of Users

A small per-request footprint can become significant when the same activity is repeated at large scale. This is one reason that the environmental impact of AI cannot be understood only by looking at an individual prompt. The relevant question is also how many requests are being processed, how computationally intensive those requests are and how efficiently the infrastructure serves them.

Google's 2025 Gemini measurement provides a useful illustration. The company estimated that the median Gemini Apps text prompt required approximately 0.24 watt-hours of electricity and consumed about 0.26 milliliters of water under the conditions measured.[10] Those quantities are very small for a single interaction.

If every prompt had the reported median footprint:

- 1 million prompts × 0.26 mL would equal about 260 liters of water.
- 100 million prompts would equal about 26,000 liters.
- 1 billion prompts would equal about 260,000 liters, or roughly 69,000 gallons.

Using the same 0.24 watt-hour energy estimate, one billion such prompts would represent approximately 240 megawatt-hours of electricity.[10]

These illustrative calculations demonstrate the effect of scale; they are not estimates of Gemini's actual daily or annual consumption. Actual totals depend on the number of requests and their mean resource use, which cannot be established from the reported median alone.

Actual AI services process mixtures of short text requests, long outputs, reasoning tasks, image generation, coding, multimodal requests and other workloads with different computational requirements.[11, 13] A billion lightweight text interactions therefore cannot be treated as environmentally equivalent to a billion long-reasoning requests or a billion generated videos.

Scale also affects infrastructure beyond the energy required for individual requests. Large numbers of concurrent users require enough servers, accelerators, networking equipment, cooling systems and electrical capacity to meet peak demand. Some of that infrastructure consumes energy even when it is not operating at full utilization, meaning total resource demand is not simply the sum of perfectly efficient individual transactions.[1]

This creates an important distinction between per-request efficiency and total resource demand. An AI system can become much more efficient per task while its overall electricity and water footprint continues to grow if usage expands faster than efficiency improves.

For example, if a future system used 80 percent less water per comparable task but the number of tasks increased tenfold, total water use would still be approximately twice the original amount. This is an illustrative mathematical example, not a prediction of future AI demand.

The environmental significance of AI therefore depends on both sides of the equation:

- lower resource use per computation × rapidly increasing amounts of computation.

A tiny footprint at the level of one interaction does not automatically imply a tiny industry-wide footprint, just as a large industry-wide total does not mean that every individual prompt consumes a large amount of water.

## 5. How Big Is AI's Total Water Footprint?

Estimating the total water footprint of artificial intelligence is considerably more difficult than estimating the footprint of an individual workload. AI operates within the much larger data-center industry, where computing infrastructure is often shared among AI, cloud services, storage, search, streaming, enterprise applications and other digital activity. Available national statistics therefore frequently measure data centers as a whole rather than AI in isolation.

The broader data-center baseline is nevertheless important. Data centers already represent a significant and rapidly growing source of electricity and water demand, and AI is contributing to that growth. But all data-center water use should not be attributed to AI. Determining AI's share requires either workload-specific measurements or modeling assumptions about how much infrastructure and electricity are actually devoted to AI.

For that reason, estimates in this section must be separated according to what the evidence actually represents. Measured facility or industry data, modeled AI estimates and company-reported figures are different forms of evidence and should not be combined as

though they provide the same degree of certainty. Future projections require an additional distinction because they depend on assumptions about deployment, hardware efficiency, cooling technology, electricity generation and the growth of AI workloads.

Scale is ultimately more consequential than the footprint of any single chatbot interaction. Even when individual AI tasks require relatively little water, rapidly expanding infrastructure can create substantial aggregate demand. Conversely, improvements in hardware, cooling and electricity can reduce the amount of water associated with each unit of computing even as total capacity grows.

This section therefore examines how much water the expanding AI infrastructure system requires in aggregate, how much of that demand can actually be distinguished from conventional data-center activity, and how that footprint could change as AI deployment accelerates.

## 5.1 U.S. Data Centers Overall

Before attempting to isolate AI's water footprint, it is useful to establish the scale of the broader data-center industry. Data centers support far more than generative AI. They provide infrastructure for websites, cloud computing, business applications, streaming services, financial transactions, social media, file storage, scientific computing and many other digital services. AI is an increasingly important source of data-center demand, but the two categories are not interchangeable.

The most comprehensive recent federal assessment comes from Lawrence Berkeley National Laboratory's United States Data Center Energy Usage Report. The 2024 report estimated that U.S. data centers consumed approximately 176 terawatt-hours of electricity in 2023, equivalent to about 4.4 percent of total U.S. electricity consumption. That was up from approximately 58 terawatt-hours in 2014.[15]

The report also documented rapid growth in electricity consumption from GPU-accelerated AI servers. Electricity consumption from GPU-accelerated AI servers increased from less than 2 terawatt-hours in 2017 to more than 40 terawatt-hours in 2023. Conventional servers remained significant, however, consuming nearly 60 terawatt-hours that year.[15] These figures illustrate why rapid data-center growth can increasingly be associated with AI without treating all data-center electricity use as AI use.

Data centers also consume water directly. Berkeley Lab estimated total direct U.S. data-center water consumption at approximately 66 billion liters, or about 17.4 billion gallons, in

2023. Most of this direct consumption was associated with cooling systems, particularly evaporative cooling.[15]

The broader water footprint was considerably larger when electricity generation was included. The report estimated that electricity consumed by U.S. data centers in 2023 was associated with approximately 800 billion liters, or about 211 billion gallons, of indirect water consumption.[15] These figures illustrate why looking only at water used within the walls of a data center can substantially understate the broader water associated with its operation.

These national totals should not be labeled as AI water consumption. They include the full range of computing performed in U.S. data centers, and the indirect estimate depends on the water intensity of electricity generation. They also represent national totals that combine facilities operating in very different climates, watersheds and electrical grids.

Berkeley Lab's 2024 report projected total data-center electricity consumption of approximately 325 to 580 terawatt-hours by 2028, equivalent to about 6.7 to 12 percent of U.S. electricity consumption under its scenarios.[15]

The estimates are already being updated as the industry changes. A Berkeley Lab update released in 2026 extended the analysis to 2030 and estimated a central value of 11.8 percent of U.S. electricity consumption, with scenarios ranging from 9.5 to 15.3 percent.[16]

These figures establish the scale of the infrastructure in which AI operates, but they do not answer how much of its water footprint belongs specifically to AI. Separating AI workloads from conventional cloud and data-center activity requires a narrower analysis.

## 5.2 What Portion Is Actually AI?

Separating AI from the broader data-center footprint remains difficult because facilities and supporting infrastructure can serve multiple workloads. National data-center electricity and water totals therefore cannot simply be labeled as AI totals.[13, 15]

There are, however, ways to estimate part of AI's contribution. Berkeley Lab separates conventional servers from servers dedicated to AI workloads. In its 2024 U.S. analysis, GPU-accelerated AI servers consumed more than 40 terawatt-hours of electricity in 2023, compared with nearly 60 terawatt-hours for conventional servers. AI-accelerated server electricity use had grown from less than 2 terawatt-hours in 2017.[15]

Those figures show that AI had already become a substantial component of server electricity demand by 2023, but they should not be divided directly by the 176 terawatt-

hours of total U.S. data-center electricity use to declare that a specific percentage of the industry's entire footprint was AI. The national total also includes storage, networking equipment and facility infrastructure such as cooling and power systems. Some of that supporting electricity serves AI workloads, but allocating it precisely requires assumptions about which equipment and infrastructure loads belong to which computing activities.[15]

Water is even harder to separate. Berkeley Lab estimated approximately 66 billion liters of direct data-center water consumption and about 800 billion liters of indirect electricity-related water consumption in 2023, but it did not report those national water totals as separate AI and non-AI categories.[15] A defensible AI-only national water total therefore cannot be obtained simply by taking the report's overall data-center water figure and assigning an assumed percentage to AI.

The International Energy Agency encounters a similar measurement problem at the global level. It distinguishes conventional computing from accelerated servers, whose growth is driven mainly by AI adoption, and estimates that accelerated servers will account for a large share of future data-center electricity growth. However, the IEA also emphasizes substantial uncertainty surrounding both current and future data-center electricity demand and uses scenario-based modeling rather than treating AI's contribution as a perfectly observable quantity.[13]

More recent evidence makes the direction of change clearer even if the exact water share remains uncertain. The IEA reported in 2026 that global electricity consumption from AI-focused data centers increased approximately 50 percent in 2025, substantially faster than the 17 percent increase for data centers overall. It also projects electricity consumption from AI-focused data centers to roughly triple by 2030.[13]

This creates an important distinction between what is known and what is inferred. Available evidence strongly supports the conclusion that AI is a major and rapidly growing driver of data-center electricity demand. It does not currently support assigning AI a similarly precise share of total U.S. data-center water consumption.

For this report, national figures are therefore labeled according to what they actually measure. Data-center totals describe the entire industry. AI-server estimates describe identifiable AI-oriented computing equipment. AI-specific water figures are used only where a study directly models or measures AI workloads. This prevents growth in AI infrastructure from being understated without attributing every gallon of water used by the broader digital economy to AI.

## 5.3 Current AI Estimates

Current estimates of AI water use do not all measure the same thing. Some come from instrumentation of operating AI infrastructure, some are produced by researchers using models and assumptions, and others are figures disclosed by the companies developing or operating AI systems. These categories overlap in some cases, but distinguishing them is important when deciding how much confidence to place in a number and what that number actually represents.

### *Measured*

The strongest current per-request evidence comes from Google's 2025 study of Gemini Apps. Google instrumented its production serving infrastructure and accounted for active AI accelerators, host CPUs and memory, idle provisioned capacity and data-center overhead. For the median Gemini Apps text prompt measured in May 2025, the researchers reported approximately 0.24 watt-hours of electricity and 0.26 milliliters of water consumption.[10]

Even this measurement has an important boundary. The 0.26-milliliter figure represents on-site operational water associated with Google's data-center cooling, rather than a complete lifecycle water footprint incorporating electricity generation, semiconductor manufacturing and other supply-chain impacts.[10] It therefore answers a narrower question than studies attempting to estimate AI's total water footprint.

Measured figures also remain unusually scarce. AI companies generally do not publish independently comparable, workload-level measurements showing the water consumed by individual models or services. Google's result is valuable partly because detailed production measurements of this kind remain uncommon.

### *Modeled*

Modeled estimates fill some of the gaps that direct measurements cannot. The GPT-3 analysis discussed earlier estimated water consumption by combining modeled inference energy with data-center cooling and electricity-related water intensity.[9] This approach allowed researchers to examine how water consumption changes across locations and times even without direct access to every underlying facility.

The 2026 Joule study similarly modeled energy demand across frontier-scale inference workloads, estimating a median 0.31 watt-hours for a standard query and approximately 3.91 watt-hours for a long reasoning workload under its modeled conditions.[11] These are

energy estimates rather than direct water measurements, so converting them into water consumption would require additional assumptions about cooling and electricity generation.

Modeled estimates are not inherently inferior to measured figures. They can answer questions that facility measurements cannot, test alternative scenarios and expose how results change when assumptions change. But their conclusions remain dependent on the models, system boundaries and input data used.

### *Company-Reported*

A third category consists of figures publicly disclosed by AI companies. These can provide information unavailable elsewhere because companies have direct access to their infrastructure, but the level of methodological detail varies substantially.

For example, OpenAI CEO Sam Altman stated in 2025 that an average ChatGPT query uses approximately 0.34 watt-hours of electricity and 0.000085 gallons of water, equivalent to about 0.32 milliliters.[17] Unlike Google's study, however, the statement was not accompanied by a detailed methodology explaining the workload definition, infrastructure boundary, cooling assumptions, geographic scope or whether indirect electricity-related water was included. It should therefore be treated as a company-reported estimate rather than an independently reproducible measurement.

Mistral AI provides a useful example of why boundaries matter. In a 2025 lifecycle assessment conducted with Carbone 4 and France's ADEME, and reviewed by environmental consultancies, Mistral reported approximately 45 milliliters of water for the marginal impact of generating a 400-token response using Le Chat. The assessment uses a broader lifecycle framework that incorporates impacts beyond immediate data-center cooling, including upstream infrastructure impacts.[18]

The difference between approximately 0.26 milliliters for Google's median Gemini prompt, 0.32 milliliters reported for an average ChatGPT query, and 45 milliliters for Mistral's 400-token Le Chat response does not establish that Le Chat uses roughly 140 times more water than ChatGPT or 170 times more than Gemini. The figures have different system boundaries, methodologies, workloads and levels of disclosure.[10, 17, 18]

That is one of the central lessons of the available evidence. A water number without its measurement boundary can be misleading even when the number itself is accurate.

For this report, AI water estimates are therefore identified according to their evidence type and scope:

- Measured — derived from instrumentation or observation of operating infrastructure.
- Modeled — calculated using empirical data combined with assumptions or models.
- Company-Reported — publicly disclosed by an AI provider but not necessarily independently reproducible from the information provided.

A figure can belong to more than one category. Google's Gemini measurement, for example, is both measured and company-produced. The labels are intended to clarify the evidence behind each number rather than rank every estimate from best to worst.

The available evidence consequently does not produce one authoritative answer to "how much water does AI use?" Instead, it produces different answers to different questions: water consumed at a particular facility, water attributable to electricity, lifecycle water associated with a model, water per defined workload, or aggregate water across an industry.

## 5.4 Future AI Water Projections

Estimates of AI's future water footprint are necessarily projections rather than measurements. Researchers must make assumptions about how quickly AI demand will grow, how much computing future models will require, what hardware will be deployed, where data centers will be located, how they will be cooled and what electricity sources will supply them. Future water estimates should therefore be interpreted as scenarios of what could occur under specified assumptions, not as predetermined outcomes.

One of the most widely cited global projections comes from the research behind Making AI Less "Thirsty." Based on an earlier estimate that global AI electricity demand could reach approximately 85 to 134 terawatt-hours in 2027, the researchers projected combined direct and electricity-related AI water withdrawal of approximately 4.2 to 6.6 billion cubic meters in 2027.[9]

That figure is frequently described incorrectly as water "consumption." The researchers separately projected water consumption of approximately 0.38 to 0.60 billion cubic meters, or 380 to 600 billion liters, in 2027.[9] The difference is substantial: the larger 4.2-to-6.6-billion-cubic-meter estimate represents water withdrawn from sources, much of which may be returned, while the smaller figure represents water projected to be consumed, primarily through evaporation.

The authors explicitly characterize these calculations as rough first-order estimates subject to future uncertainty. Water efficiency could improve, electricity generation could

change, AI demand could differ from the underlying projection, and the geographic distribution of workloads could alter both direct and indirect water requirements.[9]

More recent research has attempted to narrow the analysis specifically to AI servers in the United States. A 2025 study published in Nature Sustainability modeled U.S. AI-server deployment from 2024 through 2030 under multiple growth scenarios. The researchers projected an annual operational water footprint of approximately 731 million to 1.125 billion cubic meters, depending on the scale and characteristics of AI-server expansion.[19]

In that study, approximately 71 percent of the modeled operational water footprint was indirect water associated with electricity generation, while 29 percent was direct water use at data centers. The researchers also found substantial geographic variation. Factors including climate, cooling efficiency, electrical-grid water intensity and where AI servers are located can significantly change the resulting footprint.[19]

This geographic effect is particularly relevant to this report because the study found that southern states, including Florida, generally exhibited higher modeled data-center Power Usage Effectiveness and Water Usage Effectiveness than cooler northern states such as Washington, reflecting in part the influence of climate.[19] This does not establish the water footprint of a future Tampa Bay data center, but it reinforces why national averages cannot simply be assigned to Florida facilities.

The same research also demonstrates why future demand should not be presented as inevitable. Under modeled best practices involving improved water and energy efficiency, strategic server location and changes to the electrical grid, the researchers found that water footprints could be reduced substantially relative to less favorable scenarios. Their combined best-practice scenarios produced reductions of up to 86 percent in the modeled water footprint.[19]

These projections therefore support two conclusions at the same time. AI infrastructure could become a substantial source of future water demand if deployment expands rapidly. However, the eventual footprint is highly sensitive to decisions that have not yet been made. Hardware efficiency, cooling technology, electricity generation, facility location and water-management practices can all influence the outcome.

For the remainder of this report, future estimates are therefore treated as projected, and withdrawal and consumption are reported separately whenever the underlying research distinguishes them. A projection of billions of cubic meters of withdrawal should not be rewritten as billions of cubic meters “used up,” just as a modeled future footprint should not be presented as water that AI is already consuming today.

## 5.5 Why Scale Matters More Than the Viral Prompt

Public discussion of AI water use often focuses on the smallest unit that is easiest to imagine: a single chatbot prompt. Per-request estimates are useful for understanding how AI workloads translate into physical resource demand, but they can also distract from the larger environmental question. At industry scale, the more consequential issue is not whether one prompt consumes a few drops or several milliliters of water. It is the amount and location of infrastructure being built to support rapidly growing demand.

Data centers operate continuously and can contain thousands or tens of thousands of servers. Supporting that computing capacity requires electricity, cooling equipment, electrical substations, backup systems, networking infrastructure and, in some facilities, substantial water infrastructure. As AI demand grows, its environmental footprint is therefore shaped not only by the efficiency of individual requests but by how much total computing capacity is installed and how intensively that capacity is used.[15, 19]

This distinction helps reconcile two findings that may initially appear contradictory. Modern text inference can have a very small per-request operational footprint. Google, for example, measured approximately 0.26 milliliters of direct water consumption for its median Gemini Apps text prompt in 2025.[10] At the same time, national studies project substantial growth in electricity and water demand from AI infrastructure as the number and complexity of workloads increase.[19] Both can be true: an increasingly efficient individual activity can still produce growing aggregate demand when it is performed at enormous and expanding scale.

Infrastructure decisions also persist much longer than individual AI interactions. The cooling system selected for a new data center, the watershed supplying its water, the electrical grid serving it and the location in which it is built can influence resource demand for years. A facility designed around potable-water-intensive cooling in a water-stressed area presents a different environmental problem from a comparable facility using low-water cooling or reclaimed water in a location with adequate resources, even if both facilities run the same AI models.[1, 19]

Scale also changes the importance of local impacts. National or global totals can make water demand appear broadly distributed, but data centers are physical facilities drawing resources from particular places. A relatively small percentage of national water consumption can still matter if demand is concentrated within a watershed, utility system or community already facing competing needs. Conversely, a large computing facility does not automatically create a water crisis if its demand can be accommodated through appropriate technology, water sources and infrastructure.

This is why debates framed around whether one AI prompt “uses too much water” can miss the more actionable question. Individual users generally do not choose the cooling architecture, electricity source or geographic location serving each request. Developers, technology companies, utilities and governments make the infrastructure decisions that determine much of AI's eventual water footprint.

The shift from prompts to infrastructure also changes what effective solutions look like. Improving model efficiency matters, but so do facility siting, cooling technology, reclaimed-water access, electricity generation, drought planning, transparency and regional infrastructure capacity.[19] These decisions can affect millions or billions of future AI interactions simultaneously.

The viral prompt-level claims are therefore useful as an entry point into the issue, but they are not where the analysis should end. The environmental significance of AI water use is ultimately a question of scale, location and infrastructure, rather than simply the number of milliliters assigned to one conversation.

## 6. AI vs. the Rest of the Internet

AI is often discussed as though it represents an entirely new category of digital resource consumption. In reality, the internet already depends on a vast physical infrastructure of data centers, telecommunications networks, servers and electricity systems that supports web searches, social media, video streaming, cloud storage, uploads, video conferencing and countless other online activities. Those services already carry energy, water and material footprints.

That does not mean AI has the same footprint as conventional digital activity. Some AI workloads, particularly long reasoning, multimodal generation and other compute-intensive applications, can require substantially more computation than simpler online interactions. At the same time, the enormous variation among both AI and conventional internet activities makes broad comparisons such as “one AI prompt versus one Google search” difficult to defend without measurements produced using compatible methodologies and system boundaries.

The comparison becomes especially problematic when studies count different parts of the lifecycle. An estimate measuring only the electricity and cooling associated with AI inference cannot legitimately be compared directly with a social-media estimate that includes devices, networks, data centers, manufacturing and other lifecycle impacts. A numerical comparison can look precise while actually measuring different systems.

There is also an increasingly important conceptual problem with separating “AI” from “the Internet.” Search engines, recommendation systems, advertising, content moderation, social-media feeds and other digital services already incorporate machine learning and AI. As generative and agentic systems become embedded in existing services, the boundary between an “AI interaction” and an ordinary internet interaction will become even less distinct.

This section therefore does not attempt to manufacture a universal ranking of which digital activity “uses the most water.” Instead, it asks what can actually be compared using compatible evidence, what cannot, whether AI appears unusually resource-intensive where meaningful comparisons are possible, and where the evidence remains insufficient to answer the question at all.

## 6.1 What Conventional Digital Activity Already Uses

AI did not introduce resource consumption to the internet. Long before the recent growth of generative AI, ordinary digital activity already depended on data centers, telecommunications networks, routers and modems, and billions of end-user devices. Operating this infrastructure requires electricity, while manufacturing and replacing it requires energy, water, minerals and other materials.[20]

A 2024 peer-reviewed study published in Nature Communications examined the lifecycle environmental impacts of several major forms of everyday digital activity using a common methodology. The researchers evaluated web surfing, social media, video streaming, music streaming and video conferencing while accounting for the infrastructure required to deliver and consume that content, including data centers, transmission networks, customer-premise equipment and end-user devices.[20]

For the global average user modeled in the study, annual internet consumption included approximately 730 hours of web surfing, 894 hours of social media, 833 hours of video streaming, 566 hours of music streaming and 207 hours of video conferencing. These activities generated different impacts partly because they require different amounts of data transmission and different end-user devices. Social media, for example, was modeled at approximately 0.31 gigabytes of data traffic per hour, while video streaming required approximately 1.3 to 2.4 gigabytes per hour depending on resolution.[20]

Video streaming produced the largest share of the environmental impacts associated with the digital activities examined. Across the study's impact categories, streaming accounted for approximately 40 to 52 percent of the total, while web surfing, social media, music streaming and video conferencing generally contributed approximately 10 to 18 percent

each. This reflected not only data traffic but also differences in the devices used to consume each type of content.[20]

The study also demonstrates why the environmental footprint of internet activity extends beyond data centers. A person streaming video, scrolling social media or participating in a video conference uses a chain of physical infrastructure: servers process and store information, networks transmit it, routers and modems deliver it, and phones, televisions or computers display it. Manufacturing and operating each part of that chain contributes to the lifecycle footprint.[20]

Cloud storage and uploads operate through much of this same infrastructure, but they were not evaluated as separate user-activity categories in this particular study.[20] For that reason, this report does not assign them numerical impacts using these results. Uploading and storing files clearly require data transmission and data-center infrastructure, but a fair quantitative comparison would require compatible evidence defining factors such as file size, storage duration, redundancy and frequency of access.

The same caution applies to search. The study includes search activity within the broader category of web surfing rather than providing a separate environmental footprint for an individual Google or other search-engine query.[20] Consequently, its web-surfing results should not be converted into a supposed “water per Google search” number.

This broader context matters when evaluating AI. Digital activity has always had a physical footprint. The relevant question is not whether AI uniquely transformed a previously resource-free internet into one that consumes electricity, water and materials. It is whether AI workloads require significantly different resources from conventional digital activities, how quickly those workloads are growing, and whether their additional infrastructure creates new or intensified environmental pressures.

## 6.2 What Can Actually Be Compared

Comparing the environmental footprint of AI with conventional internet activity requires more than finding two published numbers and placing them beside each other. The measurements must describe sufficiently similar system boundaries, units and environmental impacts. Otherwise, an apparently precise comparison can be misleading.

A useful example is the digital-content study discussed in the previous section. Because web surfing, social media, video streaming, music streaming and video conferencing were evaluated using the same lifecycle-assessment methodology, their results can be compared with one another.[20] Differences among those activities reflect differences

identified within a common analytical framework rather than differences created simply by using unrelated studies.

The same principle applies within AI research. The 2024 Power Hungry Processing study measured more than 80 machine-learning models using a common experimental methodology. Its results can therefore support comparisons among the AI tasks and models it tested, including its finding that image generation was substantially more energy-intensive than text generation under those experimental conditions.[12]

Similarly, the 2026 Joule study evaluated standard text inference and long reasoning workloads within the same modeling framework. Its estimates of approximately 0.31 watt-hours for a standard query and 3.91 watt-hours for a long reasoning workload can reasonably be compared with each other because the researchers constructed both estimates using the same underlying methodology.[11]

Comparisons become much weaker when measurements with different boundaries are combined. Google's estimate of approximately 0.26 milliliters of water for a median Gemini Apps text prompt measures operational water associated with its production infrastructure.[10] Mistral's approximately 45-milliliter estimate for a 400-token Le Chat response was developed using a broader lifecycle methodology.[18] Placing those values in a chart labeled simply “water used per AI prompt” would imply that they measure the same thing when they do not.

The same problem occurs when comparing AI with conventional internet activity. A lifecycle estimate for an hour of social media use may include the manufacture and operation of a smartphone, network infrastructure and data centers.[20] A per-query AI estimate may include only electricity consumed during inference or water consumed at the data center.[10] Comparing those values directly would attribute environmental impacts to one activity that were excluded from the other.

Units also need compatible functional boundaries. One prompt, one hour of use, one gigabyte of data and one inference are different units of activity. Converting between them requires additional assumptions about user behavior and workload characteristics. Those assumptions can create the appearance of an apples-to-apples comparison without actually producing one.

For this report, quantitative comparisons are therefore made only when the underlying evidence provides a sufficiently common basis. This includes:

- comparing conventional digital activities evaluated within the same lifecycle study;[20]

- comparing AI models or task categories tested within the same experimental study;[12]
- comparing standard and reasoning inference modeled within the same study;[11]
- comparing changes within a company's own consistently reported operational metrics when the reporting boundary remains sufficiently stable.

Results from incompatible studies can still be presented together to illustrate the range of published evidence, but they should be identified by their different boundaries rather than used to rank services or activities.

This approach limits the number of direct quantitative comparisons that can be made, but reduces the risk of drawing conclusions from incompatible measurements. A comparison is only as meaningful as the consistency of the measurements being compared.

## 6.3 What Cannot Honestly Be Compared

Some of the most appealing comparisons between AI and conventional internet activity are also the least defensible. Published estimates often measure different portions of the digital system, use different units and answer different environmental questions. Putting those numbers side by side without preserving those differences can create a comparison that the underlying research never established.

For example, Google's estimate of approximately 0.26 milliliters of water for a median Gemini Apps text prompt represents operational water associated with serving that request through Google's production infrastructure.[10] The Nature Communications study of conventional digital activity uses a lifecycle framework that includes data centers, transmission networks, customer-premise equipment and end-user devices, including impacts associated with manufacturing equipment.[20] Those measurements do not share the same boundary.

It would therefore be misleading to compare the Gemini figure directly with the lifecycle footprint of social media or video streaming and conclude that one activity “uses more water” than the other. The conventional-internet calculation includes physical systems that the Gemini operational figure excludes. Operational AI inference and lifecycle digital-media consumption are not equivalent measurements.

The same problem prevents a clean comparison between an AI prompt and a Google search using many commonly repeated figures. Estimates for individual web searches have circulated for years, but they often originate from older infrastructure, company statements or calculations with different definitions and system boundaries. Meanwhile,

contemporary AI estimates range from measured operational cooling water to modeled electricity-related water and full lifecycle assessments. Without a modern study evaluating both activities under a common methodology, dividing one number by the other does not establish a reliable ratio.

Comparisons among AI providers can suffer from the same problem. As discussed earlier, Google's Gemini estimate, OpenAI's reported ChatGPT estimate and Mistral's Le Chat lifecycle estimate should not be used to rank those services by water efficiency.[10, 17, 18] Their reported figures differ in methodology, workload definition and system boundary. A larger published number may simply include environmental impacts that a smaller number excludes.

Time-based comparisons also require caution. One hour of video streaming and one AI prompt are not naturally equivalent units. A person could submit one prompt or hundreds during an hour, while a generated response could contain a few words or thousands. Converting either activity into a common time unit would require assumptions about user behavior that are separate from the measurements themselves.

Nor can data volume alone solve the problem. A gigabyte of streamed video and a gigabyte of AI-related network traffic may contain the same amount of transmitted data but require very different amounts of computation before that data is delivered. Conversely, a computationally intensive AI request might produce relatively little network traffic. Data transferred is not a universal proxy for computation performed.

These limitations do not make comparison impossible. They establish what additional evidence would be needed to make the comparison meaningful. Ideally, AI and conventional digital activities would be evaluated using the same lifecycle boundary, geographic assumptions, electricity data, hardware accounting and water methodology, with results expressed in clearly defined functional units.

Until such evidence is available, this report does not use incompatible estimates to produce claims such as:

- “one AI prompt uses X Google searches worth of water,”
- “one generated image equals X minutes of social media,” or
- “ChatGPT uses X times more water than TikTok.”

Those statements may be easy to communicate, but the available evidence reviewed here does not provide a sufficiently consistent basis for calculating them.

The more defensible conclusion is narrower: some AI workloads require substantially more computation than others, conventional internet activity already has a measurable physical

footprint, and both can contribute to electricity, water and material demand. Exactly how they compare requires measurements designed to answer that question.

## 6.4 Is AI Unusually Resource Intensive?

The available evidence does not support a simple answer that AI is either uniquely wasteful or environmentally insignificant. AI is part of the same physical digital infrastructure that already supports web browsing, streaming, social media, cloud computing and other online activity, but some AI workloads can require substantially more computation than conventional digital tasks or simpler forms of AI inference.[12, 13, 20]

At the level of an individual interaction, ordinary text inference can be relatively modest. Google measured approximately 0.24 watt-hours of electricity for its median Gemini Apps text prompt in 2025,[10] while the independent 2026 Joule study estimated a median of approximately 0.31 watt-hours for standard frontier-scale text inference.[11] These results do not suggest that every ordinary text interaction is inherently an enormous computational event.

However, “AI” encompasses much more than ordinary text inference. Within studies using consistent methodologies, resource requirements can differ dramatically by task. The Power Hungry Processing study found that image generation was substantially more energy-intensive than text generation among the models it tested,[12] while the Joule study estimated that a long reasoning workload required approximately 13 times the energy of its standard text-query scenario.[11] Emerging workloads such as video generation can be more computationally demanding still.[13]

AI also differs from many conventional internet activities in an important way: much of its workload involves creating new computation in response to a request, rather than primarily retrieving, transmitting or displaying previously stored information. A web page, stored video or cached file can often be delivered repeatedly without recreating its underlying content. Generative AI may instead perform substantial inference each time a user requests a new response, image or other output. This does not mean retrieval and streaming are resource-free, but it helps explain why some generative workloads can add significant computing demand even when the amount of data ultimately delivered to the user is relatively small.[11, 13]

Whether that makes AI “more resource intensive than the internet” cannot currently be expressed as a defensible universal ratio. Conventional digital activities have lifecycle impacts from data centers, networks and end-user devices,[20] while many current AI measurements cover only inference or data-center operations. As established in the

previous sections, those boundaries are too different to support claims that AI as a whole uses a particular multiple of the resources required by search, social media or streaming.

At the infrastructure level, however, there is stronger evidence that AI is unusually important as a source of new demand. Berkeley Lab found rapid growth in electricity consumption from GPU-accelerated AI servers in the United States,[15] and the International Energy Agency reports that AI-focused data-center electricity consumption is growing substantially faster than data-center electricity consumption overall.[13] AI's significance therefore comes not only from the intensity of particular workloads, but from the combination of increasingly compute-intensive applications and rapid expansion in their use.

Efficiency improvements complicate the picture further. Modern hardware and optimized serving systems can substantially reduce the resources required for a given computation.[11] A newer AI system can therefore perform much more computation without its resource requirements increasing proportionally. At the same time, efficiency gains can be outweighed by growth in the number, size and complexity of workloads being served.

The evidence supports a more precise conclusion than either “AI is just another internet activity” or “AI is uniquely destroying resources.” Some AI workloads are unusually compute-intensive, ordinary text inference can be relatively efficient, and AI is currently a major driver of growth in data-center infrastructure. Its environmental significance depends on the type of AI being used, the scale at which it is deployed and the physical infrastructure supporting it.

For water specifically, computational intensity alone is still not enough to determine impact. A highly compute-intensive workload running in an efficient facility with low-water cooling and a low-water electricity supply may have a different water footprint from a less computationally intensive workload operating under less favorable conditions.[1, 19] AI's computational intensity creates demand; infrastructure and location largely determine how that demand translates into water impact.

## 6.5 AI Is Becoming Part of “Normal Internet Use”

The distinction between “using AI” and “using the internet” is becoming increasingly difficult to draw. Generative chatbots are highly visible examples of AI because users intentionally interact with them, but artificial intelligence is also being incorporated into services that people may continue to experience simply as search engines, social-media feeds, online shopping, navigation, email and other ordinary digital tools.

Search provides a clear example. Modern search engines increasingly use generative AI to summarize information, answer questions and organize results in addition to retrieving traditional webpages. A person conducting what appears to be an ordinary web search may therefore trigger both conventional information-retrieval systems and generative AI computation within the same interaction.[21]

Recommendation systems create another overlap. Social-media platforms, streaming services and online marketplaces have long used machine learning to rank content and predict what users are likely to engage with. AI can determine which posts appear in a feed, which videos or songs are recommended and which products are displayed. Automated systems are also used for advertising, spam detection, content moderation, translation and other background functions that users may never consciously identify as AI.[13]

This creates an accounting problem. If AI computation becomes embedded within search, social media, cloud software and other digital services, separating the environmental footprint into neat categories of “AI use” and “non-AI internet use” becomes increasingly artificial. Part of the electricity consumed by a search engine or social-media platform may now support AI workloads even when the user never opens a dedicated chatbot.

The overlap also complicates comparisons based on user behavior. A person watching a recommended video may trigger AI systems used to select or rank that video, but most of the data transmitted during the interaction may still come from conventional video-delivery infrastructure. Similarly, an AI-assisted search can involve model inference, traditional search infrastructure, networking and the user's device. Assigning the entire environmental footprint to either “AI” or “ordinary internet use” would obscure the fact that both systems contributed to the activity.

Measuring AI's footprint nevertheless remains useful. Dedicated AI accelerators, training clusters and AI-focused data centers can still be identified and studied, and researchers can estimate the additional computing associated with particular AI functions.[15] What becomes harder is translating those infrastructure measurements into a simple division of individual internet activity between AI and everything else.

The trend also matters when interpreting future projections. Some growth in AI computing may come from people intentionally using new generative services, while another portion may result from AI being incorporated into products and services people already use. Future AI demand therefore does not necessarily require billions of people to consciously decide to “use AI” more often. AI computation can grow as existing digital activities themselves become more computationally intensive.[13]

As these technologies increasingly overlap, “AI versus the internet” is becoming less useful as a literal technological division. AI increasingly operates as one layer of the internet's existing physical and computational infrastructure. The more meaningful environmental questions are how much additional computation these systems create, whether they displace or supplement existing workloads, and how the infrastructure supporting that computation is powered and cooled.

## 6.6 What We Still Cannot Measure

Despite rapidly growing research on AI energy and water use, several comparisons that appear simple remain difficult to make reliably. In particular, the available evidence reviewed for this report does not provide a rigorous, contemporary, apples-to-apples measurement of the water footprint of an AI prompt compared with a conventional web search, an hour of social media or other common internet activities.

The problem is not a complete absence of numbers. There are estimates for all of these activities. The problem is that the numbers frequently measure different things. Google's Gemini study measures operational impacts associated with AI inference,[10] while the Nature Communications study of conventional digital activity uses a broader lifecycle framework incorporating data centers, networks, customer equipment and user devices.[20] Comparing those figures directly would produce a numerical result without establishing a scientifically consistent comparison.

Provider differences remain difficult to measure as well. Published estimates exist for ChatGPT, Gemini and Le Chat, but their methodologies and reporting boundaries differ.[10, 17, 18] Public information is even more limited for many other widely used AI services. The evidence therefore does not currently support a reliable ranking of popular AI products by water consumption.

The increasing integration of AI into conventional internet services creates an additional challenge. A modern search can involve traditional information retrieval and generative AI within the same request.[21] Social-media platforms and other online services similarly use machine-learning systems alongside conventional computing infrastructure. Separating the resource demand of the AI component from the rest of the service requires internal workload and infrastructure data that are generally not available to outside researchers.

Water accounting itself remains inconsistent. Some studies report direct cooling water, others include water associated with electricity generation, and broader lifecycle assessments may also include semiconductor manufacturing and infrastructure. Withdrawal and consumption are not always reported separately. Geographic and

temporal differences can further change results even when the same type of workload is being studied.[1, 9]

There are also major gaps in supply-chain measurement. Semiconductor fabrication is known to require substantial amounts of high-quality water, but attributing a portion of that manufacturing footprint to one model, one inference or one digital service requires assumptions about hardware lifetime, utilization and allocation among workloads.[2] As AI accelerators are frequently shared across large numbers of users and applications, there is no universally accepted method for assigning that embedded water to an individual prompt.

These limitations mean that several appealing questions cannot yet be answered with the precision their wording suggests:

- How many Google searches equal one ChatGPT prompt?
- Does an hour on social media use more or less water than an hour using AI?
- Which major chatbot uses the least water?
- How much water does one AI-generated video consume across its full lifecycle?

Existing evidence can provide partial answers under specific assumptions, but it does not currently provide a standardized basis for answering these questions universally.

This is not the same as saying that AI's water footprint is unknowable. Researchers can measure facilities, model workloads, examine electrical grids and conduct lifecycle assessments. The limitation is comparability. Better answers would require standardized reporting of workload characteristics, electricity consumption, direct water consumption, electricity-related water, location and supply-chain impacts across both AI and conventional digital services.

Until those measurements exist, this report does not manufacture precision by combining incompatible numbers. Where evidence supports a comparison, it is made. Where the evidence supports only a range, scenario or partial measurement, it is identified as such. And where a reliable comparison cannot currently be made, the uncertainty itself is part of the finding.

## 7. Where AI and Data Centers Can Actually Cause Harm

The fact that some claims about AI's water footprint are exaggerated or poorly contextualized does not mean the underlying environmental concerns are imaginary. Data centers are large physical facilities with real demands for water, electricity, land and

infrastructure, and under the wrong conditions those demands can create significant environmental and community impacts.

Water impacts are especially dependent on place and time. The same annual consumption can have very different consequences in a water-abundant region than in a stressed watershed, and demand that appears manageable across an entire year can become problematic when concentrated during drought or periods of peak summer use. The source of the water, potable supplies, groundwater, surface water or reclaimed water, also changes the nature of the impact.

Some effects occur beyond the facility itself. Electricity generation can create an indirect water footprint substantially larger than direct cooling demand, while semiconductor manufacturing requires significant quantities of highly purified water upstream. Cooling systems can produce wastewater requiring treatment, and emerging research indicates that very large facilities can also contribute to localized heat effects.

Scale can transform these individual impacts into regional ones. One facility that a utility can comfortably support does not establish that the same system can support a cluster of facilities. Multiple projects drawing on the same water supply, electrical grid and wastewater infrastructure can create cumulative demands that are not apparent when each development is evaluated separately.

The distribution of those costs and benefits also matters. Data centers can bring investment, tax revenue, construction activity and other economic benefits while placing resource or infrastructure burdens on host communities. Whether that distribution constitutes an environmental-justice concern must be established from evidence about the particular community rather than assumed from the existence of a data center alone.

The relevant question, then, is not whether AI or data centers are inherently environmentally harmful. It is where, when and under what conditions their physical infrastructure creates meaningful pressure on water resources, ecosystems and communities. This section examines the circumstances in which those risks become real.

## 7.1 Water-Stressed Communities

The environmental significance of data-center water consumption depends heavily on where and when that consumption occurs. A quantity of water that can be accommodated relatively easily in a water-abundant region may create much greater pressure in a community already experiencing drought, groundwater depletion, limited treatment capacity or competing demand from households, agriculture and other industries.[22]

Recent research illustrates how common this overlap can be. A 2026 study examining more than 4,000 U.S. data-center locations found that nearly 75 percent experienced at least one month of elevated water stress during the year, approximately 30 percent experienced at least one month classified as extreme stress under the study's methodology, and about 12 percent experienced persistent stress throughout the year.[23]

That finding should not be interpreted to mean that three-quarters of U.S. data centers are causing local water shortages. The study measures the water-stress conditions surrounding facilities, not whether each individual facility is responsible for creating those conditions. It does demonstrate that data-center water demand frequently occurs in places where water availability is already constrained at least part of the year.[23]

National and local impacts can look very different. Data-center water consumption can appear relatively modest when compared with total water use across the United States, while still representing a substantial new demand for the individual utility or watershed supplying a large facility. A 2026 peer-reviewed analysis in AGU Advances emphasizes that these localized impacts can be significant even when the industry's aggregate water footprint appears comparatively small.[22]

Individual communities demonstrate what that concentration can look like. In Council Bluffs, Iowa, Google's data center consumed approximately 980 million gallons of water in 2023 and was the municipal utility's largest customer. The AGU analysis reports that the facility represented roughly 21 percent of the city's total water use that year.[22] In The Dalles, Oregon, data-center water demand reached approximately 355 million gallons in 2021, about 29 percent of the city's total water demand, according to the same review of local records and reporting.[22]

These examples should not be generalized to every data center. Facilities differ dramatically in size, cooling technology, water source and climate. They demonstrate instead why national percentages can obscure local consequences. A single large customer can become important to a relatively small municipal water system even while representing an almost invisible fraction of national water use.

Peak demand can matter as much as annual consumption. Data centers using evaporative cooling can require the most cooling water during hot weather, when community electricity and water systems may already be experiencing elevated demand. Emerging U.S. research suggests that the infrastructure required to deliver sufficient water during these peak periods may itself become a constraint on future data-center development.[24]

Indirect water use can create pressure elsewhere as well. Electricity consumed by a data center may be generated outside the watershed where the facility is located, meaning

some of its water footprint can occur at distant power plants. The 2026 stress-adjusted analysis found that off-site electricity-related water stress can dominate the overall water impact in some major data-center markets.[23]

Assessing whether data-center water use creates harm therefore requires more than asking how many gallons a facility consumes. The relevant questions include how much water is available locally, what source is being used, what other users depend on it, whether demand coincides with drought or seasonal scarcity, whether reclaimed or lower-water alternatives are available, and whether the utility has sufficient capacity to serve the facility without reducing the resilience of the broader water system.

A large data center does not automatically create water scarcity simply by existing. But where water is already constrained, large additional demand can intensify an existing problem, reduce the system's flexibility during drought or peak demand, and increase competition for a resource shared with the surrounding community.[22, 23]

## 7.2 Competition for Potable Water

The source of a data center’s water can be as important as the amount it consumes. Many facilities have historically relied on potable municipal water, the same highly treated water supply used by homes, hospitals, schools and businesses. Using potable water for cooling is not inherently harmful, but it can become more difficult to justify where freshwater supplies are constrained or where suitable alternatives are technically and economically available.[2, 22]

Potable water offers practical advantages for cooling systems. Municipal supplies generally provide predictable quality and established delivery infrastructure, while lower-quality water can contain minerals, salts, microorganisms or other constituents that increase scaling, corrosion and treatment requirements. For data-center operators, potable water can therefore be simpler to use than alternative supplies.[5]

That convenience also means using water that has already been treated to drinking-water standards. Drinking water has already undergone treatment to meet public-health standards, consuming infrastructure, chemicals and energy before it reaches the data center. When potable water is subsequently evaporated for cooling, that treated water is removed from immediate local availability rather than remaining available for other municipal uses.

Alternatives can include reclaimed wastewater, treated municipal effluent, captured rainwater, non-potable surface or groundwater where appropriate, and cooling

technologies that reduce or eliminate routine water consumption. The suitability of each option depends on local infrastructure, water chemistry, regulation, climate and facility design.[2, 5]

Reclaimed wastewater is particularly relevant because data-center cooling generally does not require drinking-water quality. Where adequate treatment and distribution infrastructure exist, reclaimed water can substitute for potable supplies and reduce competition with community drinking-water demand. It can also provide a productive second use for wastewater that has already passed through the municipal system.[25]

However, reclaimed water is not environmentally free. It must be collected, treated and transported, and additional treatment may be necessary before it can be used safely in cooling equipment. Reclaimed water can also have competing beneficial uses, including irrigation, industrial applications and environmental purposes. Its availability therefore does not automatically mean that a data center should receive unlimited access to it.

Alternative water sources can also create tradeoffs. Lower-quality water may require additional treatment or more frequent blowdown, while some treatment processes require additional energy or produce concentrated waste streams.[7] A responsible comparison should therefore consider the entire water-management system rather than assuming that replacing potable water automatically eliminates environmental impacts.

The strongest concern arises where large volumes of potable water are consumed in a water-stressed community despite practical lower-impact alternatives. In that situation, the issue is not merely that the data center “uses water.” It is that highly treated freshwater is being allocated to an industrial cooling demand while households, ecosystems and other users depend on the same constrained resource.[22, 23]

This makes water-source selection an important part of data-center planning. Utilities and governments can evaluate whether proposed facilities require potable water, whether reclaimed-water infrastructure can serve them, whether dry or hybrid cooling is practical, and how additional demand fits within long-term water-supply and drought plans.

The appropriate solution will differ by location. In some communities, potable-water use by a data center may represent a manageable fraction of an abundant supply. In others, using drinking-quality water for cooling when feasible alternatives exist can create avoidable competition for a limited public resource. The distinction depends on local evidence rather than the assumption that all data-center water use is equally harmful.

## 7.3 Drought and Peak Demand

Annual water-consumption totals can conceal periods when data-center demand places substantially greater pressure on a local water system. Water availability changes throughout the year, while data centers generally require reliable cooling regardless of whether the surrounding region is experiencing normal conditions, extreme heat or drought.[22, 23]

This matters particularly for facilities that rely on evaporative cooling. Cooling demand generally increases as outdoor temperatures rise, potentially increasing water consumption during hot periods. Those same conditions can coincide with higher community water demand, greater electricity demand and reduced water availability. A facility's annual average therefore may not reveal its demand during the periods when the local system is under the greatest stress.[5, 23]

Drought can intensify this problem. Reduced precipitation, lower streamflow, declining reservoir levels or stressed groundwater supplies can decrease the amount of water available to utilities and other users. A data center that represents a manageable demand during normal conditions may become more consequential when the available supply contracts. Water impact therefore depends not only on how much a facility consumes, but also on when that consumption occurs relative to local water availability.[4, 22]

Recent research reinforces the importance of examining water stress over shorter time periods. The 2026 study of more than 4,000 U.S. data-center locations discussed earlier evaluated water stress monthly rather than relying only on annual averages. Nearly three-quarters of the facilities studied experienced at least one month of elevated water stress, while a smaller portion experienced persistent stress throughout the year.[23] This distinction shows how an annual classification can overlook temporary but potentially important periods of scarcity.

Peak demand also creates an infrastructure problem separate from annual water availability. Public water systems must be capable of delivering enough water when demand is highest, not merely enough to satisfy an annual average. A large facility may therefore require utilities to maintain additional pumping, treatment, storage and distribution capacity even if its total annual consumption remains within the region's overall water supply.[24]

This difference can be illustrated simply. Two facilities could each consume the same amount of water over a year, while one uses water relatively consistently and the other requires much larger volumes during the hottest weeks of summer. Their annual totals

would be identical, but the second facility could impose a greater challenge for a water system that is already approaching its maximum delivery capacity during those periods.

Climate change can further complicate long-term planning by altering temperature, precipitation, drought patterns and water availability. At the same time, hotter conditions can increase cooling requirements. Facilities expected to operate for decades therefore need to be evaluated against future water conditions as well as historical averages.[4, 22]

These risks do not mean that data centers inevitably deprive communities of water during drought. Utilities can use drought-response plans, water-use restrictions, alternative supplies, storage, reclaimed water and cooling strategies designed to reduce demand during constrained periods. But those protections are most effective when peak and drought conditions are considered before large new demands are added to a water system.

Annual gallons alone are therefore an incomplete measure of local impact. A more meaningful assessment asks how much water a facility requires during the hottest and driest periods, whether that demand coincides with other peak users, how much capacity the utility has available, and what happens to facility operations if water supplies become constrained.

## 7.4 Groundwater and Surface-Water Pressure

Data centers ultimately depend on specific water sources. Even when a facility purchases water from a municipal utility rather than withdrawing it directly, that utility may obtain its supply from groundwater aquifers, rivers, reservoirs or combinations of these sources. Large additional demand can therefore affect groundwater and surface-water systems indirectly through the public water supply as well as through direct industrial withdrawals.[22]

Groundwater presents particular concerns because aquifers are replenished at different rates. Pumping water faster than it is recharged can lower groundwater levels over time, while concentrated withdrawals can affect nearby wells and alter interactions between groundwater and connected rivers, wetlands, springs or other surface waters. Whether a data center meaningfully contributes to those effects depends on the aquifer, withdrawal rate, existing users and local hydrology.[25]

Surface-water supplies present a different set of constraints. Rivers, lakes and reservoirs fluctuate with precipitation, runoff, upstream withdrawals and seasonal conditions. A withdrawal that can be accommodated during periods of high flow may represent greater pressure during drought or low-flow conditions. The sustainable availability of surface water therefore cannot necessarily be determined from an annual supply total alone.[25]

Consumption through evaporation adds another dimension. Water evaporated by a cooling system eventually returns to the global water cycle as precipitation, but it is not necessarily returned at the same place or time from which it was withdrawn. From the perspective of the supplying watershed or aquifer, that water has therefore been consumed even though it has not disappeared from Earth. This is why consumptive use can matter locally even when statements that data centers are permanently “destroying” water would be physically inaccurate.[22]

The relationship between groundwater and surface water can also make attempts to separate the two misleading. In many locations, groundwater contributes to streamflow, springs and wetlands, while surface water can recharge aquifers. Sustained changes to one part of the system can therefore affect another. Assessing a proposed facility requires understanding the particular hydrologic system supplying it rather than treating “groundwater” and “surface water” as unlimited independent reservoirs.[25]

Real-world experience also demonstrates that water-source decisions can change when local conditions are considered. The 2026 AGU review describes opposition to proposed groundwater use by a Google facility in South Carolina that ultimately resulted in a negotiated shift toward surface-water alternatives.[22] The significance of that example is not that surface water is universally preferable to groundwater, but that the suitability of a water source depends on local resource constraints.

Water reuse can reduce some of this pressure. In Quincy, Washington, data centers historically used potable groundwater for cooling. The city and Microsoft subsequently developed a reuse system that treats industrial wastewater for reuse as cooling water, reducing reliance on the potable groundwater supply. The case also illustrates the complexity of reuse: concentrated minerals in cooling-tower blowdown created treatment and groundwater-recharge challenges that required additional infrastructure.[26]

The relevant question is therefore not simply whether a data center “uses groundwater” or “uses surface water.” A meaningful assessment asks how much is withdrawn and consumed, how quickly the source is replenished, what other users and ecosystems depend on it, how conditions change seasonally, whether groundwater and surface water are connected, and whether alternative supplies or reuse are feasible.

Where withdrawals remain within the sustainable capacity of the local water system, data-center demand may be manageable. Where large new withdrawals are concentrated in already constrained aquifers or surface-water systems, however, data centers can add genuine pressure to water supplies and the communities and ecosystems that depend on them.[22, 25]

## 7.5 Wastewater and Water Quality

Water quantity is not the only potential local impact of data-center cooling. Facilities that use water-based cooling systems can also generate wastewater, and the composition, treatment and ultimate destination of that wastewater determine whether it creates additional environmental concerns.[7]

Cooling towers provide the clearest example. As water evaporates, dissolved minerals and salts remain behind and become increasingly concentrated in the recirculating cooling water. Operators periodically remove a portion of this water as blowdown to prevent excessive scaling, corrosion and biological growth. Replacement water is then added to the system.[5, 7]

Cooling systems may also use treatment chemicals to maintain equipment and water quality, including corrosion inhibitors, scale-control agents and biocides. Blowdown can therefore contain concentrated constituents originally present in the source water as well as residual chemicals associated with cooling-water treatment. Its composition depends on the source water, treatment program, cooling technology and number of times the water has been recirculated.[7]

What happens next varies by facility. Blowdown may be discharged to a municipal wastewater system, treated on site, reused, or managed through another permitted disposal pathway. Treatment and discharge requirements depend on the composition of the wastewater and applicable local, state and federal regulations. A gallon of wastewater discharged after cooling therefore presents a different environmental question from a gallon of water consumed through evaporation.

The Quincy, Washington, case discussed earlier demonstrates how these issues can intersect. Data-center cooling produced blowdown containing concentrated minerals that complicated disposal and groundwater recharge. The city, Microsoft and other partners developed treatment and reuse infrastructure that allowed water to be recycled for additional cooling while separately managing concentrated waste streams.[26] The case demonstrates both the potential environmental challenge and the possibility of engineering systems designed to reduce it.

Water temperature can matter as well. Where heated water is discharged to surface waters, changes in temperature can affect aquatic environments and are subject to regulatory controls. However, not every data center directly discharges heated cooling water into a river or lake. The environmental implications depend on the facility's cooling

architecture and discharge pathway, so thermal pollution should not be assumed simply because a data center uses water.

Newer liquid-cooling technologies introduce additional questions. Direct-to-chip systems generally circulate coolant within contained loops, while immersion cooling can use specialized dielectric fluids.[8] These technologies may reduce dependence on conventional evaporative cooling, but their environmental performance also depends on coolant chemistry, system integrity, maintenance, disposal and what happens if fluids are released.

Claims involving PFAS and other persistent chemicals require particular caution. Some specialized fluorinated heat-transfer fluids have raised legitimate environmental concerns, but not all liquid-cooling or immersion fluids are fluorinated, and the use of liquid cooling does not by itself establish PFAS contamination of wastewater or surrounding water supplies. Evidence should identify the actual chemical used, its environmental behavior and a plausible release pathway before attributing such contamination to a facility.

Wastewater can also create a tradeoff with water conservation. Increasing the number of times cooling water is recirculated can reduce the amount of freshwater required, but it also increases the concentration of dissolved constituents. Advanced treatment can allow additional reuse, while requiring energy, equipment and management of the resulting concentrated waste.[7, 26] A system optimized solely to minimize freshwater withdrawals may therefore still have water-quality consequences that need to be considered.

For this reason, responsible data-center water management requires attention to both quantity and quality. Important questions include what water enters the facility, what chemicals are added, how many times the water is reused, what leaves as wastewater, what contaminants or concentrated constituents it contains, how it is treated and where it ultimately goes.

A data center's wastewater should not automatically be described as pollution simply because it contains cooling-system discharge. But inadequate treatment, inappropriate disposal, spills or other releases of harmful substances can create genuine environmental impacts. Reducing water consumption is only part of responsible water management; protecting the quality of the water that remains or returns to the environment matters as well.

## 7.6 Electricity-Related Water Use

Some of a data center's water footprint can occur far from the facility itself. As discussed in Section 5, Berkeley Lab estimated that U.S. data centers directly consumed approximately 66 billion liters of water in 2023, while electricity generation associated with those facilities consumed approximately 800 billion liters.[15]

The indirect footprint can therefore be substantial, but its location and environmental significance depend on the electricity sources supplying the facility.

## 7.7 Semiconductor Manufacturing

AI's water footprint begins before a model is trained or a prompt is processed. The processors, memory and other electronic components used in AI servers must first be manufactured, and semiconductor fabrication is itself a water-intensive industrial process. This upstream demand forms part of the embedded or supply-chain water footprint discussed earlier in this report.[2]

Semiconductor fabrication requires exceptionally clean water. During production, silicon wafers undergo repeated processing steps that can leave particles, chemicals and other contaminants on their surfaces. Manufacturers use ultrapure water, or UPW, extensively for rinsing and cleaning wafers, as well as in other fabrication processes where even extremely small concentrations of contaminants can interfere with semiconductor production.[27]

Producing ultrapure water requires additional resources. Incoming water must undergo multiple purification processes to remove dissolved minerals, organic material, particles, microorganisms and other contaminants. A 2024 U.S. Department of Commerce environmental assessment reports that producing 1,000 gallons of ultrapure water can require approximately 1,400 to 1,600 gallons of municipal water, meaning the purification process itself can contribute to the industry's overall water demand.[27]

The volumes involved can be substantial. Using data from 29 semiconductor facilities, the same federal assessment reported median ultrapure-water use equivalent to approximately 1.16 million gallons per day per facility in 2021. Water is also required for facility cooling, emissions-abatement equipment and other operations beyond the fabrication process itself.[27]

Industry-wide research reinforces the scale of this demand. A peer-reviewed analysis of environmental disclosures from 28 semiconductor companies reported approximately 789 million cubic meters of total water withdrawal in 2021. The study found that surface water

and municipal supplies represented the majority of reported withdrawals, while groundwater, third-party supplies and reclaimed water provided smaller shares.[28]

These industry figures should not be described as the water footprint of AI chips specifically. Semiconductor fabrication supplies processors and components for smartphones, vehicles, consumer electronics, industrial equipment, conventional servers and many other technologies in addition to AI. Even within a fabrication plant, attributing water consumption to an individual GPU or other chip requires decisions about how facility-wide water demand should be allocated among products.

AI can nevertheless increase this upstream demand by driving deployment of specialized accelerators and the construction and replacement of servers containing them. A complete lifecycle assessment of AI infrastructure therefore needs to consider hardware manufacturing in addition to the water consumed while that hardware is operating.[2]

Manufacturing location matters here just as it does for data centers. A semiconductor fabrication facility requiring millions of gallons of water per day can have very different environmental consequences depending on the availability of local supplies, competing users, drought conditions and the amount of water that can be reclaimed. Some semiconductor manufacturers recycle substantial portions of their process water, demonstrating that gross withdrawal does not necessarily equal consumptive loss.[27]

Water quality is also part of the issue. Semiconductor manufacturing produces wastewater from processes including wafer rinsing, etching, polishing and ultrapure-water production. EPA documentation identifies wastewater streams containing substances that require treatment, including fluoride, metals and other process-related contaminants.[29]

Semiconductor manufacturing therefore represents a real but frequently overlooked part of AI's broader water footprint. Operational measurements such as “water per prompt” generally do not capture the water required to manufacture the hardware performing that computation. At the same time, the entire semiconductor industry's water footprint should not be attributed to AI.

The appropriate question is how much additional semiconductor manufacturing AI infrastructure creates, where that manufacturing occurs, how water-intensive the fabrication processes are and how effectively manufacturers reduce, recycle and treat their water.

## 7.8 Heat

Data centers do not only consume electricity; they ultimately release most of that energy as heat. Servers, networking equipment and electrical systems generate heat while operating, and cooling systems transfer that heat away from the equipment. Much of it is eventually rejected into the outdoor environment. Until recently, research on data-center environmental impacts focused much more heavily on electricity, carbon emissions and water consumption than on what this concentrated release of heat might do to surrounding communities.[30]

A 2026 peer-reviewed study measured air temperatures in residential neighborhoods downwind of four operating data-center facilities in the Phoenix metropolitan area. The facilities ranged from a 36-megawatt data center to a 169-megawatt campus. Researchers conducted five vehicle-based measurement traverses, comparing downwind temperatures with corresponding upwind areas to investigate the facilities' localized thermal effects.[30]

The researchers found average downwind air temperatures approximately 0.7 to 0.9°C, or 1.3 to 1.6°F, warmer than corresponding upwind areas. During individual measurements, localized differences reached as high as 2.2°C, or about 4°F, and detectable thermal signatures extended as far as approximately 500 meters from facility boundaries.[30]

These results require careful interpretation. The study does not show that data centers have raised the temperature of Phoenix by 4°F, nor that communities surrounding every data center experience persistent warming of that magnitude. The measurements captured localized temperature differences during a limited number of daytime observations around four facilities in a hot, dry metropolitan environment. The 2.2°C result was the largest observed local difference, not an annual or citywide average.[30]

Other emerging research is beginning to examine the phenomenon at larger geographic scales, although much of this work remains preliminary. A 2026 global satellite analysis of nearly 10,000 data centers reported a mean localized land-surface-temperature difference of approximately 0.95°C relative to nearby reference areas. However, that research remains a preprint and has not yet undergone peer review, and satellite-measured land-surface temperature is not the same measurement as the air temperature experienced by people.[31]

A separate 2026 modeling study examined concentrated data-center heat emissions in California and projected localized warming hotspots, with stronger effects on land-surface temperature than near-surface air temperature and greater warming at night. Its maximum modeled July-mean effects reached 3.11°C for surface temperature and 0.80°C for near-surface air temperature. That study is also currently a preprint, meaning its results should

be treated as emerging evidence rather than established measurements of observed community temperatures.[32]

Several mechanisms can contribute to a data center's local thermal footprint. In addition to the direct rejection of heat generated by computing equipment, construction can replace vegetation and permeable land with large buildings, roofs, pavement and other surfaces that absorb and retain heat differently. These effects should not automatically be attributed to server waste heat alone. Separating heat generated by computing from changes in land cover is an important challenge for future research.

The potential consequences also depend strongly on location. Additional heat may be less consequential in some industrial areas than in densely populated neighborhoods already experiencing severe urban heat. Hot climates such as Phoenix are particularly relevant because even relatively small increases can occur on top of already high temperatures, while cooling demand can itself increase as outdoor temperatures rise.[30]

The evidence base remains much younger than the research on data-center electricity and water consumption. More measurements are needed across different climates, facility designs, seasons, times of day and distances from facilities before a representative “temperature increase per data center” could be established.

The emerging evidence nevertheless identifies localized heat as a legitimate environmental impact worth monitoring. Large computing facilities concentrate substantial electricity consumption into relatively small areas, and at least one peer-reviewed field study has now directly measured warmer air downwind of operating data centers.[30] The appropriate conclusion is not that data centers are broadly heating entire cities by several degrees, but that concentrated waste heat can create measurable localized thermal effects whose magnitude and community significance depend on facility size, design, climate and surrounding land use.

## 7.9 Cumulative Development

The impact of a single data center cannot always be evaluated independently from other facilities developing around it. One facility may represent a manageable addition to a local water and energy system, while a cluster of facilities drawing from the same utilities, watershed and electrical grid can create a substantially different infrastructure challenge.

Data centers often cluster geographically because operators benefit from existing fiber networks, electrical infrastructure, available land, skilled labor and proximity to other technology facilities. Northern Virginia provides the most prominent U.S. example, with a

particularly dense concentration of facilities in and around Loudoun and Prince William counties.[33]

This concentration matters because facilities within the same region can depend on the same underlying resources. Several individually manageable projects can collectively increase demand on the same water-treatment plants, pipelines, reservoirs, rivers, wastewater systems and electrical infrastructure. Evaluating each project only against current conditions can therefore underestimate the demand created after other approved or proposed facilities become operational.

The Potomac River basin illustrates why cumulative planning matters. The Interstate Commission on the Potomac River Basin's 2026 analysis identified potential water-supply reliability challenges from projected data-center growth, particularly during low-flow periods.[33] Section 8.2 examines the region's current demand, seasonal variation and future projections.

Cumulative impacts can also extend beyond direct water consumption. Multiple facilities can simultaneously increase electricity demand, and the power generation required to serve them may carry its own water footprint. Additional development can require expansions to water-treatment capacity, pumping systems, transmission mains, wastewater infrastructure and electrical generation and transmission. The infrastructure consequences of a cluster can therefore extend beyond the sum of the water meters at individual facilities.[22, 23]

Cumulative development can also complicate drought planning. Several facilities may individually represent modest shares of annual water demand while collectively requiring substantial water during the same periods of extreme heat and low water availability. Because peak demand can matter more to system reliability than annual averages, regional planning needs to consider existing facilities, projects under construction and reasonably foreseeable development together.[24, 33]

The same principle applies to wastewater and environmental impacts. A receiving water body may be capable of accommodating discharge from an individual facility while the combined effects of multiple discharges, reduced streamflow and increasing consumptive use require additional analysis. ICPRB specifically identifies potential cumulative concerns involving temperature, salinity, minerals and other wastewater constituents as data-center development expands in the Potomac basin.[33]

The infrastructure demands of a data-center cluster depend on facility size, cooling technology, water source, electricity supply, utility capacity, climate and the pace of development. Project reviews should evaluate these factors across the facilities expected

to share regional systems, including their combined demand during drought and peak-use periods.

For communities experiencing rapid data-center development, the relevant question is therefore not only “Can our infrastructure support this facility?” It is also “Can our infrastructure support the total amount of development being built and planned here, including during drought and peak-demand conditions?”

## 7.10 Environmental Justice / Community Distribution

The environmental impacts and economic benefits of data-center development are not necessarily distributed among the same people. A facility can generate investment, tax revenue, construction employment and permanent jobs while also creating demands for electricity, water, land and public infrastructure. An environmental-justice analysis asks who receives those benefits, who bears the environmental and infrastructure burdens, and whether affected communities have meaningful opportunities to influence the decisions.[34, 35]

This question should not begin with the assumption that data centers are systematically located in disadvantaged communities. The available evidence is still developing, and siting patterns vary considerably by region. Instead, environmental-justice concerns should be evaluated using evidence about the socioeconomic characteristics of host communities, existing environmental burdens, resource availability and the distribution of project costs and benefits.

Recent research in California provides one of the clearest attempts to examine this relationship directly. A 2026 analysis led by researchers at Santa Clara University mapped known operating and planned data centers across the state and evaluated their locations alongside water availability and indicators of environmental and socioeconomic vulnerability. The researchers found increasing overlap between data-center development, water-constrained areas and communities experiencing greater socioeconomic or environmental vulnerability.[36] This is evidence of a California pattern, not proof that the same distribution exists nationally.

Research from Northern Virginia provides another perspective. A 2025 mixed-methods study of the region's concentrated data-center development examined community-level effects including water use, noise, infrastructure strain, health concerns and economic burdens. Rather than treating regional economic growth as evidence that benefits are distributed evenly, the researchers specifically examined the power relationships that determine who benefits from development and who experiences its costs.[34]

The economic side of the equation also requires evidence. Data centers can provide substantial capital investment, tax revenue, construction activity and employment, but those benefits vary by project and by the agreements negotiated with state and local governments. Permanent employment can be relatively limited compared with the capital investment and construction activity associated with hyperscale facilities. Tax exemptions or other incentives can also reduce the public revenue otherwise collected from development. The appropriate question is therefore not simply whether a project "creates jobs" or "pays taxes," but how many jobs are created, who can access them, what tax revenue remains after incentives, and how those benefits compare with public costs and resource demands.[64]

Utility costs are another potential distributional concern. New generation, transmission, substations and other infrastructure required for large data-center loads create costs that utilities and regulators must allocate. Tariffs, cost-recovery agreements and other mechanisms can reduce the risk of transferring project-specific costs to existing customers.[65]

Proper cost allocation does not necessarily prevent broader increases in electricity prices. Virginia's Joint Legislative Audit and Review Commission found that data centers were paying their full cost of service under the rate structures examined, while projected growth in electricity demand was still likely to increase costs for other customers through additional infrastructure investment and higher energy prices. This Virginia finding illustrates a distinction between assigning a facility's costs appropriately and managing the wider effects of demand growth; it is not a forecast of Tampa Bay electricity rates.[64]

Procedural fairness matters as well. Communities cannot effectively evaluate a project when basic information about expected electricity demand, water use, tax incentives or infrastructure requirements is unavailable. TBRPC identifies confidentiality agreements as a potential barrier to accurate impact assessment and effective community-benefit negotiations and discusses CBAs as one mechanism for establishing measurable developer commitments.[56]

Environmental justice therefore does not require assuming that a data center is harmful simply because it is proposed near a lower-income or otherwise vulnerable community. Nor should economic investment automatically be treated as sufficient compensation for environmental impacts. The analysis requires examining both sides of the relationship.

For this report, the relevant questions are: Who receives the jobs and economic benefits? Who pays for infrastructure expansion? Which communities experience additional water, electricity, noise, heat or land-use pressures? What environmental burdens already exist

there? Were residents given meaningful information and participation before decisions were made? And are protections and benefits enforceable over the lifetime of the facility?

A data center can provide genuine economic benefits while also creating genuine environmental burdens. The environmental-justice question is whether those benefits and burdens are distributed fairly, particularly when communities with fewer resources or existing environmental pressures are being asked to accommodate additional infrastructure.

## 8. Case Studies: Where Data Center Water Became a Real Issue

National averages and modeled projections can identify the scale of data-center water demand, but they do not show how those demands are experienced by individual communities. Water impacts become concrete at the local level, where a facility interacts with a particular municipal supply, watershed, climate, wastewater system and regulatory environment.

The consequences have not been uniform. In some places, controversy has centered on the amount of municipal water consumed and whether residents were given adequate information about it. Elsewhere, the larger concern is the cumulative effect of many facilities developing within the same region. In water-stressed areas, the central questions involve scarcity, cooling technology and whether potable supplies should be used at all. Other communities have demonstrated how reclaimed water and purpose-built infrastructure can reduce those pressures.

The cases examined here were selected because they illustrate different dimensions of the problem rather than four versions of the same story. The Dalles, Oregon, demonstrates the importance of transparency and public accountability. Northern Virginia shows how individual facilities can become a regional infrastructure issue when development accumulates at enormous scale. Arizona demonstrates why water scarcity, climate and cooling technology need to be considered together. Quincy, Washington, provides a contrasting example of how engineering, treatment and water reuse can change the outcome.

These cases also caution against treating the presence of a data center as evidence of environmental harm by itself. The relevant questions are more specific: How much water was involved? Where did it come from? How significant was the demand relative to local conditions? What made the situation problematic or manageable? What changed in response?

Most importantly for this report, the case studies provide real-world experience that can inform decisions elsewhere. Tampa Bay does not share the same climate, water system or development history as any one of these locations. The purpose is therefore not to predict that Tampa Bay will reproduce their outcomes, but to identify which problems emerged, which solutions were attempted and which lessons can be applied before comparable development occurs locally.

## 8.1 The Dalles, Oregon: Water Demand and Public Transparency

Google opened its first company-owned data center in The Dalles, Oregon, in 2006 and has continued expanding its campus there. Over time, the facilities became a significant industrial user of the city's municipal water system. For years, however, the amount of municipal water supplied to Google's facilities was not publicly disclosed.

That lack of information became a public-records dispute in 2021. Oregonian/OregonLive reporter Mike Rogoway requested records showing the annual amount of water the city supplied to Design LLC, the entity associated with Google's facilities. The city denied the request, arguing that the information constituted a trade secret. After the Wasco County District Attorney determined that the city had not established that the records qualified for the exemption and ordered their disclosure, the City of The Dalles filed a lawsuit seeking to prevent their release.[35]

The case ended in a settlement in December 2022. Under the agreement, the city agreed to disclose aggregate annual meter readings associated with Google's utility accounts for 2012 through 2021 and comparable annual information in response to future public-records requests, subject to Oregon public-records law. The city also agreed to pay $53,173 for the newspaper's legal fees and costs.[37] Columbia Gorge News reported that Google would reimburse the city for that payment.[77]

The records revealed why the information was relevant to local water planning. The Oregonian reported that Google's facilities used approximately 355 million gallons of municipally supplied water in 2021, representing roughly 29 percent of The Dalles' total municipal water use that year.[78]

The approximately 355-million-gallon figure measures municipally supplied water, not consumptive loss. The settlement defined annual water usage as aggregate meter readings associated with Google's utility accounts.[37] Determining consumption also requires accounting for water discharged, as Google's environmental reporting does separately.[79]

Google’s 2025 Environmental Report lists 461.1 million gallons withdrawn at The Dalles in 2024, 99.7 million gallons discharged and 361.4 million gallons consumed. Consumption represented approximately 78 percent of reported withdrawal. Google calculates consumption as withdrawal minus discharge.[79]

The numbers also show that the issue did not end when the lawsuit did. Peer-reviewed research published in PLOS Water reports that Google's water demand in The Dalles increased approximately 342 percent between 2012 and 2024. In 2024, the 461.1-million-gallon withdrawal equaled approximately 1.26 million gallons per day, compared with citywide average demand of approximately 4.2 million gallons per day.[35]

A 2021 expansion agreement committed Google to fund $28.5 million in water-system improvements and transfer water rights associated with its industrial land to the city in exchange for additional municipal water supply. City officials and Google maintained that the improvements would accommodate growth and increase water availability for other users. The Oregonian also reported concerns from residents, WaterWatch and a Yakama Nation Fisheries representative about whether continued expansion could strain water supplies and affect fish and wildlife during prolonged drought.[78]

City records obtained by Oregon Public Broadcasting show Google-related municipal water use of approximately 383.9 million gallons in 2023 and 434.4 million gallons in 2024.[80] The 2024 municipal figure differs from Google’s reported site-level withdrawal of 461.1 million gallons, so the two should remain separately identified unless their reporting boundaries can be reconciled.[79, 80] Separately, The Dalles’ 2024 water master plan includes approximately 1 million gallons per day of projected demand from an unnamed industrial user. A former city public works director told OPB that this projected user was not Google.[80]

What made The Dalles particularly important as a case study, then, was not simply that a data center used a large quantity of water. A major industrial user had become a significant participant in the municipal water system while the public was unable to see the basic annual usage figures needed to evaluate that demand. The eventual disclosure did not prove that Google's water use was unsustainable; it made informed evaluation of that question possible.

The case also demonstrates that transparency and water sustainability are separate issues. A community may ultimately determine that a large industrial water demand is manageable. But residents, researchers and policymakers cannot independently evaluate scarcity, infrastructure requirements, conservation measures or future expansion when basic water-use information is unavailable.

For Tampa Bay, the case demonstrates why large industrial water commitments should be transparent before agreements are finalized. The public should be able to evaluate how much water is involved, where it comes from, how much is consumed, what infrastructure is required and how future growth fits within long-term water planning. Access to that information enables scrutiny of both developer commitments and government decisions.

## 8.2 Northern Virginia: When Data Centers Become a Regional Industry

Northern Virginia illustrates what happens when data-center development moves beyond individual projects and becomes a major regional industry. The area contains one of the world's largest concentrations of data centers, with development historically centered in Loudoun County and increasingly expanding into Prince William County and elsewhere in the Potomac River basin.[33]

This concentration changes the water question. Rather than asking how much one facility consumes, regional planners must consider the combined demand of many existing facilities and more than 100 million square feet of additional data-center development currently planned across the region.[33]

The Potomac River makes that cumulative demand particularly important. It supplies approximately 75 percent of the water used by the Washington metropolitan area's three primary water utilities and is the sole drinking-water source for Washington, D.C., and Arlington County. Together, the major regional systems relying on the river serve approximately five million people.[33]

Current data-center water demand remains relatively manageable at the regional level. The Interstate Commission on the Potomac River Basin estimated that data centers within the Washington Metropolitan Area consumed approximately 4 million gallons per day on average in 2025, with peak-day consumption of approximately 15 million gallons per day. Upstream facilities consumed less than 0.1 million gallons per day on average and approximately 0.3 million gallons per day at peak.[33]

Those figures also demonstrate why percentages require context. Within the Washington Metropolitan Area, data centers currently account for only about 1 percent of total water withdrawals, but approximately 9 percent of annual consumptive use and as much as 12 percent of consumptive use during summer. Across the broader Potomac basin, data centers represent approximately 0.3 percent of withdrawals and 3 percent of consumptive use.[33]

The difference reflects the distinction established earlier in this report: withdrawal measures water taken from a source, while consumption describes the portion not promptly returned to that source. Data-center cooling can therefore represent a relatively small share of regional withdrawals while having a larger share of consumptive demand.

Seasonality creates another important difference. ICPRB found that monthly data-center water use during summer can approach three times average annual demand, while peak daily use can reach as much as ten times average demand. That increase occurs when outdoor water demand is also elevated and Potomac River flows may be lower.[33]

Cooling technology partly explains both the region's current demand and uncertainty about its future. Approximately 40 percent of existing facilities in the basin can rely exclusively on air cooling, according to ICPRB, while many others use hybrid systems that can take advantage of cooler weather during part of the year. Future hyperscale facilities with denser computing equipment may use different cooling systems, however, and the balance between air and water cooling will significantly affect future demand.[33]

Under ICPRB's baseline assumptions, average data-center water consumption within the Washington Metropolitan Area could increase from approximately 4 million gallons per day in 2025 to about 22 million gallons per day by 2050. Peak demand could rise from approximately 15 million gallons per day to more than 80 million gallons per day. Upstream consumptive demand is projected to reach approximately 5 million gallons per day on average and 17 million gallons per day at peak.[33] These are PROJECTED, not current, quantities.

ICPRB identifies potential water-supply reliability challenges if projected development materializes, particularly during low-flow conditions. The severity of those challenges depends on cooling technologies, development rates, efficiency improvements and water-management policies.[33]

Electricity planning illustrates the cumulative demands of data-center expansion. In its 2025 planning work, PJM identified data centers as the primary driver of projected electricity-demand growth across its service territory and estimated up to approximately 30 gigawatts of additional data-center demand between 2025 and 2030. This was a planning projection, rather than measured demand already added to the grid.[38]

Within Dominion Energy Virginia's territory, data-center demand was already measured in gigawatts rather than megawatts. Dominion projected approximately 4.2 gigawatts of data-center billing demand for 2025, while continuing expansion required additional transmission infrastructure.[38]

The consequences extend beyond the facilities themselves. PJM's regional planning has identified major transmission expansion needs driven heavily by rapidly increasing electricity demand, with particularly strong concentration in Northern Virginia. A 2025 transmission-planning window produced a proposed package of backbone upgrades estimated at approximately $11.6 billion, intended in part to move additional generation into Northern Virginia's growing load centers.[38]

Electricity growth also matters to water because, as established in Section 7.6, the water footprint of a data center does not necessarily stop at its cooling system. Electricity generation can consume water elsewhere in the grid. Northern Virginia therefore demonstrates both sides of the water-energy relationship: direct cooling demand occurs within the Potomac water system, while some additional water associated with supplying electricity may occur elsewhere. The two should not be combined without compatible accounting boundaries.

Concentrated data-center demand can also create operational challenges for the electricity grid. On July 22, 2026, nearly 4,000 megawatts of data-center load unexpectedly disconnected from the grid in Northern Virginia and switched to backup generation. PJM and Dominion operators responded to imbalances between generation and demand and spikes in voltage and frequency. PJM reported that operators managed the event without unresolvable reliability impacts, but identified it as the third measurable event of its kind in two years. The incident helped prompt proposed requirements for large computing facilities to remain connected during certain transmission-system disturbances.[81]

That incident is not a water event, but it illustrates why Northern Virginia is useful as a cumulative-development case. Once enough facilities are concentrated in one region, they stop functioning only as individual industrial customers. Their combined demand becomes a regional infrastructure-planning issue.

The Potomac analysis also exposes a regulatory complication. Data centers in the basin are generally supplied by public utilities rather than holding their own direct water-withdrawal permits. ICPRB notes that this can leave large utility-supplied facilities outside existing consumptive-use mitigation requirements that would apply to certain direct withdrawals. The commission has recommended examining policies addressing cumulative demand during low-flow and drought conditions and improving reporting of water use, energy demand and cooling technology.[33]

Northern Virginia therefore does not provide a simple story of data centers exhausting a regional water supply. Current average water demand remains comparatively modest. The concern comes from the trajectory and concentration of development: rapidly expanding

computing capacity, large projected increases in electricity demand, more than 100 million square feet of planned data-center construction, sharply higher summer water demand and a shared river system that must remain reliable during drought and low-flow periods.

For Tampa Bay, the lesson is particularly relevant. The question is not only whether a utility can accommodate one proposed data center. Regional planning must consider the cumulative demand of facilities already operating, under construction, approved and reasonably expected to follow. Water supply, peak demand, wastewater capacity, electricity infrastructure and drought resilience should be evaluated at that regional scale before individually manageable projects collectively become a much larger infrastructure problem.

## 8.3 Arizona: Data Centers in a Water-Stressed Climate

Arizona provides a useful test of one of this report's central arguments: the environmental significance of data-center water use cannot be separated from location. A cooling strategy that is reasonable in a cool, water-abundant region may be a poor choice in a hot, arid region where water supplies are already under long-term stress.

The state's water situation is serious, but describing Arizona simply as “running out of water” would be inaccurate. Arizona obtains water from multiple sources, including the Colorado River, the Salt and Verde river systems and groundwater, and individual cities have different supply portfolios and water rights. Phoenix, for example, reports that nearly 60 percent of its supply comes from the Salt and Verde rivers, approximately 40 percent from the Colorado River, and only a small portion from groundwater.[39]

At the same time, Arizona's Department of Water Resources identifies persistent long-term drought as an ongoing concern. The Colorado River, which provides roughly 40 percent of Arizona's overall water supply, has experienced severe drought conditions since 2000. In July 2026, the Colorado River system was approximately 33 percent full, with Lake Mead at approximately 28 percent and Lake Powell at approximately 24 percent. Arizona also remained under both short- and long-term drought conditions.[39]

This context does not mean that every gallon consumed by an Arizona data center deprives a household of drinking water. It does mean that water-intensive industrial development in Arizona deserves a different level of scrutiny than the same demand might receive in a wetter region.

The data-center industry's response also demonstrates that cooling technology is not fixed.

Microsoft's original Arizona data-center design provides one example. The company designed its facilities to use outside air without cooling water when temperatures remained below 85°F (29.4°C). Above that temperature, the facilities could use direct evaporative cooling. Microsoft stated that this approach used substantially less water than conventional water-cooled cooling towers while also benefiting from the lower electricity requirements of evaporative cooling.[40]

Arizona's climate, however, limits how often that water-free operating mode can be used. Microsoft now reports that its Phoenix facilities may require water for direct evaporative cooling as much as 40 percent of the year, compared with approximately 10 percent in Virginia and less than 5 percent in some cooler European locations.[40] The same cooling architecture therefore produces different water requirements simply because it is deployed in a different climate.

Microsoft's Arizona facilities purchase water from EPCOR and the City of Goodyear. When evaporative cooling is operating, the company reports that water cycles through the cooling system approximately two to five times; some evaporates and the remainder is generally discharged to the local wastewater-treatment system.[40]

The evidence available here does not establish that Microsoft's Arizona cooling supply is broadly reclaimed water. Microsoft's current Arizona fact sheet identifies purchases from local utilities but does not characterize the supplied cooling water as reclaimed. That distinction matters: this report should not infer a water source simply because reclaimed water is technically possible.

More importantly, Microsoft's design is now changing. In 2024, the company announced a new data-center architecture using closed-loop chip-level liquid cooling and air-cooled chillers that requires zero water evaporation for cooling during normal operations after the cooling loop is initially filled. Microsoft estimates that the design can avoid more than 125 million liters, approximately 33 million gallons, of cooling water annually per data center compared with its evaporative designs.[41]

Phoenix was selected as one of the first locations for this architecture. Microsoft announced that new projects in Phoenix and Mount Pleasant, Wisconsin, would pilot the zero-water-evaporation design beginning in 2026. Its current Arizona materials also state that future Arizona facilities will use air-cooled chillers rather than the direct evaporative cooling used by its existing facilities.[41]

Google made an even more explicit siting decision in nearby Mesa. Before developing its Mesa data center, Google conducted a watershed-risk assessment and concluded that the proposed local water source faced high risk of depletion and scarcity. According to Google,

that finding directly affected the facility design: the company chose to air-cool the Mesa data center to minimize its impact on the local watershed.[42]

That example is important because it provides something stronger than a general corporate promise to conserve water. It documents a case in which local water conditions influenced the cooling technology selected for a specific data center.

Meta's Mesa development provides another variation. Meta reports that its cooling technologies are selected according to local climatic and environmental conditions and that its newer AI-optimized architecture uses closed-loop liquid cooling with dry coolers. The company has also funded Arizona watershed-restoration projects and previously reported that its Mesa facility was designed to be approximately 60 percent more water-efficient than an average data center under its comparison methodology.[43]

These examples also expose the water-energy tradeoff discussed earlier in this report. Evaporating water is an effective method of rejecting heat and can reduce the electricity required for mechanical cooling. Eliminating that evaporation can therefore increase electricity requirements.

Microsoft acknowledges this directly. Its zero-water cooling design replaces evaporative heat rejection with mechanical cooling, which increases Power Usage Effectiveness relative to its evaporative designs. The company says improvements in chip-level cooling and higher-temperature chillers limit the increase to what it describes as a nominal amount across its fleet.[41]

That tradeoff prevents a simplistic conclusion that the cooling system consuming the least water is automatically environmentally superior. In a water-stressed location, accepting somewhat higher electricity consumption may be reasonable to protect scarce water supplies. In another location, efficient evaporative cooling supplied by abundant or reclaimed water could potentially produce a lower overall environmental burden. The electricity source also matters because additional electricity can carry its own indirect water footprint.

Arizona therefore demonstrates both the risk and adaptability of data-center development.

The risk is real: large computing facilities are being constructed in a hot region facing persistent drought and substantial uncertainty surrounding one of its major water sources. Increasing industrial demand must therefore be evaluated alongside municipal, agricultural, ecological and other existing demands.[39]

But the evidence also shows that high data-center water consumption is not technologically inevitable. Google chose air cooling after identifying local water scarcity.

Microsoft moved from climate-dependent evaporative cooling toward closed-loop systems designed to eliminate routine cooling-water evaporation. These decisions demonstrate that facility design can respond to watershed conditions rather than treating water consumption as an unavoidable consequence of computing.

Arizona consequently provides a more useful lesson than either “data centers are stealing desert water” or “technology companies have solved the problem.” Water scarcity is genuine, different facilities make different choices, and reducing direct water consumption can create energy tradeoffs that must also be evaluated.

For Tampa Bay, the lesson is that cooling and siting decisions should reflect the environmental conditions of the place where the facility will actually operate. Climate, water source, drought vulnerability, reclaimed-water availability, electricity supply and seasonal peak demand should influence facility design before construction. A cooling system should not be considered sustainable merely because it performs well somewhere else.

## 8.4 Quincy, Washington: Can Data Center Water Be Reused?

Quincy, Washington, provides a useful counterpoint to cases where data-center water demand primarily became a source of conflict. In Quincy, substantial data-center water use created genuine water-supply and wastewater-management problems, but those problems eventually produced a large-scale reuse system designed specifically around the industry's cooling water. The result demonstrates that data-center water can be recycled, but doing so requires substantial infrastructure, treatment, financing and management of the waste products that remain.[26]

Quincy is located in the arid portion of eastern Washington and receives only about 7.8 inches of precipitation annually. Five groundwater wells provide the city's potable supply. Although Quincy has only about 8,100 residents, its large industrial sector gives it water demand comparable to a community of roughly 30,000 people. Before the reuse project, the city pumped approximately 2.1 billion gallons of groundwater per year. About 57 percent went to food processors, 33 percent to residential and other public or commercial uses, and approximately 10 percent to data centers for cooling.[26]

Microsoft has operated a large data-center campus in Quincy since 2007. Historically, the facility received potable groundwater from the city for cooling. The water typically circulated through the cooling system four or five times before being discharged as blowdown to the city's wastewater system.[26]

That arrangement created problems on both sides of the water cycle.

Quincy's groundwater naturally contains relatively high concentrations of dissolved minerals including calcium, magnesium and silica. As cooling water evaporated, those minerals became increasingly concentrated. Allowing the water to circulate indefinitely could cause scaling and impair heat transfer, so increasingly concentrated water had to be removed from the cooling system as blowdown.[26]

Sending that wastewater to Quincy's municipal wastewater reclamation facility created another problem. The facility was designed primarily for municipal wastewater rather than industrial cooling water containing high concentrations of total dissolved solids. The city's monitoring found total dissolved solids in influent and effluent consistently above the Washington Department of Ecology's 500-milligram-per-liter groundwater guideline during the period when substantial data-center wastewater was entering the system.[26]

That mattered because Quincy reused treated municipal wastewater for irrigation and groundwater recharge. Mineral accumulation created visible deposits on groundwater-recharge basins and complicated compliance with requirements intended to protect groundwater that could eventually become drinking water.[26]

The problem was therefore larger than simply “data centers use too much water.” Quincy faced three connected challenges: potable groundwater was being used for an industrial process that did not require drinking-water quality; the groundwater's mineral content made it less than ideal for cooling; and the resulting mineral-concentrated blowdown was difficult for the existing municipal wastewater system to manage.

Microsoft and the City of Quincy responded by developing the Quincy Water Reuse Utility (QWRU). After more than ten years of planning, construction and testing, the system became operational on June 30, 2021. The approximately $31 million project consists of ten interconnected treatment systems and more than 30 miles of piping across Quincy's industrial area.[26]

Instead of treating cooling water as a single-use supply, the QWRU creates a loop. After water cycles through Microsoft's cooling equipment four or five times, the resulting blowdown enters a series of treatment processes that can include lime softening, ultrafiltration, high-efficiency softening and reverse osmosis. Treated water is then blended to achieve the required water quality and returned to the data center for cooling again.[26]

The system does not make water consumption disappear. Some water still evaporates during cooling, and additional water must periodically enter the loop to replace those losses. What changes is the source and repeated use of the remaining water.

Quincy obtains approximately 260 million gallons per year of Columbia Basin Project water from an irrigation canal for make-up water, primarily during the warmer months when cooling demand is highest. According to EPA's case study, only about 5 percent of annual make-up water remained potable groundwater under the system described. The city has sought additional access to canal water with the goal of potentially eliminating routine potable-groundwater dependence for data-center cooling.[26]

EPA estimates that the reuse system offsets approximately 138 million gallons of potable-water demand per year, or about 522 million liters.[26] Microsoft's more recent reporting states that the Quincy reuse system has reduced its potable-water use there by approximately 97 percent on average.[43] These figures should be presented separately rather than combined into a single efficiency calculation because they use different metrics. EPA reports an estimated annual demand offset, while Microsoft reports a percentage reduction in its potable-water use.

Water reuse also did not make the dissolved minerals disappear. Treatment concentrates calcium, magnesium, silica and other salts into brine streams. These are transferred to lined evaporation ponds, where water evaporates and the remaining solids are ultimately removed and disposed of in a manner intended to prevent their entry into groundwater or surface water. EPA reports that Quincy expects accumulated residuals to require removal approximately every two or three years.[26]

That distinction is important. Recycling water transfers some of the environmental-management challenge from water quantity to treatment and residual management. The system substantially reduces demand for potable groundwater and prevents high-TDS industrial wastewater from entering the municipal reclamation facility, but it requires energy, chemical treatment, specialized equipment and appropriate disposal of the concentrated material removed from the water.

The project also demonstrates the scale of infrastructure required for sophisticated reuse. Microsoft and Quincy entered a 30-year agreement under which the city owns and operates the QWRU while Microsoft financed the approximately $31 million capital investment and pays its operating costs. Other data centers can potentially connect, but they must contribute connection costs, required upgrades and their share of continuing operational expenses.[26]

This financing arrangement matters from a community perspective. The infrastructure required to accommodate the industrial wastewater was not simply transferred to ordinary municipal ratepayers. The industrial user generating the demand financed the specialized system needed to manage it.

The solution nevertheless has limits. As of EPA's case-study period, only Microsoft's campus was connected to the QWRU. Other Quincy data centers located farther away continued using potable groundwater because extending the reuse network to them would be expensive.[26] The existence of a successful reuse project therefore does not mean that all data-center cooling in Quincy has become circular.

The system also retains potable groundwater as a backup source. During a hot and dry period in 2021, pumps supplying canal water to the reuse system were shut off. QWRU was able to switch temporarily to potable groundwater. That redundancy demonstrates another important principle: water resilience can involve diversifying supplies rather than permanently eliminating every conventional source.[26]

Quincy therefore answers the question posed by this case study with a qualified yes. Data-center cooling water can be treated and reused at substantial scale. The QWRU reduces reliance on potable groundwater, separates difficult industrial wastewater from the municipal treatment stream and allows cooling water to circulate repeatedly rather than being discarded after several cycles.[26]

But Quincy also shows why “just use reclaimed water” is not a complete solution. Water chemistry matters. Wastewater still requires treatment. Salts and contaminants removed from reused water still require disposal. Replacement water is still needed for evaporation. Pipelines and treatment facilities cost money. And facilities located too far from reuse infrastructure may not be able to participate economically.

For Tampa Bay, the lesson is particularly useful. Alternative water supplies and industrial reuse can reduce competition for drinking-quality water, but these systems are most effective when incorporated into regional infrastructure planning rather than added after facilities have already been built. A Tampa Bay data-center strategy should therefore ask not only whether reclaimed or other non-potable water exists, but where it exists, whether sufficient transmission infrastructure connects it to proposed facilities, what treatment the cooling system requires, what happens to concentrated residuals, who pays for the necessary infrastructure, and what backup supply would be used during interruptions.

Quincy's experience suggests that water efficiency should be evaluated at the system level. It should be to design a system in which the quality of water supplied matches the quality actually required, usable water is circulated as many times as practical, potable supplies are preserved where alternatives are feasible, industrial wastewater is treated appropriately, and the industry creating the infrastructure demand bears the cost of managing it.

### 8.5 What the Case Studies Show

Across the four case studies, data-center water impacts are highly dependent on local conditions and infrastructure choices. The Dalles illustrates the importance of public disclosure; Northern Virginia shows why cumulative development must be evaluated regionally; Arizona demonstrates how water scarcity can influence cooling design; and Quincy shows how treatment and reuse infrastructure can reduce potable-water demand while creating its own costs and management requirements.[26, 33, 35, 37, 39-43]

None of these cases establishes a universal outcome for data-center development. A facility that creates manageable demand in one location may create greater pressure in another, while changes in cooling technology, water source, electricity supply, wastewater management and regional planning can materially change the result. The cases therefore provide examples of conditions and responses, not templates that can be transferred unchanged to Tampa Bay.

## 9. Are AI Companies Making It Better?

The water footprint of AI is not fixed. The amount of water associated with computing can change as hardware becomes more efficient, models require less computation, cooling systems improve, electricity generation changes and companies adopt alternative water sources or conservation strategies.

Major technology companies have announced efforts to reduce those impacts through more efficient computing, improved data-center design, water-replenishment programs, reclaimed-water projects and other resource-management strategies. These efforts matter because rapid growth in AI infrastructure could otherwise offset improvements achieved at the level of individual models or facilities.

At the same time, corporate sustainability commitments require careful interpretation. Improvements in efficiency do not necessarily mean that total water use is declining, and replenishment commitments are not equivalent to eliminating water consumption at the facilities where it occurs. Company-wide environmental targets can also obscure substantial differences among individual data centers and regions.

This section therefore examines what AI and data-center companies are actually doing to reduce their water footprint, how much progress can be established from available evidence, and where the limitations of those strategies remain. The relevant question is not simply whether companies are “making it better,” but whether improvements in computing

efficiency, cooling, water sourcing and stewardship are occurring quickly enough to meaningfully change the industry's growing resource demands.

## 9.1 Computing Efficiency

There is strong evidence that AI computation is becoming more efficient. Improvements are occurring simultaneously in hardware, model architecture and inference-serving systems, allowing newer systems to perform a given amount of computation with less electricity than earlier approaches.[11, 13]

The rate of improvement has been substantial. The International Energy Agency reported in 2026 that the electricity required for an individual AI task had fallen by at least an order of magnitude annually in recent years, attributing the decline to advances in both hardware and software.[13] The IEA describes the recent rate of task-level AI efficiency improvement as unusually rapid by historical energy-efficiency standards.

Hardware is one contributor. AI workloads increasingly run on specialized accelerators such as GPUs and TPUs rather than relying entirely on general-purpose processors. New generations of accelerators can perform more AI computation for each unit of electricity, while techniques such as lower-precision arithmetic allow some workloads to perform calculations using fewer bits without requiring equivalent reductions in useful model performance.[13]

Efficiency can also improve at the model level. Larger models are not always necessary for every task. Techniques including quantization, pruning, knowledge distillation and model compression can reduce the computation or memory required during inference. A 2026 systematic review of Green AI techniques found evidence of substantial efficiency improvements from these approaches, including reductions in model size and energy requirements under the applications examined.[44]

Model architecture matters as well. Some systems can route requests toward smaller or specialized models when a larger model is unnecessary, while mixture-of-experts architectures can activate only portions of a model for a particular computation rather than using every parameter for every request. These techniques do not eliminate computation, but they can reduce the amount required to produce a useful result.

A third source of improvement occurs during inference serving. Production AI systems do not necessarily process every user request independently on an otherwise idle accelerator. Operators can batch requests, improve accelerator utilization, optimize memory

management, cache reusable computations and increase the number of tokens produced for a given amount of hardware and electricity.

The 2026 Joule analysis examined these improvements together and estimated that advances in models, serving systems and hardware could collectively reduce inference energy per comparable query by approximately 8 to 20 times relative to its baseline assumptions.[11] The study also found that optimized frontier-scale inference can already require substantially less electricity than estimates based on poorly utilized or non-production hardware.

These improvements are important for water because electricity and water footprints are connected. Reducing the electricity required for a comparable workload can reduce heat that must be removed from the servers and can reduce electricity-related water consumption. Greater computational efficiency can therefore lower both energy and some components of water demand, although the resulting water savings depend on cooling technology and electricity source.

These improvements provide strong evidence that the resource intensity of comparable AI computation can continue to decline. They do not, by themselves, establish that the industry's total water footprint will decline as deployment expands. That separate question is examined in Section 10.

## 9.2 Cooling Innovation

Cooling technology is becoming one of the most important opportunities for reducing the water footprint of AI infrastructure. As AI accelerators become more powerful, they also concentrate more heat into individual servers and racks, making traditional room-level air cooling increasingly difficult for the highest-density systems. This challenge is accelerating development of direct-to-chip liquid cooling, closed-loop systems, dry heat rejection, hybrid cooling and other approaches designed to remove heat more efficiently while reducing or eliminating evaporative water consumption.[41, 45]

Cooling innovation provides another pathway for reducing direct water consumption. As AI hardware increases rack power density, developers are expanding direct-to-chip liquid cooling, closed-loop systems, dry heat rejection and hybrid approaches capable of reducing or eliminating routine evaporative cooling.[41, 45]

Microsoft's newer data-center architecture illustrates this distinction. Its design circulates liquid directly to computing equipment in a closed loop and uses mechanical heat rejection rather than evaporative cooling. Once initially filled, the loop continuously

recirculates its cooling water. Microsoft estimates that the design can avoid more than 125 million liters (approximately 33 million gallons) of cooling water annually per data center compared with its previous evaporative designs.[41] This is a company-reported design estimate, not an independently measured industry-wide result.

The transition is already moving beyond experimental systems. Microsoft reports that all of its new data-center designs introduced since August 2024 use its next-generation zero-water-evaporation cooling architecture, with Phoenix and Mount Pleasant among its initial deployments. The company reported in 2026 that approximately 90 percent of its existing owned fleet operates with low- to zero-water cooling systems, although that fleetwide figure is also company-reported and encompasses several different cooling technologies.[41]

Google is developing another approach for bringing liquid cooling into facilities originally designed around air cooling. Its 2026 Brazos system places a closed-loop liquid-to-air cooling unit at the rack, allowing high-density liquid-cooled AI equipment to operate inside existing air-cooled data centers without requiring the entire building to be converted to a facility-wide chilled-water loop. Google developed the system in response to newer AI and high-performance-computing chips exceeding 1,000 watts of thermal design power.[45]

Cooling innovation is also occurring outside individual technology companies. The U.S. Department of Energy's ARPA-E COOLERCHIPS program is supporting development of advanced thermal-management technologies for high-power data centers. Its 2026 second phase is testing systems intended to handle AI heat loads reaching 1 megawatt per rack while using less energy and no water for cooling. These are research and development targets rather than demonstrated performance across commercial data centers, but they show how rapidly cooling requirements are changing as AI hardware becomes denser.[45]

The water-energy tradeoff remains. Microsoft acknowledges that its zero-water cooling architecture increases power usage effectiveness compared with its evaporative designs, although the company says warmer operating temperatures and more efficient chillers limit the increase.[41] “Zero water for cooling” should therefore not be interpreted as “zero environmental impact.” Mechanical cooling still requires electricity, and the electricity supplying the system can carry its own indirect water footprint.

Nevertheless, cooling represents an area where technological change can directly address one of the environmental impacts identified earlier in this report. High direct water consumption is not an unavoidable requirement of AI computing. Facilities can already be designed to operate with little or no routine evaporative cooling, and emerging systems are being developed specifically for the much greater heat densities expected from future AI hardware.

## 9.3 Reclaimed and Non-Potable Water

Alternative water sources can improve conventional cooling systems as well. Reclaimed wastewater, recycled process water and captured rainwater can reduce dependence on potable supplies even where evaporative cooling remains in use. The Quincy case demonstrates how reuse can substantially reduce potable-water demand, although treatment, pipelines and residual management remain necessary.[26]

Cooling innovation therefore involves more than searching for a single technology that uses the least water. The appropriate system depends on chip density, climate, water availability, electricity supply, existing infrastructure and the temperature at which heat can be rejected. Google's decision to air-cool its Mesa facility after identifying high local water-scarcity risk demonstrates how those factors can influence an actual design decision.[42]

Reducing data-center demand for potable water does not always require eliminating water-based cooling. Another strategy is to match the quality of the water supply to the quality actually required for the task. Cooling equipment generally does not require water treated to drinking-water standards, creating opportunities in some locations to use reclaimed wastewater, recycled industrial water and other non-potable supplies instead.[2, 26]

This distinction matters because potable water represents more than raw water. Municipal drinking water has already undergone treatment and is maintained to standards intended to protect human health. Where a data center can safely operate on a lower-quality alternative, using that alternative can preserve drinking-quality supplies for households and other uses that actually require them. The environmental benefit can be particularly important in water-stressed regions or where a large facility would otherwise become a major municipal potable-water customer.

Reclaimed water is typically wastewater that has been treated to a quality appropriate for a subsequent beneficial use. Depending on local regulations and treatment levels, reclaimed water can be used for irrigation, industrial processes, cooling and other non-potable applications. Using it for data-center cooling can reduce demand on freshwater supplies while giving already-treated wastewater another productive use before it is discharged or otherwise managed.[46]

The approach is already being used commercially. Google's data center in Douglas County, Georgia, has used recycled wastewater for cooling through a partnership with the Douglasville-Douglas County Water and Sewer Authority. Rather than relying entirely on

drinking water, treated wastewater that would otherwise be discharged is further treated and supplied to the data center for cooling. Google has described the system as an example of selecting alternative water sources according to local conditions.[47]

The Quincy, Washington, case examined in Section 8.4 demonstrates a more elaborate version of the same principle. Quincy's system treats data-center cooling blowdown and returns treated water to Microsoft's facility while supplementing the system primarily with non-potable Columbia Basin Project water. EPA estimates that the resulting reuse system offsets approximately 138 million gallons of potable-water demand annually.[26]

These examples demonstrate an important distinction between water use and potable-water use. A facility can continue using substantial quantities of water for cooling while greatly reducing its dependence on drinking-quality supplies. Consequently, a claim that a facility has reduced its “potable water use” should not automatically be interpreted as an equivalent reduction in its total water consumption.

Reclaimed water also should not automatically be treated as environmentally free or unlimited. Wastewater may already have other beneficial uses, including irrigation, industrial supply or environmental flows. Diverting it to a data center can therefore create competing demands of its own. The environmental value of reuse depends on what would otherwise have happened to that water and whether sufficient supply exists after other needs are considered.

Water chemistry presents another constraint. Reclaimed and other non-potable sources can contain salts, minerals, nutrients, microorganisms or other constituents at concentrations unsuitable for cooling equipment without additional treatment. Higher concentrations can increase scaling, corrosion or biological growth and may require filtration, disinfection, softening, reverse osmosis or other treatment.[26] The resulting treatment can consume energy and produce concentrated residual streams that still require management.

Infrastructure can be just as important as water availability. A community may produce large quantities of reclaimed water without having pipelines capable of delivering it to an industrial site. Building separate non-potable distribution systems, storage, pumping and treatment facilities can require substantial capital investment. Quincy's approximately $31 million reuse project and more than 30 miles of piping demonstrate how extensive that infrastructure can become.[26]

Some communities are beginning to incorporate alternative water supplies directly into data-center development rather than retrofitting them later. Meta reports using reclaimed water at several facilities and evaluating water source, quality and availability when

selecting cooling strategies. Google similarly reports evaluating alternatives including reclaimed wastewater where local conditions make them appropriate.[47, 48] These are company-reported practices, however, and facility-level claims should be evaluated using actual withdrawal, consumption and water-source data where available.

Alternative supplies can also improve drought resilience, but they do not eliminate the need for contingency planning. Reclaimed-water production itself can fluctuate, infrastructure can fail and treatment systems require maintenance. Quincy's reuse system, for example, retained potable groundwater as a backup and temporarily relied on it when canal-water pumps were unavailable during hot, dry conditions.[26]

A more useful measure is therefore not simply whether a data center can advertise that it "uses reclaimed water." A stronger assessment asks what percentage of its supply is potable versus non-potable, how much water is actually consumed, where the alternative water originates, what treatment it requires, what competing uses exist, what happens to the resulting wastewater or residuals, and what supply is used when the alternative system is unavailable.

Reclaimed and non-potable water can therefore be an important part of reducing the industry's pressure on drinking-water supplies, particularly where evaporative cooling remains appropriate. But it is a water-management strategy rather than a universal solution. In some locations, closed-loop or dry cooling may be preferable. In others, abundant reclaimed water and existing distribution infrastructure may make water-based cooling comparatively reasonable.

The larger lesson is that data centers do not necessarily require potable water for every cooling application. Where suitable non-potable supplies exist, matching water quality to its actual use can preserve potable resources while allowing water-efficient cooling technologies to remain available. The strongest projects are those in which alternative supply, treatment, distribution, wastewater management and drought contingencies are planned together rather than treating reclaimed water as an unlimited substitute for freshwater.

## 9.4 Better Siting

Some environmental impacts of data centers cannot be solved through efficiency alone. A highly efficient facility can still create significant local pressure if it is built where water is scarce, electricity infrastructure is constrained or numerous other facilities already compete for the same resources. Choosing an appropriate location is therefore itself an environmental-management strategy.[22, 23]

Water availability is one important siting consideration, and annual water-supply totals alone may not capture local risk. Groundwater recharge, surface-water availability, drought vulnerability, seasonal demand, existing users, projected population growth and climate-related changes in future supply can all affect whether a proposed demand is sustainable. DOE water-management guidance similarly recommends evaluating future water-availability risks and planning for drought and other shortages.

Climate also affects which cooling technologies make sense. DOE's data-center design guidance emphasizes that there is no universally optimal design because environmental conditions differ by location. Cooler climates can provide more opportunities for outside-air or water-side economization, while hot climates can require substantially different cooling strategies.[45]

The Arizona case demonstrates how this principle can operate in practice. Google reported that its watershed assessment for Mesa identified high depletion and scarcity risk and consequently chose air cooling for the facility.[42] Microsoft is similarly deploying cooling architecture in Phoenix designed to eliminate routine evaporative water consumption.[41] These examples show that siting analysis and cooling design do not need to be separate decisions: local water conditions can influence the technology ultimately installed.

Available water sources can also affect siting. A location with access to adequate reclaimed-water infrastructure may be better suited to water-based cooling than one requiring additional potable-water withdrawals. The existence of a wastewater-treatment plant alone does not establish that reclaimed water is available, however; competing uses, treatment requirements and distribution infrastructure also matter.

Electricity creates another siting constraint. AI facilities require large, continuous electrical loads, and concentrating facilities in one region can require additional generation and transmission infrastructure. The electricity sources serving a location can also affect the facility's indirect water footprint. Water and electricity availability are therefore interconnected siting considerations.

Cumulative development is particularly important. Northern Virginia demonstrates that a location capable of accommodating early data-center projects can eventually face a different planning problem as dozens of facilities concentrate around the same infrastructure.[33] The suitability of a location can therefore change as additional facilities are developed, making existing facilities, projects under construction and reasonably foreseeable development relevant to siting analysis.[33]

Wastewater capacity can constrain a site even when adequate incoming water is available. Cooling blowdown or other industrial wastewater may require sufficient treatment capacity, and local water chemistry can affect those requirements.[26]

Local heat and surrounding land use are emerging considerations as well. The measured thermal effects discussed in Section 7.8 suggest that locating very large facilities immediately adjacent to heat-vulnerable populations may warrant additional evaluation, particularly in already hot urban environments. The evidence is not yet sufficient to establish a universal setback distance, but it identifies localized heat as another potential siting consideration, particularly in already hot urban environments.[30]

Community conditions can also affect the significance of a proposed location. Existing environmental burdens, infrastructure capacity and the distribution of a project's economic benefits and costs can differ substantially among potential host communities.[34-36] Development in a lower-income community is not automatically unjust, but existing resource or environmental pressures and opportunities for meaningful public participation can be relevant to evaluating a site.

Better siting can also reduce infrastructure requirements rather than merely environmental impacts. Existing transmission capacity, appropriate industrial land, sufficient water and wastewater infrastructure, fiber connectivity and proximity to suitable generation can reduce the need for expensive new supporting systems. DOE's current data-center siting work similarly considers the availability of land and existing energy infrastructure when identifying potential development locations.

No location will optimize every variable simultaneously. A site with abundant water may have a carbon-intensive electrical grid. A dry site may have excellent solar resources but require water-saving cooling technologies. A location with reclaimed water may require substantial new transmission infrastructure, while an electrically advantageous location may have limited wastewater capacity.

For that reason, better siting does not mean finding a place where a data center has no environmental impact. It means choosing locations where the required resources can be supplied with the least avoidable harm and then designing the facility around the actual conditions of that location.

## 9.5 Cleaner / Lower-Water Electricity

As established earlier in this report, a substantial portion of the water associated with data-center operation can occur through electricity generation. Reducing the water intensity of

the electricity supplying a facility can therefore reduce its indirect water footprint even when nothing changes inside the data center itself.

Electricity technologies differ substantially in their water requirements. Thermoelectric power plants, including nuclear, coal and natural-gas facilities, can require water for cooling, although the quantities withdrawn and consumed vary greatly according to plant and cooling technology. Once-through systems can withdraw very large quantities while returning much of the water, whereas recirculating systems generally withdraw less but consume a greater proportion through evaporation.[48]

Wind and solar photovoltaic generation generally have very low operational water requirements because they do not require conventional steam-cycle cooling. NREL's lifecycle research finds that wind and solar PV are typically among the lowest-water electricity technologies, although neither has literally zero lifecycle water use: manufacturing, construction and maintenance still require water.[48]

Hydroelectricity requires additional caution. Although hydropower does not consume cooling water in the same way as a thermoelectric plant, reservoirs can lose substantial quantities of water through evaporation. How that evaporation should be attributed to electricity becomes complicated when reservoirs simultaneously provide drinking water, irrigation, flood control, recreation and other services. For this reason, a simple statement that all renewable electricity is automatically “low-water” would be inaccurate.

Nuclear energy presents another important distinction between low carbon and low water. Nuclear generation has very low operational greenhouse-gas emissions but commonly relies on thermal cooling systems that can withdraw or consume significant quantities of water. NREL's comparative research places nuclear among the higher-water electricity technologies, particularly where water-intensive cooling systems are used.[48]

This reinforces a principle established earlier in the report: low carbon does not automatically mean low water, and low water does not automatically mean low overall environmental impact. Electricity sources need to be evaluated across multiple dimensions rather than placed into a single category of “clean.”

The electricity currently supplying data centers remains mixed. The IEA estimates that globally, coal supplies approximately 30 percent of electricity physically consumed by data centers, renewables approximately 27 percent, natural gas 26 percent and nuclear 15 percent. In the United States, natural gas supplies more than 40 percent, renewables approximately 24 percent, nuclear around 20 percent and coal approximately 15 percent.[13]

That distinction between electricity physically supplied and electricity companies contract to purchase is important. Technology companies frequently announce renewable-energy

purchases, power-purchase agreements and targets such as matching annual electricity consumption with renewable generation. Those arrangements can finance additional renewable capacity, but they do not necessarily mean that a particular data center is physically operating on wind or solar electricity every hour. The IEA therefore bases its analysis on the electricity physically consumed through local grids rather than companies' contractual procurement claims.[13]

The industry's enormous growth also means that cleaner electricity procurement must be considered alongside additional electricity demand. The IEA projects global data-center electricity consumption to rise from approximately 485 TWh in 2025 to about 950 TWh in 2030, with electricity use by AI-focused data centers growing even faster.[13]

Renewables are expected to supply a substantial part of that growth. The IEA projects them to meet nearly half of additional global data-center electricity demand through 2030. However, fossil fuels are not disappearing from the expansion: natural gas and coal together are projected to provide more than 40 percent of the additional electricity required by data centers through 2030.[13]

The projected mix indicates that near-term data-center growth is expected to increase demand for both renewable and fossil-fuel generation. Technology companies are helping finance renewable projects, and renewable generation serving data centers is increasing rapidly, but under the IEA's current outlook low-carbon generation does not supply all additional demand through 2030.[13]

Cleaner electricity can nevertheless address multiple components of the environmental problem simultaneously. Increasing electricity supplied by wind and solar PV can reduce greenhouse-gas emissions while generally reducing operational water requirements relative to water-cooled thermoelectric generation. Greater energy efficiency can reduce both electricity demand and the water associated with producing that electricity. The IEA specifically identifies expansion of solar PV and wind, together with efficiency, as mechanisms capable of reducing water withdrawals across the power sector.[49]

But even here, location matters. Adding renewable generation somewhere on a national grid is not necessarily equivalent to reducing fossil generation serving a constrained data-center region at the time the facility operates. Transmission capacity, grid congestion, storage, hourly generation profiles and the availability of dispatchable electricity all influence what actually supplies the load.

For water accounting, the same principle applies. A company purchasing renewable-energy credits cannot automatically assign the operational water intensity of wind or solar to every kilowatt-hour consumed by its data center unless the accounting methodology

actually supports that attribution. Berkeley Lab's national data-center analysis deliberately used the regional grid mix associated with facility locations and did not incorporate individual corporate power-purchase agreements.[15]

Cleaner and lower-water electricity therefore represents a genuine opportunity, but not a complete solution. Reducing the water intensity of the grid can reduce AI's indirect water footprint, while reducing its carbon intensity can reduce climate impacts. Ideally, electricity expansion accomplishes both. Where those objectives conflict, the tradeoff should be reported rather than hidden behind a generic “clean energy” label.

## 9.6 Corporate Water Commitments

Major technology companies have increasingly adopted formal water commitments alongside their carbon and renewable-energy goals. Microsoft, Meta and Amazon Web Services have each committed to becoming “water positive” by 2030, although the exact definitions, accounting boundaries and strategies differ among companies.[50-52]

The phrase water positive does not mean that a company stops consuming water. Generally, these commitments combine reductions in operational water use with projects intended to replenish, restore or improve water resources elsewhere. Consequently, a company can continue consuming substantial quantities of water while reporting progress toward water positivity if qualifying restoration or replenishment projects generate sufficient credited water benefits.

Microsoft committed in 2020 to becoming water positive by 2030. Its strategy includes reducing operational water intensity, replenishing water in stressed watersheds, expanding access to water and sanitation, investing in water-related innovation and supporting water policy. Microsoft reported approximately 50 million cubic meters of contracted annual water-replenishment volume in FY2024, up from approximately 2.8 million cubic meters in FY2020.[50] These figures represent contracted replenishment volumes rather than evidence that Microsoft's physical water consumption has fallen by the same amount.

Meta has similarly committed to becoming water positive in 2030. Its framework is explicitly sensitive to local water stress: Meta aims to restore 200 percent of the water it consumes in high-water-stress regions and 100 percent in medium-water-stress regions. The company reports that its operational restoration projects returned more than 1.6 billion gallons of water to high- and medium-water-stress regions in 2024 and that it has funded more than 40 restoration projects.[51]

AWS announced its water-positive commitment in 2022, promising to return more water to communities and the environment than it uses in its direct data-center operations globally by 2030. Amazon reports that AWS improved its average data-center Water Usage Effectiveness from earlier levels to 0.12 liters per kilowatt-hour in 2025, a 52 percent improvement since 2021. Amazon also reports that AWS reached 75 percent of the way toward its water-positive target in 2025, up from 53 percent in 2024.[52] These are company-reported metrics.

These commitments reflect substantial changes in how major technology companies measure, manage and address water use. Companies are measuring water efficiency, financing watershed projects, deploying alternative water sources and increasingly considering water stress when making infrastructure decisions. The cooling and reuse examples discussed earlier, including Microsoft's newer zero-water-evaporation cooling architecture, Google's decision to air-cool its Mesa facility and the Quincy reuse system, demonstrate that at least some corporate commitments are accompanied by changes to physical infrastructure rather than existing solely as replenishment pledges.[26, 41, 42]

However, replenishment is not physically equivalent to eliminating consumption at a data center. Water consumed through evaporation in one watershed cannot literally be returned to that same place and time simply by funding an equivalent volume of water benefits somewhere else. A restoration project may produce genuine ecological or community benefits, but volumetric accounting should not obscure the location and timing of the original consumption.

This is particularly important because water is fundamentally local. Saving one million gallons in a water-abundant watershed is not necessarily environmentally equivalent to consuming one million gallons from a severely stressed watershed. Meta's decision to apply a 200-percent restoration target in high-water-stress regions implicitly recognizes this problem by weighting restoration expectations according to local conditions.[51]

The meaning of “returned” or “replenished” water also deserves scrutiny. Corporate portfolios can include projects such as irrigation-efficiency improvements, wetland and watershed restoration, groundwater recharge, leak reduction and improved community water access. These activities can create real benefits, but their effects are not identical. A gallon estimated to have been conserved through an irrigation project is not physically the same intervention as a gallon of treated wastewater returned to an aquifer.

Additionality is another important question: would the claimed water benefit have occurred without the company's investment? A replenishment project provides a stronger environmental claim when corporate financing causes an improvement that otherwise

would not have occurred. Baseline assumptions, project duration and monitoring therefore affect how much confidence should be placed in reported volumetric benefits.

Timing matters as well. Microsoft's reported contracted replenishment volume, for example, describes the expected benefits of projects under contract rather than necessarily the amount of water physically replenished during that reporting year.[50] Amazon similarly distinguishes water already returned through active projects from larger annual volumes expected from its contracted portfolio.[52]

Corporate reporting boundaries can further complicate comparisons. AWS's commitment concerns water used in its direct data-center operations, while Microsoft describes a broader global operational commitment, and Meta applies its restoration framework according to water stress in watersheds where it operates.[50-52] A percentage reported by one company therefore should not automatically be ranked against another company's percentage as though the accounting systems were identical.

This is especially relevant to AI. The largest cloud providers operate enormous conventional cloud-computing businesses in addition to AI infrastructure. Their corporate water reports generally do not provide enough information to isolate the water consumption associated specifically with AI from databases, storage, web hosting, enterprise software and other workloads sharing the same infrastructure. A company's progress toward a data-center water goal therefore should not be described as an AI-specific water reduction unless the underlying reporting actually separates AI.

Corporate water commitments should consequently be treated as neither meaningless public relations nor proof that the industry's water problem has been solved. They can finance genuine conservation and restoration, encourage operational efficiency and influence facility design. But the strength of a commitment depends on what is actually measured, where the benefits occur, whether projects create additional benefits, and whether direct consumption is being reduced rather than merely balanced through accounting.

For this report, the strongest evidence of progress is therefore not the phrase “water positive” itself. It is measurable change underneath that commitment: lower water consumption per unit of computing, reduced potable-water dependence, appropriate use of reclaimed water, cooling technologies adapted to local scarcity, transparent facility-level reporting and restoration projects that produce independently verifiable benefits in the watersheds where pressure occurs.

## 9.7 What “Water Positive” Actually Means

“Water positive” can sound as though a company produces more water than it consumes. That is not what the term means. Water is not being manufactured, and water consumed at a data center is not necessarily being physically returned to the same source. Corporate water-positive strategies generally combine efficiency, reduction, substitution and replenishment, each of which affects water resources in fundamentally different ways.[50-52]

***Efficiency means accomplishing the same or greater amount of useful computing with less water per unit of activity.***

For data centers, this is commonly expressed through Water Usage Effectiveness (WUE), which relates facility water use to the energy consumed by computing equipment. AWS, for example, reports improving its global average WUE to 0.12 liters per kilowatt-hour in 2025.[52]

Efficiency is valuable, but it is a rate rather than a total. A rapidly expanding company can improve water efficiency while its total water demand still increases. The same distinction appeared earlier with AI energy efficiency: using fewer resources per unit of computing does not guarantee lower aggregate resource use when the amount of computing grows rapidly.

***Reduction means actually decreasing the amount of water withdrawn or consumed.***

This can result from eliminating evaporative cooling, improving cycles of concentration, repairing leaks, changing operating practices or reducing the underlying cooling requirement. Microsoft's closed-loop cooling architecture discussed earlier provides one example of a design intended to eliminate routine water evaporation for cooling.[41]

Reduction is therefore different from efficiency. A facility could improve its WUE while still consuming more water because its computing load increased. Conversely, a facility that changes cooling systems might reduce its absolute water consumption even while its computing capacity expands. Both measurements are useful, but they answer different questions.

***Substitution means replacing one water source with another rather than necessarily reducing total water use.***

A data center that switches from potable drinking water to reclaimed wastewater may continue consuming a similar volume of water while substantially reducing its demand on

potable supplies. The Quincy reuse system demonstrates this distinction: alternative supply and reuse substantially reduced Microsoft's dependence on potable groundwater without eliminating the facility's need for cooling water.[26]

AWS similarly identifies recycled water, reclaimed wastewater and rainwater among the alternative supplies it uses to reduce dependence on potable sources.[52] The company reports using reclaimed water at data centers in multiple locations, including Northern Virginia and California.

Substitution can produce significant environmental benefits, particularly where drinking-water supplies are constrained. But “less potable water” does not automatically mean “less total water consumed.” The source, scarcity and competing uses of the replacement water still need to be evaluated.

### *Replenishment is different from all three.*

Instead of changing the amount or type of water used directly by a facility, replenishment projects are intended to create measurable water benefits outside the facility. Examples can include restoring wetlands, improving irrigation efficiency, repairing municipal leaks, increasing groundwater recharge or improving watershed conditions.[50-52]

These benefits can be quantified through methodologies such as Volumetric Water Benefit Accounting (VWBA). The World Resources Institute and its partners developed VWBA to provide a standardized framework for estimating the water benefits produced by stewardship projects. The updated VWBA 2.0 guidance emphasizes local watershed context, scientifically defensible methods, appropriate time periods and avoiding double counting.[53]

This distinction is critical because a volumetric water benefit is not necessarily a physical gallon returned to a data center's original water source. If a company funds an irrigation-efficiency project that prevents one million gallons from being withdrawn from a watershed, that project may create an estimated one-million-gallon water benefit even though no tanker or pipeline literally carries one million gallons back to the data center's municipal supply.

Likewise, restoring wetlands, reducing leaks and improving groundwater recharge can provide legitimate hydrologic and ecological benefits without reversing the specific evaporation that occurred at a cooling tower. WRI's methodology therefore treats water stewardship at the catchment or watershed scale and emphasizes understanding the particular water challenges where projects occur.

This explains how a company can continue consuming water and nevertheless become “water positive” under its stated accounting framework. AWS, for example, defines its 2030 commitment as returning more water to communities and the environment than it uses in its direct data-center operations. Its strategy combines greater efficiency, alternative water sources, reuse and replenishment projects.[52]

These categories should therefore not be collapsed into a single number without explanation:

- Efficiency = less water required per unit of activity.
- Reduction = less water actually withdrawn or consumed.
- Substitution = replacing a higher-value or more stressed water source with another source.
- Replenishment = creating water benefits through projects outside the facility.

A company could improve all four simultaneously, but they are not interchangeable. A 20-percent improvement in water efficiency is not necessarily a 20-percent reduction in total consumption. Replacing 100 million gallons of potable water with 100 million gallons of reclaimed water does not mean 100 million fewer gallons were consumed. And a replenishment project credited with 100 million gallons of volumetric benefit does not mean the company physically returned 100 million gallons to the exact source from which its facility consumed water.

This does not make water-positive commitments meaningless. Replenishment projects can produce real benefits, and standardized accounting frameworks are becoming more sophisticated about requiring projects to address actual local water challenges. WRI's updated guidance specifically emphasizes local catchment conditions, appropriate temporal accounting, conservative assumptions and prevention of double counting.[53]

But the terminology can become misleading when a water-positive claim is presented without explaining what produced it. A company reaching its target primarily through reduced physical consumption represents a different outcome from one reaching the same numerical target primarily through externally credited replenishment projects.

For this report, corporate water performance should therefore be examined in layers rather than reduced to a single “water positive” label. The strongest evidence asks: Is the company using less water per unit of computing? Is its total withdrawal and consumption decreasing? Is it reducing reliance on potable or stressed supplies? And separately, what measurable benefits are its replenishment projects producing, where are those benefits occurring and how are they being calculated?

“Water positive” is most useful when understood as a water-stewardship goal rather than a statement that a company's physical water consumption has become negative. The underlying efficiency, reduction, substitution and replenishment measures tell us considerably more about what is actually changing.

## 9.8 Limits of Water Replenishment

Water-replenishment projects can produce real benefits, but they do not make the location and timing of a data center's original water consumption irrelevant. A gallon of water conserved, restored or replenished somewhere else is not automatically hydrologically equivalent to a gallon consumed at the facility, particularly when the two occur in different watersheds or at different times.[53]

This limitation follows from a basic characteristic of freshwater: water impacts are local. Unlike greenhouse gases, where a ton of carbon dioxide has global climatic significance regardless of where it is emitted, the consequences of consuming water depend strongly on the particular watershed, aquifer, season and competing demands involved. Removing water from a stressed basin can therefore create a different impact from removing the same volume in a water-abundant basin.

### *Watershed matching*

The strongest replenishment projects begin with the watershed where the water challenge actually exists. WRI's Volumetric Water Benefit Accounting 2.0 framework begins by requiring practitioners to understand the local catchment context, including hydrological, social, political and governance conditions and the shared water challenges affecting that location.[53]

This means that volume alone is insufficient. Imagine a data center consumes 100 million gallons annually from a water-stressed watershed while the company finances a project credited with conserving 100 million gallons in a water-abundant watershed hundreds of miles away. The corporate accounting may record equal volumes, but the second project has not restored the first watershed's lost water availability.

A project outside the facility's watershed can still provide genuine environmental or community benefits. It simply should not be presented as though geographic separation does not matter.

This is one reason corporate commitments increasingly prioritize replenishment in water-stressed areas. As discussed in Section 9.6, Meta, for example, applies different

restoration targets according to local water stress.[51] That approach acknowledges that equal volumes can carry unequal environmental significance.

### *Timing*

When water is replenished can matter almost as much as where.

A facility may create its greatest water demand during summer heat or drought, precisely when a watershed has the least capacity to accommodate additional consumption. A restoration project that generates its calculated benefit during a wetter season may improve annual water availability without replacing the water that was unavailable during the critical period.

WRI therefore requires the use of an appropriate temporal scale when calculating volumetric water benefits.[53]

This creates an important distinction between annual accounting and real-time water availability. A company could theoretically consume 100 million gallons during a watershed's most stressed months and generate a credited 100-million-gallon benefit over the entire year. Those annual quantities may balance numerically while their effects on seasonal water security do not.

The same issue applies across years. A company might begin consuming water when a facility opens while a restoration project takes several years to reach its expected benefit. Conversely, some projects may create improvements that persist for many years. Credible accounting therefore needs to identify when benefits actually begin, how long they persist and whether reported volumes represent benefits already delivered or benefits expected in the future.[53]

### *Additionality*

A replenishment claim is also stronger when the company's intervention creates a benefit that would not otherwise have occurred.

Consider a municipal water utility already planning and fully funding a pipeline-repair project. If a company later contributes money to the project, it would be difficult to attribute the entire volume of water saved by those repairs to that company if the work would have happened anyway.

This is the problem of additionality.

VWBA 2.0 addresses this through project eligibility, baseline conditions and attribution requirements. Projects must establish an appropriate comparison between conditions with and without the intervention, and corporate claims should represent the company's actual contribution to the activity rather than automatically claiming the project's entire benefit.[53]

Additionality can sometimes be difficult to establish with certainty. Watershed projects frequently involve governments, utilities, nonprofits, landowners and multiple corporate sponsors. The appropriate question is therefore not simply whether the company provided money, but what changed because of that contribution and how much of the resulting benefit can reasonably be attributed to it.

### *Verification*

Finally, projected water benefits need to be measured and tracked.

Replenishment projects can involve activities as different as repairing leaking pipes, improving agricultural irrigation, restoring wetlands, changing reservoir operations or increasing groundwater recharge.[53] These interventions require different measurement methods and contain different uncertainties. A calculated volumetric benefit is therefore an estimate produced using a defined methodology, not necessarily a directly metered quantity of water.

WRI recommends scientifically defensible methods, documented and conservative assumptions, appropriate time scales and safeguards against double counting. Its framework also calls for implementation and tracking plans and for companies to confirm that claimed benefits reflect the project's actual status, duration and the company's contribution before communicating them.[53]

Avoiding double counting is particularly important when several organizations finance the same project. If a watershed-restoration project produces an estimated benefit of 100 million gallons, three participating companies cannot each claim the entire 100 million gallons without collectively turning one physical benefit into 300 million gallons of corporate accounting. Attribution rules are needed to determine which organization can claim what portion.[53]

Verification also matters because expected and realized benefits can differ. A groundwater-recharge project may perform differently depending on rainfall and soil conditions. Irrigation improvements may produce different savings depending on how farmers subsequently operate the system. Wetland restoration may take years to mature.

Monitoring allows estimates and claims to reflect what the project actually accomplishes rather than only what was predicted when it was announced.

These limitations do not make replenishment ineffective. Well-designed projects can reduce water withdrawals, restore aquatic habitat, recharge groundwater, improve watershed resilience and address shared water challenges. WRI developed its accounting framework specifically to make those benefits more credible and comparable.

But replenishment should be understood as complementary to reducing the original impact, not a license to ignore it. The strongest water strategy follows a practical hierarchy: first improve efficiency and avoid unnecessary consumption; reduce demand where feasible; substitute reclaimed or less-stressed supplies where appropriate; and then use credible replenishment projects to address remaining impacts and shared watershed challenges.

A company's claim that it has replenished a volume equal to or greater than its consumption therefore does not establish that every local consequence of that consumption has been reversed. The more meaningful questions are where the original water was consumed, where the claimed benefit occurred, whether the timing addresses the same period of stress, whether the project created an additional benefit, how that benefit was measured and whether the claimed volume has been independently or otherwise credibly verified.

For this report, “water positive” should therefore never be interpreted as “water impact erased.” A credible replenishment program can make a company a better water steward while its facilities still create local water demands requiring their own management.

## 9.9 The Transparency Problem

Evaluating whether AI companies are improving their water performance ultimately depends on the information they make available. Major technology companies now publish considerably more environmental data than they did a decade ago, including corporate water withdrawals, consumption, Water Usage Effectiveness, replenishment projects and, in some cases, facility-level information.[50-52] This represents genuine progress. It still does not provide enough standardized information to answer many of the questions raised throughout this report.

One problem is that companies do not necessarily report the same measurement boundaries. A water figure may describe an entire corporation, a cloud-computing division, owned data centers, a particular facility or only water used directly for cooling. Some

companies report withdrawal and consumption separately; others emphasize WUE, potable-water reductions or replenishment volumes. Supply-chain water and electricity-related water may be reported separately, estimated using different methodologies or excluded entirely.[50-53]

Consequently, apparently comparable numbers may measure different things. A company reporting lower WUE than another company has not necessarily demonstrated that its total water footprint is smaller. WUE measures the relationship between facility water use and IT energy under a particular accounting methodology; it does not by itself capture the facility's absolute consumption, local water scarcity, electricity-related water or embedded water associated with manufacturing hardware.

Corporate totals create another limitation. Microsoft, Google, Amazon and Meta operate data centers supporting far more than generative AI. Their infrastructure also provides cloud storage, databases, search, advertising, social media, streaming, enterprise software and other digital services. Conversely, AI workloads may operate on infrastructure owned by another company. Corporate water use therefore cannot simply be labeled “AI water use.”

Even within a data center, publicly available information generally does not reveal exactly how much water should be attributed to individual workloads. Servers can process different models and applications, cooling systems serve shared equipment, and facility infrastructure supports the entire building. Without workload-level operational data, researchers cannot reliably divide the facility's water meter among ChatGPT requests, search queries, image generation, cloud storage and other services sharing that infrastructure.

Geographic aggregation can hide important differences as well. Water is a local resource, so a global corporate total can combine consumption from water-abundant and water-stressed locations into one number. A company could improve its global water efficiency while increasing consumption at a particular facility in a constrained watershed. Conversely, increasing global consumption does not establish that every individual facility is creating local water stress. Facility and watershed-level information is necessary to distinguish those outcomes.

The Dalles case demonstrates why this information can matter. Google's water use became a substantial portion of the city's municipal demand while facility-specific information remained unavailable to the public until litigation resulted in disclosure.[35, 37] The eventual records made it possible to discuss the scale of the demand using actual measurements rather than speculation. The controversy itself demonstrates how secrecy

can make an environmental question harder to evaluate even when the underlying water use may ultimately prove manageable.

Some disclosure gaps arise from legitimate commercial and security concerns. Detailed information about server utilization, hardware configurations, model architecture and facility operations can reveal competitively sensitive information. However, protecting proprietary technical information does not necessarily require withholding basic environmental indicators such as annual withdrawal, consumption, water source, cooling technology and watershed location.

AI creates an additional transparency problem because models change rapidly. A published measurement for one model, hardware generation or serving configuration may become outdated as providers replace models, modify inference systems or deploy new accelerators. Even when a company publishes a credible “water per prompt” or “energy per query” measurement, that number should not automatically be applied to other models or treated as a permanent characteristic of the service.[10, 11]

Independent researchers therefore frequently have to estimate what companies do not disclose. These studies can provide valuable ranges and reveal the factors controlling AI's environmental footprint, but a modeled estimate is not the same thing as a measured company disclosure. Throughout this report, those categories have been kept separate for precisely this reason.

Transparency problems also complicate corporate water-positive claims. Replenishment portfolios introduce additional measurements involving project baselines, location, timing, attribution and estimated volumetric benefits.[53] A company may publish a large replenishment number without that figure answering how much water its individual data centers consume in stressed watersheds. Both pieces of information can be useful, but one should not substitute for the other.

For the same reason, this report does not rank AI companies from “best” to “worst” on water performance using currently available public numbers. Such a ranking would imply a level of comparability the underlying disclosures do not provide. A provider reporting 0.12 L/kWh WUE cannot responsibly be declared more water-efficient overall than another provider whose published figure uses a different facility boundary, cooling methodology, geographic portfolio or accounting period.[52]

Nor should absence of disclosure automatically be interpreted as evidence of worse environmental performance. Unknown is not the same as bad. A company providing less public information deserves a lower level of confidence in claims about its water

performance, but missing data cannot legitimately be converted into an estimate of environmental harm.

A useful transparency standard would make several pieces of information consistently available: annual facility-level water withdrawal and consumption; potable, reclaimed and other water sources; cooling technology; monthly or seasonal demand where water stress is relevant; wastewater and discharge volumes; electricity consumption and grid assumptions; and clear definitions of the organizational and operational boundaries used in reporting. For AI-specific claims, providers would additionally need standardized workload categories and enough information about hardware and utilization to make comparisons meaningful.

Independent verification would further strengthen those disclosures. Corporate sustainability reports are valuable primary sources for understanding what companies say they consume, conserve and replenish, but independent measurement, regulatory reporting and peer-reviewed research provide different forms of evidence. No single category should automatically replace the others.

Greater transparency would allow independent researchers, regulators and communities to evaluate reported water performance against actual operating data. Consistent disclosure would make it easier to identify where efficiency measures reduce consumption, where total demand is increasing and where facilities place pressure on shared resources. The purpose is to establish an evidence base that supports scrutiny and accountability.

The findings support two conclusions simultaneously: corporate environmental reporting has improved, and major gaps remain. Companies increasingly disclose water efficiency, total use, alternative-water projects and replenishment efforts, while workload-level AI water consumption, standardized facility-level reporting and comparable accounting boundaries remain limited.

Until those gaps close, comparisons need to retain their labels: Measured, Modeled, Projected and Company-Reported. A precise-looking ranking assembled from incompatible numbers would not make the AI-water debate more informative. It would simply replace missing information with false precision.

## 10. Will AI's Water Problem Get Better or Worse?

AI's future water footprint will be shaped by two trends moving in opposite directions. The technology is becoming more efficient, but the amount and intensity of AI computation are

also increasing rapidly. Improvements in hardware, models, inference serving and cooling can reduce the water associated with a comparable amount of useful computation, while wider adoption, reasoning, agents and multimodal generation can push total resource demand upward.

This distinction between water intensity and total water demand is central to forecasting AI's environmental impact. A future AI system may require substantially less energy and water to perform the same task than a system operating today. But if the number of users, requests and computationally demanding workloads grows faster than those efficiency gains, aggregate electricity and water demand can still increase.

The outcome is also not determined by AI models alone. Cooling technology affects direct water consumption; the electricity mix affects indirect water use; reclaimed and non-potable supplies can reduce competition for drinking water; and siting determines the watershed and infrastructure that ultimately experience the demand.[19, 26, 46] Peer-reviewed scenario research shows that changing these variables can produce dramatically different future water footprints even under continued growth.[19]

Forecasts should therefore be treated as conditional scenarios rather than predictions of an inevitable outcome. Extrapolating today's water intensity indefinitely ignores technological improvement, while assuming efficiency will automatically overcome growth ignores the rebound effect and the increasing computational requirements of emerging AI applications.

This leads to a more useful question than simply whether AI's water use will rise or fall: Can improvements in hardware, cooling, electricity, water sourcing and siting reduce resource intensity faster than the scale and computational demands of AI expand? The answer to that question will largely determine whether AI's water footprint becomes more manageable or develops into a substantially larger infrastructure challenge.

## 10.1 AI Is Becoming More Efficient

One reason AI's future water footprint is difficult to predict is that the technology itself is not standing still. As hardware, models and inference systems improve, the energy required to perform a given AI task is declining, potentially reducing the water needed for cooling and electricity generation.[11, 13]

At the hardware level, newer AI accelerators can perform substantially more computation for each unit of electricity consumed. Specialized GPUs and other accelerators, lower-precision computation and other hardware improvements allow more useful AI work to be

performed without electricity demand increasing proportionally with computing capability.[13]

Software improvements can compound those gains. Quantization, model compression, knowledge distillation and more efficient model architectures can reduce the computational requirements of inference,[44] while production-serving techniques such as batching, caching and improved accelerator utilization allow infrastructure to process more requests using the same hardware.[11]

The Joule analysis discussed earlier demonstrates how significant these combined improvements could be. Its researchers estimated that improvements in models, serving systems and hardware could collectively reduce the energy required for comparable AI inference by approximately 8 to 20 times relative to the study's baseline assumptions.[11] The estimate should not be interpreted as meaning that every future AI query will automatically use one-twentieth as much energy; it demonstrates the scale of efficiency improvement technically available when multiple layers of the system improve together.

Cooling technology is evolving at the same time. Direct-to-chip liquid cooling can remove heat more efficiently from increasingly dense AI hardware, while closed-loop and dry heat-rejection systems can greatly reduce or eliminate routine evaporative water consumption.[41, 45] Improvements in computational efficiency can also reduce the amount of heat generated for a comparable workload, lowering the cooling requirement in the first place.

These changes matter when interpreting projections. A projection that assumes tomorrow's AI systems require exactly the same energy and water per unit of computation as today's systems can overstate future demand if efficiency continues improving rapidly. Conversely, assuming that future efficiency gains will automatically translate into proportional reductions in total water use can understate demand if AI deployment grows even faster.

There is already precedent for this distinction elsewhere in computing. Data-center computing output increased enormously during the 2010s while improvements in servers, virtualization and facility efficiency restrained electricity growth far more than would have occurred if energy intensity had remained constant. The current AI expansion is placing new pressure on that historical relationship, but it demonstrates why growth in computing activity and growth in resource consumption do not necessarily occur at the same rate.[13, 15]

Efficiency improvements can reduce multiple components of AI's water footprint simultaneously. Lower electricity requirements can reduce the water associated with electricity generation. Lower heat production can reduce cooling requirements. More

efficient cooling can reduce direct consumption. And better utilization can reduce the amount of physical infrastructure required to perform a given amount of useful computing.

But efficiency should be measured against a comparable task. A newer reasoning model may be substantially more efficient per unit of computation than an older model while still using more total energy for a particular request because it performs much more computation. As Section 4 demonstrated, ordinary text inference and long reasoning can differ dramatically in resource requirements.[11] An efficiency improvement does not necessarily mean that every newer or more capable AI interaction has a smaller footprint than every older one.

These developments give us a genuine reason to expect AI's water intensity to improve. Hardware, software, inference serving and cooling are all moving toward accomplishing more computing with fewer resources per unit of work.[11, 13, 41, 44, 45]

That is different, however, from saying AI's total water footprint will decline. Efficiency determines how many resources are required for a given amount of AI activity. Total demand depends on how much AI activity ultimately occurs.

The distinction between falling resource intensity and potentially rising total demand is the central tension for the rest of Section 10.

## 10.2 Workloads Are Also Becoming More Demanding

Efficiency is only one side of AI's future resource equation. At the same time that individual computations are becoming more efficient, AI systems are being asked to perform increasingly computationally demanding tasks. Long reasoning, agentic workflows, multimodal processing and generative image, audio and video applications can require substantially more computation than an ordinary text response.[11, 13]

Reasoning provides one of the clearest measured examples. Language-model inference generates responses token by token, while reasoning systems can perform substantially more computation at inference time and produce much longer outputs before completing a task. This approach is often described as test-time scaling: additional computation during inference can improve performance on difficult problems, but the additional computation also requires additional energy.[11]

The 2026 Joule study modeled a conventional frontier-scale text query with a median output of approximately 300 tokens and estimated median energy use of 0.31 Wh per query. When the researchers modeled a long-reasoning workload with approximately 5,000

median output tokens, estimated median energy increased to 3.91 Wh per query, about 13 times the standard text-query estimate.[11]

The approximately thirteenfold increase applies to the workloads modeled in this study. Energy requirements vary with model architecture, hardware, output length and serving configuration. By applying a consistent methodology to different workloads, the researchers demonstrated that extended reasoning can increase energy requirements by more than an order of magnitude within text-based inference.[11]

The effect becomes significant at scale. In the Joule study's illustrative scenario, serving one billion conventional queries per day required approximately 0.7 GWh per day under baseline assumptions. If only 10 percent of those requests became long reasoning queries, estimated electricity demand increased to approximately 1.7 GWh per day, more than double the original total.[11]

Agentic AI can increase computational demand for a related reason. Instead of receiving one request and producing one response, an AI agent may perform a sequence of operations: planning a task, searching or retrieving information, calling tools, evaluating intermediate results, revising its approach and continuing until the objective is completed. One apparent user request can therefore produce multiple model inference steps and substantially more generated tokens behind the scenes.[11]

The Joule researchers specifically identify long reasoning and token-intensive agentic workflows as an important source of future inference-energy growth. Longer outputs can also reduce serving concurrency, meaning infrastructure may process fewer simultaneous requests as efficiently as it can with shorter conventional interactions.[11]

Multimodal AI expands the range of workloads further. AI systems increasingly process combinations of text, images, audio and video rather than text alone. Generating these media can involve fundamentally different computational processes from generating a short text response, and the resource requirements should not be assumed to be interchangeable.

Video generation is particularly important. The IEA identifies video generation, reasoning and agentic tasks as emerging energy-intensive AI applications and reports that these workloads can consume hundreds or thousands of times more electricity per query than simple text generation, depending on the task.[13]

That range should not be converted into a universal “video uses X times more water than text” claim. As Section 4.5 established, comparable high-quality water measurements for different generative media remain limited. Video length, resolution, model, hardware and

generation method can all substantially affect computation. The IEA's finding establishes the potential difference in energy demand, not a standardized water-per-video number.

This also exposes a limitation of generalized per-prompt comparisons. Two requests submitted through the same interface can represent radically different computational workloads. Asking a model to answer a short factual question is not equivalent to asking it to reason through a difficult problem, analyze multiple documents, generate images or video, or operate as an agent completing a multistep task. Assigning all of them the same water or electricity number conceals precisely the workload variation that determines their resource requirements.[11]

The IEA describes AI's future energy demand as the interaction of three rapidly changing variables: efficiency improvements, increasing adoption and changing model capabilities that enable more energy-intensive applications. Although energy use per individual AI task has been falling rapidly, electricity consumption from AI-focused data centers still increased approximately 50 percent in 2025.[13]

Efficiency improvements are not therefore failing. Without those improvements, the resource requirements of increasingly capable AI could be substantially greater. The problem is that the task itself is changing while its efficiency is improving. Comparing an efficient future video-generation system with today's text chatbot is not the same thing as comparing two generations performing an equivalent workload.

Water follows this trend indirectly as well as directly. Greater computational demand requires more electricity and produces more heat that must be removed from computing equipment. Depending on facility cooling technology and electricity source, that can increase both direct cooling-water requirements and electricity-related water consumption. The relationship is not a fixed number, but the physical connection remains.

## 10.3 The Rebound Effect

Greater efficiency does not necessarily produce lower total resource consumption. If AI becomes cheaper and more efficient to operate, those improvements can also make it practical to use AI more frequently, for more purposes and at much larger scale. This tension between efficiency and expanding use is commonly described as a rebound effect.[13, 54]

The basic arithmetic is straightforward. Suppose technological improvements reduce the energy required for a comparable AI task by 50 percent. If the number of tasks remains unchanged, total energy demand falls by approximately half. But if the number of tasks

triples, total demand increases despite the substantial efficiency improvement. The same principle can apply to cooling and water.

In this example, the efficiency improvement limits total demand to 1.5 times its original level; without it, tripling activity would triple demand. Efficiency therefore reduces consumption relative to what it would otherwise have been, even though total consumption still rises.

AI creates particularly strong conditions for this effect because efficiency improvements can lower both the financial and technical barriers to deployment. More efficient models, hardware and serving systems can make AI economical for applications that previously would have been too expensive or computationally demanding.[11, 13] At the same time, AI is being integrated into existing digital activities while emerging applications such as extended reasoning, agents and generative video can require substantially more computation than conventional text inference.[11, 13] Efficiency therefore operates against both more AI interactions and potentially more computation within each interaction.

The IEA's projections illustrate the result at the infrastructure level. Despite rapid improvements in AI hardware and software efficiency, electricity consumption from data centers is expected to continue increasing substantially as deployment expands. The IEA projects global data-center electricity consumption to reach approximately 945 TWh by 2030, more than double its 2022 level, with AI identified as the principal driver of that growth.[13]

The IEA also demonstrates why efficiency still matters enormously. In its High Efficiency Case, stronger improvements in software, hardware and infrastructure substantially reduce projected data-center electricity demand relative to its Base Case. In the opposite direction, slower efficiency progress produces considerably higher demand.[13]

Water adds another layer because the rebound effect does not translate into water demand at a fixed rate. More computation means more electricity and heat, but the resulting water requirement depends on where that computation occurs, the cooling system used and the electricity supplying it. A new workload operating in a dry-cooled facility powered largely by low-water electricity can have a very different water consequence from the same workload served by evaporative cooling and water-intensive thermal generation.

Cooling improvements can themselves experience a form of scale pressure. If a new cooling design cuts direct water consumption per megawatt dramatically but the company subsequently builds several times as much computing capacity, total corporate or regional water consumption may remain unchanged or increase even though every new megawatt

is substantially more water-efficient. The technology succeeded; growth simply exceeded the savings it produced.

Water Usage Effectiveness and similar intensity metrics require the same distinction. Falling WUE is evidence that a facility or fleet requires less water relative to its IT energy use, but it does not establish that absolute water consumption is falling. A company can simultaneously report its best-ever WUE and its highest-ever total water consumption if its computing infrastructure expands sufficiently quickly. Reporting only an improving intensity metric can therefore obscure rising absolute demand, while reporting only rising total consumption can conceal efficiency improvements that prevented demand from becoming even larger. Both measurements are necessary to understand what is happening.

For water planning, absolute demand ultimately matters locally. A municipal utility must supply actual gallons, not improvements in gallons per unit of computing. A watershed experiencing drought responds to total withdrawals and consumption during the period of scarcity, regardless of how efficient the underlying servers have become. Efficiency remains extremely valuable, but infrastructure planning must account for the demand that remains after those efficiency gains.

If efficiency, low-water cooling and cleaner electricity improve faster than AI deployment expands, total water impacts could stabilize or decline. If computing demand grows faster, total water demand can continue rising even as the footprint of an equivalent individual task becomes dramatically smaller. And because both sides of that equation are currently changing rapidly, long-range projections carry substantial uncertainty.

The rebound effect therefore does not establish that AI efficiency improvements are pointless. It establishes almost the opposite: efficiency determines how resource-intensive AI's expansion will be, but it does not determine how large that expansion will become.

## 10.4 Growth Scenarios

Because both AI efficiency and AI demand are changing rapidly, long-term water forecasts should be treated as scenarios rather than predictions of a single inevitable future. Recent peer-reviewed studies therefore present ranges based on different assumptions about server deployment, efficiency, location, cooling and electricity supply rather than claiming to know precisely how much water AI will require years from now.[19]

One of the most detailed recent analyses was published in Nature Sustainability in 2025. Xiao and colleagues modeled the operational energy, water and carbon impacts of AI servers in the United States from 2024 through 2030 under multiple deployment scenarios. Their scenarios include low demand, low power, a mid-case, high application and high demand, allowing both expansion and technological efficiency to vary.[19]

Across the modeled expansion scenarios, the researchers estimated a U.S. AI-server annual water footprint of approximately 731 million to 1.125 billion cubic meters over the 2024–2030 modeling period.[19] That is approximately 193 billion to 297 billion gallons annually. These figures are modeled projections, not measurements of current AI water use.

The study's definition of water footprint is also important. Approximately 71 percent of the modeled water footprint was indirect, associated principally with electricity generation, while approximately 29 percent was direct.[19] The figures therefore should not be described as hundreds of billions of gallons physically flowing into AI data centers. Most of the modeled footprint occurs outside the facility boundary.

The scenarios also show how strongly the outcome depends on where AI infrastructure is built. Under the researchers' modeled siting strategies, concentrating servers in locations with more favorable water conditions could reduce the water footprint by approximately 52 percent, while unfavorable spatial distribution could increase it by as much as 354 percent relative to the study's baseline allocation.[19]

Efficiency also materially changes the scenarios. The Nature Sustainability researchers modeled improvements in facility efficiency, advanced liquid cooling and server-utilization optimization. Collectively, the best industry practices examined could reduce the modeled water footprint by approximately 32 percent compared with the study's baseline assumptions.[19]

When the researchers combined favorable industry efficiency, siting and grid-development assumptions, the modeled water footprint could be reduced by as much as 86 percent relative to the corresponding unmitigated trajectory.[19] This is a large potential improvement, but it represents a modeled best-practice scenario. It does not establish that the industry has already achieved an 86-percent reduction or that such an outcome will automatically occur.

The opposite scenarios are considerably less favorable. Under combinations of rapid deployment, inefficient practices, unfavorable locations and slower improvements to the electricity system, the researchers found substantially larger residual water footprints. The study concludes that achieving corporate net-zero-water ambitions by 2030 would be

difficult without substantial reliance on water-restoration mechanisms, particularly if AI-server growth follows the higher-demand trajectories.[19]

A separate peer-reviewed estimate provides a useful comparison. Li and colleagues, whose work was published in Communications of the ACM in 2025, projected that global AI could reach approximately 4.2–6.6 billion cubic meters of combined Scope 1 and Scope 2 water withdrawal in 2027, with approximately 0.38–0.60 billion cubic meters of water consumption.[10]

Those numbers should not be directly compared as though they measure the same thing as the Nature Sustainability figures. Li and colleagues estimate global AI operational withdrawal and consumption using their particular Scope 1 and Scope 2 framework, whereas Xiao and colleagues model the water footprint of U.S. AI servers using a different geographical boundary, methodology and time horizon.[9, 19]

Li and colleagues themselves caution that their 2027 projection is a rough first-order estimate subject to substantial uncertainty. Future electricity-generation water intensity could improve, AI energy demand may differ from the underlying projection, and the geographical distribution of AI workloads remains uncertain.[10]

That uncertainty is not a weakness unique to these studies. It reflects the reality of forecasting an industry whose hardware, models, cooling systems and demand are changing simultaneously. A scenario in which AI deployment grows rapidly but efficiency improves slowly produces a fundamentally different result from one in which the same deployment occurs alongside aggressive improvements in hardware, utilization, cooling and electricity supply.

The peer-reviewed projections therefore do not establish that AI is destined to consume a particular number of gallons in 2030. They establish something more important: under plausible growth trajectories, AI can become a significant water demand, but the size of that demand remains highly sensitive to decisions being made now about efficiency, cooling, electricity and location.

## 10.5 Siting Scenarios

The location of future AI infrastructure may be one of the strongest determinants of its water footprint. The Nature Sustainability analysis discussed in Section 10.4 found that the same growth in AI computing can produce dramatically different water impacts depending on where servers are deployed.[19] Future water demand therefore depends not only on how

much AI infrastructure is built, but also on the environmental and infrastructure characteristics of the places where it operates.

The resulting range was substantial. Relative to the study's baseline geographical allocation, its water-conscious siting scenario reduced the modeled AI-server water footprint by approximately 52 percent, while an unfavorable distribution increased it by as much as 354 percent.[19] These are modeled scenarios, not measured changes at existing facilities.

Siting also affects indirect water use because regional electricity systems differ in their water intensity. This was particularly important in the Nature Sustainability analysis, in which approximately 71 percent of the modeled AI-server water footprint was indirect, principally through electricity generation.[19] A location that minimizes on-site cooling water does not therefore necessarily minimize the total modeled water footprint.

Volumetric footprint is not the same as local water stress. One million gallons consumed in a water-abundant watershed does not necessarily have the same environmental significance as one million gallons consumed in a basin experiencing shortages or intense competition among users. The Nature Sustainability analysis demonstrates how geography can change modeled water footprint, while the watershed framework discussed earlier establishes why local conditions remain relevant even when volumetric footprints are similar.[19, 53]

The study also found that siting can interact powerfully with technological improvements. Better cooling, improved server utilization and other efficiency measures can reduce water intensity, while favorable geographical allocation can reduce it further. Under the researchers' most favorable combination of industry practices, siting and electricity-system assumptions, the modeled water footprint could be reduced by as much as 86 percent relative to the corresponding unmitigated trajectory.[19]

There is therefore no single nationally optimal location for every environmental objective. A region favorable for water may have limited transmission capacity; a location with abundant low-water electricity may face land, ecological or community constraints; and a cool climate may reduce cooling demand while providing limited reclaimed-water infrastructure. Siting cannot be reduced to moving facilities somewhere with “more water.” Water availability, water stress, electricity, wastewater infrastructure, cumulative development and other local constraints interact.

The Nature Sustainability scenarios nevertheless establish an unusually important point for this report: AI's future water footprint is geographically malleable.[19] The computing

demand may be global, but the physical infrastructure serving it must exist somewhere, and choosing that somewhere can dramatically change the resulting water footprint.

For Tampa Bay, this finding changes the question from simply “How much water does AI use?” to a more locally useful one: “What would the water footprint of AI infrastructure look like if it were built here?” Answering that requires Tampa Bay's actual water sources, climate, electricity mix, reclaimed-water infrastructure, seasonal conditions and cumulative development capacity, not a national average.

## 10.6 Mitigation Scenarios

The growth scenarios do not describe an unavoidable water future. The same research projecting substantial increases in AI infrastructure also shows that multiple interventions can reduce the water footprint of that growth, particularly when improvements in computing hardware, cooling, electricity supply and siting occur together.[19]

These interventions can reinforce one another. The Nature Sustainability analysis modeled combinations of improved server utilization, more efficient infrastructure, advanced cooling and changes to the electricity system rather than treating each measure independently. Under favorable combinations of industry and grid improvements, the researchers found substantial reductions in the projected water footprint relative to less efficient development pathways.[19] These are mitigation scenarios, not evidence that those reductions have already occurred.

Reclaimed or other appropriate non-potable water addresses a somewhat different problem. Substitution may not substantially reduce total water consumption, but it can reduce pressure on potable supplies where water-based cooling remains appropriate. As the Quincy case demonstrates, reuse can offset potable-water demand without eliminating the underlying cooling-water requirement.[26, 46] Reduction and substitution should therefore not be treated as equivalent forms of mitigation.

The greatest potential improvements occur when these measures are combined rather than treated as alternatives. More efficient hardware reduces the underlying heat and electricity requirement. Better cooling reduces the water required to remove the remaining heat. Lower-water electricity reduces the indirect footprint of the electricity still required. Reclaimed supplies can reduce pressure on potable water where water-based cooling remains appropriate. Better siting places the resulting facility where its remaining demands can be accommodated with less environmental stress.

The Nature Sustainability study illustrates the scale of this combined effect. Under its most favorable combination of industry practices, geographical allocation and electricity-system development, the modeled U.S. AI-server water footprint could be reduced by as much as 86 percent relative to the corresponding unmitigated pathway.[19] That figure is not a forecast of an 86-percent decline in actual AI water consumption. It represents the difference between modeled futures and demonstrates how much decisions about infrastructure can change the outcome.

Even the best mitigation scenario does not make AI physically impact-free. Semiconductor manufacturing still requires water. Equipment and facilities still have embodied impacts. Electricity infrastructure still requires materials and land. Some cooling systems still consume water, and reclaimed-water treatment requires energy and produces residuals. Growth can also partially or completely offset efficiency gains.[11, 26]

A percentage reduction alone does not fully describe the environmental effect of a mitigation measure. A more informative comparison asks how much water otherwise similar infrastructure would require with and without the intervention, what type of water was affected, where the change occurred and whether reducing one impact increased another.

The available research supports neither a predetermined high-water future nor the assumption that technological improvement makes scale irrelevant. Hardware, cooling, electricity, water source and geography can materially reduce the resource intensity of growth, but even a dramatically more efficient facility can contribute to increasing regional demand if enough new capacity is built. Mitigation determines the resource intensity of growth; it does not determine how much growth ultimately occurs.[19]

The mitigation scenarios help answer the question posed at the beginning of Section 10. AI's water footprint can become substantially smaller per unit of comparable activity even as the industry grows. Whether its total impact also declines depends on whether those improvements are deployed broadly and quickly enough to outpace growth in computing demand.

## 10.7 The Most Defensible Forecast

AI may become less water-intensive for comparable workloads as technology improves, while total AI-related water demand may rise if growth in deployment and computational intensity outpaces those improvements.[11, 13, 19]

The first half of that conclusion is supported by technological trends already underway. AI hardware is becoming more energy efficient, models and inference-serving systems are improving, and newer cooling technologies can substantially reduce direct water consumption.[11, 13, 41, 44, 45] These improvements make it reasonable to expect the water intensity of comparable AI computation to continue declining, although the rate of improvement remains uncertain.

The second half follows from scale. AI adoption is increasing while workloads themselves are changing. Long reasoning, agentic systems and multimodal generation can require substantially more computation than ordinary text inference.[11, 13] Consequently, falling resource requirements for a comparable task can coexist with rising aggregate electricity, cooling and water demand.

The peer-reviewed growth scenarios reinforce this conclusion. The Nature Sustainability analysis finds substantial potential U.S. AI-server water footprints under continued expansion but also finds that the outcome changes dramatically under different assumptions about efficiency, electricity and geographical deployment.[19] Its scenarios show that technological and infrastructure decisions can materially reduce the footprint without establishing that total demand will necessarily decline.

Location may determine whether these broad technological improvements translate into manageable local impacts. The Nature Sustainability analysis found substantial differences among geographical deployment scenarios, while the case studies discussed earlier demonstrate why those differences can matter in practice.[19, 33, 39-43] A nationally modest water footprint can still create a significant local problem if development becomes concentrated in a constrained watershed.

For that reason, national or global totals alone cannot determine whether AI has a “water problem.” Total consumption could increase nationally while individual facilities become less water-intensive and some watersheds experience little additional stress. Conversely, strong global efficiency improvements would not protect a particular community if intensive development occurs where water or infrastructure is already constrained.

Long-range numerical forecasts should therefore be treated cautiously. Future AI demand depends on variables that are unusually difficult to predict: adoption, model capabilities, reasoning length, agentic activity, hardware efficiency, server utilization, cooling technology, electricity generation, siting and the pace at which facilities are constructed. A projection is useful for identifying possible infrastructure requirements; it is not a predetermined water bill for the future.

The most defensible forecast is therefore conditional:

> AI will probably become substantially less water-intensive for comparable workloads. Total AI-related water demand may nevertheless increase, potentially substantially, if growth in deployment and computational intensity outpaces improvements in hardware, cooling, electricity and infrastructure. Where that growth occurs will help determine whether rising demand remains manageable or becomes a significant local water challenge.

Neither outcome is technologically predetermined. The peer-reviewed scenarios show substantial differences between better- and worse-managed development pathways.[19] Decisions about efficiency, cooling, electricity, water source and siting can materially change AI's future water footprint.

The question is therefore not simply whether AI will use more or less water. It is whether the water intensity of computing can decline faster than society expands the amount and intensity of computing, and whether the remaining infrastructure is located where local water systems can sustainably support it.

## 11. AI, Data Centers and Tampa Bay

The national debate over AI and water becomes more useful when translated into the conditions of a specific region. Tampa Bay has its own water-supply system, climate, electricity infrastructure, reclaimed-water networks, development patterns and environmental constraints. Those conditions determine whether future data-center growth would create significant local water impacts and what measures could reduce them.

Importantly, Tampa Bay is not currently Northern Virginia, Phoenix or The Dalles. The region has data centers, but it has not yet developed the concentration of large and hyperscale facilities driving many of the water and infrastructure controversies examined earlier in this report. That creates an opportunity to consider those experiences before comparable development occurs locally.

### 11.1 Tampa Bay's Existing Data-Center Landscape

Data centers already operate throughout the Tampa Bay region, supporting telecommunications, cloud services, colocation, enterprise computing and other digital infrastructure. However, the region's existing facilities are generally much smaller than the hyperscale campuses increasingly associated with large cloud and AI companies.

The Tampa Bay Regional Planning Council examined this distinction in its 2026 Data Center Debrief, developed specifically to help local governments understand the implications of potential data-center growth. TBRPC reports that the Tampa Bay region currently contains small- and medium-scale data centers but no large or hyperscale facilities.[56]

That distinction matters because the term data center covers facilities with dramatically different infrastructure requirements. A relatively small colocation or enterprise facility cannot be assumed to have the electricity, water, land or cooling requirements of a hyperscale campus containing hundreds of thousands of servers. TBRPC's regional analysis therefore distinguishes Tampa Bay's existing facilities from the much larger developments appearing elsewhere in the United States.[56]

TBRPC's accompanying Data Center StoryMap compares Tampa Bay's current landscape with hyperscale development nationally. The purpose is partly prospective: although the region does not currently host facilities at that scale, rapidly increasing demand for cloud computing and artificial intelligence is contributing to data-center expansion into new markets. TBRPC launched a dedicated Data Center Resource Hub in 2026 to provide local governments with information on water, heat, noise, land-use and policy considerations before those proposals arrive.[56]

Tampa Bay's current data-center water footprint therefore should not be inferred from the national hyperscale examples examined earlier. The existence of data centers in the region does not establish that Tampa Bay is already experiencing the water demands associated with large AI and hyperscale campuses elsewhere in the United States.

It also means that national estimates of data-center growth should not automatically be assigned to Tampa Bay. There is currently no evidence that the region will necessarily develop into a major hyperscale cluster, how many large facilities might eventually be proposed, or what companies and computing workloads those facilities would support.

The absence of existing hyperscale facilities also makes advance planning possible. TBRPC's Data Center Resource Hub takes a prospective approach, providing local governments with information on potential water, heat, noise and land-use impacts, mitigation strategies, zoning examples and policy approaches drawn from communities that have already experienced substantial development.[56] Establishing information requirements and development standards before large facilities are proposed can give the region more options than attempting to address impacts after sites, water sources and major infrastructure have already been selected.

The remainder of this section therefore does not assume that AI data centers are currently creating a major Tampa Bay water problem. The relevant question is whether future large-

scale development could create significant local impacts given the region's actual water resources and infrastructure, and what planning options are available before such development becomes established.

## 11.2 What We Know About Their Water Use

Although Tampa Bay's existing data centers can be identified, determining how much water individual facilities use is considerably more difficult. Our review of the region's existing facilities did not identify publicly available facility-level water-consumption data sufficient to calculate a reliable water footprint for Tampa Bay's current data-center industry.

That absence is important to distinguish from evidence of high water use. As established in Section 11.1, the Tampa Bay Regional Planning Council reports that the region currently has small- and medium-scale facilities but no large or hyperscale data centers.[56] The national water-demand figures associated with hyperscale and AI infrastructure therefore cannot simply be assigned to Tampa Bay's existing facilities.

Our review of public information sought facility-specific details on water sources, cooling technology, water withdrawal and consumption, wastewater discharge and Water Usage Effectiveness for existing Tampa Bay data centers. The information located was insufficient to calculate actual water consumption for the facilities reviewed.

This creates an important limitation. A data center's size or electrical capacity alone cannot reliably determine its direct water use. As established earlier in this report, facilities with similar computing loads can have very different water footprints depending on cooling technology, climate, utilization, electricity source and other operating conditions. Estimating gallons solely from building size, megawatts or a national-average WUE would therefore create a level of precision that the available local evidence does not support.

The same applies to cooling. Without verified facility-specific information, the presence of a data center does not establish that it operates evaporative cooling towers or consumes large quantities of municipal water. Facilities may use air-cooled, evaporative, hybrid or other cooling configurations, with substantially different consumptive requirements. Where the cooling architecture is unknown, this report treats it as unknown rather than assigning one by assumption.

Indirect water use is even more difficult to determine at the facility level. TBRPC notes that a substantial portion of the water associated with large and hyperscale data centers can occur through electricity generation and that this indirect component is frequently omitted from water-use estimates.[56] Calculating that footprint for an existing Tampa Bay facility

would require reliable electricity-consumption information together with an appropriate regional electricity-water intensity. The facility's address alone would not be sufficient.

Wastewater presents a similar information gap. TBRPC identifies wastewater volume and composition as considerations for water-cooled facilities and recommends greater standardized and transparent reporting of direct and indirect water use.[56] But those regional concerns do not establish the wastewater characteristics of Tampa Bay's existing facilities without facility-specific records.

The public information reviewed supports the following conclusion:

> We know that data centers currently operate in Tampa Bay. We do not currently have sufficient public facility-level evidence to determine how much water those facilities collectively withdraw or consume, what portion is potable or reclaimed, or how much indirect water should be attributed to their electricity use.

That uncertainty should remain visible rather than being filled with national averages. It also establishes a baseline against which future large or hyperscale proposals can be evaluated if more detailed facility-level information becomes available.

## 11.3 What We Don't Know

Beyond the information gaps described in Section 11.2, two unresolved questions are particularly relevant to this report: when facilities place demand on regional water systems, and how much of their activity supports AI.

The public information reviewed was insufficient to establish peak or seasonal water demand for existing Tampa Bay facilities. Even if annual totals were available, they would not show whether cooling demand coincides with drought, high community demand or other periods of pressure on regional supplies.

The review also could not reliably separate AI from non-AI computing. Facilities may support telecommunications, storage, enterprise applications, cloud services and AI workloads simultaneously. A facility's total water consumption would therefore not, by itself, establish the portion attributable to AI. The information identified does not support calculating a precise AI water footprint for Tampa Bay.

These gaps matter for future planning. TBRPC identifies access to water-use information and standardized reporting of direct and indirect consumption as policy considerations for large and hyperscale development.[56] Evaluating future proposals would require information on expected operating loads, seasonal demand, cooling, water sources, discharge and

electricity use, together with a clear explanation of what any reported AI-specific estimate includes.

## 11.4 Tampa Bay's Water System

Evaluating future data-center development in Tampa Bay requires understanding the water system into which that demand would be introduced. The region does not rely on a single reservoir, river or aquifer. Instead, drinking water is supplied through a regional network combining groundwater, surface water and desalinated seawater, while local utilities also operate extensive reclaimed-water systems for non-potable uses.[57, 58]

Tampa Bay Water is the regional wholesale drinking-water supplier for Hillsborough, Pasco and Pinellas counties and the cities of Tampa, St. Petersburg and New Port Richey. It supplies water to those member governments, whose local utilities then distribute it to homes and businesses. Tampa Bay Water reports that its member governments' service areas serve more than 2.6 million residents across the region.[57]

The regional system is unusual because it deliberately combines three different drinking-water sources. Groundwater is withdrawn from the Floridan Aquifer through regional wellfields; surface water is collected from the Hillsborough and Alafia rivers and Tampa Bypass Canal when sufficient flows are available; and seawater is treated at the Tampa Bay Seawater Desalination Plant. The C.W. Bill Young Regional Reservoir, with approximately 15.5 billion gallons of storage capacity, allows surface water captured during wetter periods to be stored for later use.[57]

The proportions supplied by those sources are not fixed. Tampa Bay Water adjusts production according to rainfall, river flows, groundwater conditions, reservoir storage, demand and other operational constraints. Its Optimized Regional Operations Plan uses monitoring and forecasting to shift production among available supplies while attempting to meet demand and limit environmental impacts.[57]

That flexibility is important because Tampa Bay's water availability is seasonal and climate-dependent. Surface-water availability depends heavily on rainfall and river flows. During dry periods, less river water may be available for treatment or reservoir replenishment, increasing reliance on groundwater, stored surface water and desalinated seawater.

The 2026 drought illustrates how these supply constraints affect regional operations. Section 11.5 examines the shortage declarations, rainfall deficits and changes in reservoir storage.

The regional supply mix changed accordingly. In August 2026, Tampa Bay Water reported average production of approximately 144.2 million gallons per day from groundwater, 30.0 MGD from surface water and 8.8 MGD from desalinated seawater. Groundwater therefore supplied approximately 79 percent of regional production that month. Fiscal-year-to-date figures relied less heavily on groundwater, with approximately 70 percent groundwater, 24 percent surface water and 6 percent desalination, illustrating how strongly the mix can change with conditions.[57]

Groundwater itself is not unlimited. Tampa Bay Water is permitted to withdraw an annual average of 120 MGD from 13 regional wellfields, and the utility notes that historically high pumping rates contributed to declining water levels in some lakes and wetlands.[57] Regional operations now rotate production among available sources rather than relying on unrestricted increases in groundwater pumping.

This history is important when considering large new industrial users. The Southwest Florida Water Management District's 2025 Regional Water Supply Plan for the Tampa Bay Planning Region states that sufficient alternative water sources exist to meet projected future demands while replacing some existing fresh-groundwater withdrawals that contribute to hydrologic stress. The plan therefore emphasizes a combination of alternative supplies, conservation and carefully managed groundwater rather than assuming unlimited expansion of traditional fresh-groundwater withdrawals.[58]

Reclaimed water is another major part of the regional picture, but it should not be confused with Tampa Bay Water's wholesale drinking-water supply. Reclaimed water is highly treated wastewater distributed by local utilities and other reuse systems for uses including landscape irrigation, industrial applications, groundwater recharge and environmental enhancement. SWFWMD identifies reclaimed water as an important alternative water supply and reports that more than half of wastewater-treatment flows districtwide have been beneficially reused for non-potable purposes in recent years.[58]

That distinction could become particularly important for data centers. A future facility requiring water for cooling would not necessarily need drinking-quality water if appropriate reclaimed supply were available. But the existence of reclaimed water somewhere in Tampa Bay does not mean that an unlimited supply is available at every potential data-center site. Distribution pipelines, treatment capacity, existing customers, seasonal demand and competing future uses determine whether reclaimed water can actually serve a particular project.[58]

SWFWMD provides the broader regulatory and planning framework. Its Tampa Bay Planning Region consists of Hillsborough, Pasco and Pinellas counties, and its 2025 Regional Water

Supply Plan evaluates expected demand and potential supplies through 2045.[58] Local governments and utilities remain responsible for their own systems and development decisions, while the District evaluates regional water resources, issues water-use permits where applicable, establishes water-resource protections and helps fund alternative-water and conservation projects.[58]

The region therefore does not face a simple condition of either having or not having enough water. Tampa Bay has a diversified and actively managed supply system, but that system operates within hydrological and infrastructure constraints. Fresh groundwater has environmental limits, surface-water availability varies with rainfall and river flows, desalination provides a more drought-resistant source but only part of regional supply, and reclaimed water may face competing uses.[57, 58]

The 2026 shortage makes the distinction particularly relevant to this report. A proposed data center's annual water demand might appear modest relative to total regional drinking-water production. But an annual average alone would not establish whether that demand is easily accommodated. Its source, location, peak demand and behavior during drought could matter more than the annual number by itself.

For Tampa Bay, the appropriate question is therefore not simply, “How many gallons would a data center use?” It is which part of the regional water system would supply those gallons, whether potable-quality water is necessary, what other demands compete for that source and what happens to the facility's demand when rainfall and river flows are constrained.

## 11.5 Current Water Conditions

Tampa Bay is considering the implications of future data-center development during a period of significant regional water-supply stress. As of September 2026, the region remains under water-shortage conditions following an extended rainfall deficit and drought. These conditions were not caused by data centers; Tampa Bay currently has no large or hyperscale facilities, as established in Section 11.1.[56]

Tampa Bay Water entered a Stage 3 Extreme Regional Supply Shortage on March 1, 2026, after persistent rainfall and river-flow deficits reduced available surface-water supplies and threatened storage in the C.W. Bill Young Regional Reservoir. At the time, the agency described the situation as one of the region's most severe water-supply shortages in approximately 50 years.[76]

On September 22, 2026, the Southwest Florida Water Management District's Governing Board extended its Modified Phase III "Extreme" Water Shortage restrictions through March 31, 2027, citing ongoing drought conditions. The restrictions apply to Hillsborough, Pasco and Pinellas counties, along with other areas of the District, and generally limit landscape irrigation to one day per week with restricted watering hours.[74]

The cause of the shortage is well documented. The region experienced prolonged below-normal rainfall that reduced river flows, lowered groundwater and surface-water levels and limited the amount of surface water available for regional drinking-water production. As of August 2026, Tampa Bay Water reported a 12-month rainfall deficit of 14.1 inches and an average streamflow deficit of approximately 28 million gallons per day.[75]

Conditions improved during the summer. August rainfall averaged approximately 9.0 inches, 1.1 inches above normal, helping reduce demand and replenish regional storage. The C.W. Bill Young Regional Reservoir increased from approximately 15 percent of capacity at the end of July to 31 percent at the end of August.[75]

Reservoir storage continued to recover in September. The Southwest Florida Water Management District reported that the C.W. Bill Young Regional Reservoir held approximately 7.87 billion gallons on September 20, 2026, about half of its 15.5-billion-gallon capacity.[74, 75]

Despite improving reservoir storage, Tampa Bay Water reported that the region remained in a Stage 3 Regional Supply Shortage in its September 4 update, and the District's September 22 announcement cited ongoing drought conditions.[74, 75] For future data-center planning, these conditions demonstrate why water-supply assessments must account for prolonged drought as well as annual-average availability.

A future facility might operate for decades, experiencing both wet years and droughts. Planning should establish how its demand would be supplied during extended shortages, what operating adjustments would be required and whether additional infrastructure would be needed.

Peak demand also matters. Hot, dry conditions can simultaneously increase regional water demand and raise the cooling requirements of some data centers. A proposed facility's annual average water consumption should therefore be considered alongside maximum daily demand, seasonal demand and expected operation during declared shortages or drought emergencies.

Water source and cooling technology could change that calculation. A facility capable of using reclaimed water or operating with low-water cooling would interact differently with a drought-constrained potable-water system than one dependent on drinking-quality water

for evaporative cooling. But reclaimed-water availability and reliability would themselves need to be evaluated under local conditions rather than assumed.

For future large or hyperscale development, this creates a drought-specific planning question: What happens to a facility's water demand when Tampa Bay enters Stage 2 or Stage 3 shortage conditions?

## 11.6 What Hyperscale Development Could Change

Tampa Bay's existing data-center landscape provides a poor proxy for what future hyperscale development could mean for the region. As established in Section 11.1, TBRPC identifies existing facilities as small or medium scale and reports no current large or hyperscale data centers in the region.[56] The facilities identified in our review are generally in the single-digit-megawatt range, including facilities around 4–9 MW.

The scale now being discussed for new AI infrastructure elsewhere in the United States is fundamentally different. The U.S. Department of Energy has reported grid-interconnection requests for hyperscale facilities in the 300–1,000 MW range or larger, driven in significant part by expanding AI demand. A future 100 MW facility would therefore not simply represent a somewhat larger version of a 5 MW Tampa Bay data center. It could represent roughly twenty times the electrical capacity, while a 500 MW campus could represent roughly one hundred times that scale.

Those ratios should not be converted directly into water-use ratios. A 500 MW facility does not necessarily consume one hundred times as much water as a 5 MW facility because cooling technology, utilization, climate and electricity supply can differ substantially. But the comparison demonstrates why the scale of the infrastructure changes the planning question.

Electricity is the most obvious difference. A 100–500+ MW campus represents a major new continuous load on the electrical system. DOE has identified hyperscale AI data centers as one of the sources of rapidly accelerating U.S. load growth,[13] while Berkeley Lab's latest national modeling projects data centers could account for roughly 9.5 to 15.3 percent of U.S. electricity consumption by 2030 under its range of scenarios.[16]

Water demand can scale alongside that computing capacity when facilities use water-intensive cooling. The precise amount cannot be known from megawatts alone, but a hyperscale campus operating evaporative cooling could introduce a municipal or regional water demand unlike anything represented by Tampa Bay's current small and medium

facilities. The significance would depend on cooling technology, WUE, operating load, water source and seasonal conditions.

A hypothetical example illustrates the scale without predicting the design of any future Tampa Bay project. Using an illustrative WUE of 0.5 L/kWh,[66] consistent with UNEP's recommended 2029 water-efficiency benchmark for data centers, a 100 MW IT load operating continuously at that intensity would correspond to roughly 1.2 million liters, or about 317,000 gallons, of direct water consumption per day. At 500 MW, the same assumptions would produce roughly 1.6 million gallons per day. Those are illustrative calculations, not forecasts: a dry-cooled facility could use dramatically less direct water, while a different evaporative system could use more.

Scale also increases the importance of peak and drought-period demand. As Section 11.5 demonstrates, a future project's maximum requirements during water-shortage conditions may be more consequential than its annual average alone.

At hyperscale, reclaimed-water availability also becomes an infrastructure question. A large campus could require enough supply to warrant new transmission pipelines, storage, pumping or treatment capacity, while existing customers and planned future uses could limit the amount actually available to a new industrial user.

Wastewater requirements can scale as well. Evaporative systems produce blowdown as dissolved minerals become concentrated, meaning a much larger facility can change both the quantity and characteristics of wastewater that local infrastructure must accommodate.

Land use and heat can similarly become different questions at hyperscale. A campus can contain multiple data-center buildings, substations, backup-generation equipment, cooling infrastructure and electrical transmission facilities. TBRPC specifically developed its 2026 resource hub to help Tampa Bay governments consider the water, heat, noise and land-use implications of large and hyperscale facilities, rather than assuming experience with existing smaller facilities adequately represents them.[56]

Perhaps most importantly, hyperscale development can change the issue from facility planning to regional planning. Even if one 100 MW facility proved supportable, several 100–500 MW campuses proposed within the same utility territory could produce a substantially different cumulative demand for electricity, potable or reclaimed water, wastewater capacity and supporting infrastructure. Northern Virginia demonstrates why evaluating each facility independently can miss the consequences of an emerging regional cluster.[33]

This does not establish that Tampa Bay will receive 500 MW data-center campuses, much less that such development is inevitable. TBRPC's current work is prospective, and the

region presently lacks large and hyperscale facilities.[56] The purpose of examining hyperscale development is therefore planning, not prediction.

For water planning, the key question changes accordingly. It is no longer simply whether Tampa Bay can provide water to a data center. It becomes:

> How much large-scale data-center development could Tampa Bay accommodate, using which water sources and cooling technologies, under normal and drought conditions, before cumulative demand requires additional infrastructure or places significant additional pressure on the regional water system?

That is a fundamentally different question from the one posed by the region's existing facilities.

## 11.7 Florida's New Data-Center Water Requirements

Tampa Bay's consideration of future data-center growth is occurring just as Florida has created a new statewide permitting framework specifically for the water demands of large-scale data centers. Effective July 1, 2026, Florida law establishes additional requirements for large-scale data centers seeking allocations of water and explicitly directs regulators to consider reclaimed water where it is a feasible alternative.[59]

The legislation, enacted as Chapter 2026-65, Laws of Florida, created section 373.262 of the Florida Statutes. The Legislature states that large-scale data-center development should be managed so that Florida's water resources are used in the public interest, are not harmed by the proposed use and remain consistent with local zoning regulations and comprehensive plans.[59]

The law does not prohibit large-scale data centers from using Florida water. Instead, it subjects their proposed allocations to the state's existing core standards for consumptive use. A water management district or the Florida Department of Environmental Protection may not issue the allocation if the proposed use would harm the area's water resources. To receive a permit, the applicant must establish that its proposed water use is reasonable-beneficial, will not interfere with existing legal uses of water, and is consistent with the public interest.[59] These are established principles of Florida water law that the new statute expressly applies to large-scale data centers.

Of particular relevance to Tampa Bay, the law creates a specific requirement concerning reclaimed water. Regulators must require reclaimed water to replace all or part of a proposed surface-water or groundwater allocation when several conditions are satisfied: a suitable permitted reclaimed-water source must be available; adequate distribution or

supply lines must reach the property's boundary; the applicant must be capable of accessing them; reclaimed-water use must be environmentally, economically and technically feasible; and its use cannot conflict with applicable surface-water discharge requirements.[59]

That provision is stronger than simply encouraging developers to consider reclaimed water. When the statutory conditions are satisfied, the governing board or department must require reclaimed water in lieu of all or a portion of the proposed surface-water or groundwater use.[59] At the same time, those conditions prevent the law from assuming that reclaimed water is available everywhere. The statute requires, among other things, that suitable reclaimed water be available and permitted and that distribution or supply lines with sufficient capacity and quality be available at the property boundary.[59]

This distinction is particularly relevant to the regional system described in Section 11.4. Tampa Bay has extensive reclaimed-water resources, but availability varies among utilities and locations. Florida's new framework makes the existence, capacity and accessibility of that infrastructure a material consideration when a large-scale data center seeks a water allocation.

The law imposes additional requirements once a large-scale data center requests an allocation averaging at least 100,000 gallons per day. Its permit application must disclose all sources and amounts of water, as well as water losses, associated with cooling, industrial and treatment processes, employee sanitary needs and landscape irrigation.[59]

The applicant must also submit a water-conservation plan. At minimum, that plan must provide for recycling cooling water before discharge or disposal, a leak-detection and repair program, water-efficient fixtures and employee water-conservation awareness and education.[59]

Those requirements address several of the information gaps identified in Sections 11.2 and 11.3. For existing Tampa Bay facilities, our review generally could not identify facility-level information sufficient to establish water sources, consumption or cooling losses. For a future large-scale facility seeking at least 100,000 gallons per day through this permitting framework, substantially more information about its proposed water use would have to be provided to regulators.

The law also creates a procedural safeguard: a large-scale data-center water permit cannot be approved without a hearing.[59] In addition, a later modification of a consumptive-use permit proposed by a large-scale data center must be treated in the same manner as an initial permit application rather than through the simplified modification process otherwise available under certain circumstances.[59]

These provisions are important for transparency, but they should not be overstated. Requiring information in a permit application does not necessarily mean that every operational water metric will subsequently appear in a standardized public database or corporate sustainability report. Nor does the law create a universal maximum number of gallons that a Florida data center may consume. It creates a permitting and review framework rather than a fixed statewide water-use cap.

The 100,000-gallon threshold also deserves careful wording. It triggers the statute's additional application disclosures and conservation-plan requirements for large-scale data centers; it should not be described as a legal limit of 100,000 gallons per day.[59] A project may seek a larger allocation if it can satisfy the applicable permitting requirements.

The framework does not eliminate every transparency question identified in Sections 11.2 and 11.3. Projected permit information is different from standardized public reporting of actual operational performance after a facility begins operating. A fuller picture could include expected and actual annual water withdrawal and consumption; peak and seasonal demand; potable, reclaimed and other supply sources; cooling technology and WUE; wastewater volume and characteristics; electricity demand; drought-period operations; and subsequent measured performance.

For Tampa Bay, the timing is notable. TBRPC reports that the region currently has no large or hyperscale facilities, meaning Florida established these requirements before comparable large-scale development became established locally.[56]

Several provisions address factors identified earlier in this report as relevant to data-center water impacts, including water source, reclaimed-water availability, cooling-water recycling, existing legal uses and the effects of proposed withdrawals on water resources.[59] The statute therefore incorporates several variables that would otherwise need to be considered independently when evaluating a future large-scale project.

Florida's new framework therefore does not answer whether a particular future Tampa Bay data center would be environmentally appropriate. That determination would still depend on its location, size, cooling technology, water source, electricity requirements and local conditions. But it changes the regulatory baseline under which such a proposal would be evaluated.

For Tampa Bay, the practical question is no longer merely whether local governments should ask a future large-scale data center where its cooling water will come from. For projects seeking significant water allocations, Florida law now requires many of those questions to be addressed through the permitting process.[59]

## 11.8 Hillsborough County's Response

Tampa Bay's data-center debate is no longer entirely hypothetical. In August 2026, the Hillsborough County Board of County Commissioners directed staff to draft a proposed moratorium ordinance on AI data-center development in unincorporated Hillsborough County and to evaluate a broader land-use regulatory framework for future facilities. The Board also directed staff to evaluate existing policies governing water demand from water-intensive development and potential water-conservation and capacity-management strategies.[60]

The Board directed staff to draft an ordinance that would impose a moratorium on the development of AI data centers in unincorporated Hillsborough County. The proposed moratorium was structured to take effect after October 1, 2027, or upon expiration of the restrictions associated with Florida Senate Bill 180, whichever allowed the County to proceed.[60]

That timing is important. The commissioners' action should not be described as an AI data-center ban currently in effect. As of September 2026, the Board has directed preparation of a future ordinance and regulatory framework; the County's own August meeting record describes these measures as proposals and staff directions rather than an existing prohibition.[60]

Water was explicitly part of the County's concern. Commissioners directed staff to evaluate existing policies governing water demand from water-intensive development and identify legally available strategies for water conservation and utility-capacity management.[60]

The Board also directed staff to evaluate land-use regulations for future AI data centers, including the possibility of limiting or prohibiting the use of County potable water for cooling.[60] Such a restriction would address competition for drinking-quality water rather than necessarily eliminating data-center water use.

Depending on the final regulations and available infrastructure, a facility could instead rely on reclaimed water, another approved non-potable source or low-water cooling. The resulting water impacts would depend on the alternative used.

Hillsborough County’s data-center planning takes place alongside substantial investment in its existing water and wastewater systems. In September 2026, the County listed 58 Water Resources capital improvement projects totaling approximately $2.44 billion, including approximately $1.5 billion for its One Water Program. The investments address infrastructure, maintenance, resiliency, regulatory requirements and service needs across the County.[82]

Those investments are not evidence that data centers caused Hillsborough County's infrastructure needs. They primarily respond to existing and projected regional growth and utility requirements. Their relevance is instead that any future hyperscale campus would enter a county already making substantial long-term investments to expand and maintain water and wastewater capacity.

The County's deliberations are also occurring during the regional water shortage described in Section 11.5. Although that shortage is unrelated to hyperscale AI development, it provides immediate context for evaluating how large new industrial demands would operate during constrained supply conditions.[58]

Hillsborough County's approach is therefore notably prospective rather than reactive. There is no existing hyperscale AI campus whose water demand the County is attempting to correct. Instead, officials are considering what rules should apply before facilities potentially hundreds of megawatts in scale arrive.

The County's response also complements, rather than replaces, Florida's new statewide requirements discussed in Section 11.7. State law governs consumptive-use permitting and can require reclaimed water when specified conditions are satisfied.[59] Hillsborough County is considering additional questions within its own authority, including land use, utility capacity and whether County potable water should be available for data-center cooling at all.[60]

The eventual policy remains unsettled. A moratorium could change before adoption, and the details of future land-use or potable-water restrictions have not yet been finalized. This report therefore should not evaluate Hillsborough County as though it has already adopted a permanent data-center policy.

What can be said is that one of Tampa Bay's largest local governments has begun treating potential hyperscale AI infrastructure as a water-planning issue before such development has become established locally.[60]

The relevant policy question is therefore not whether Hillsborough County should treat data centers as inherently good or bad. It is what information, water-source, conservation and infrastructure requirements are appropriate for evaluating individual proposals before their long-term demands become part of the regional system.

## 11.9 TBRPC's Regional Planning Work

Tampa Bay is not approaching potential large-scale data-center development without regional planning. In August 2026, the Tampa Bay Regional Planning Council (TBRPC)

published Data Center Debrief: Information & Considerations for the Tampa Bay Region, a white paper developed specifically to help its member governments evaluate the potential impacts and opportunities associated with large and hyperscale data centers.[56]

TBRPC reports that Tampa Bay currently has no hyperscale data centers.[56] Its work is therefore prospective, providing local governments with a planning framework before development at that scale becomes established in the region.

TBRPC explicitly describes this as an effort to help governments remain “proactive rather than reactive.” The Council notes that local governments are being asked to make rapid decisions about an evolving industry with limited resources and that examples from other parts of the country can provide a foundation for considering what may occur locally. At the same time, it cautions that Tampa Bay has its own natural and physical characteristics and may require locally tailored strategies.

The Council's work extends beyond water. Its white paper evaluates three principal categories of potential community impact from large and hyperscale facilities: noise, water and heat. It also considers mitigation strategies, regulatory approaches and potential economic benefits. TBRPC's purpose is therefore broader than the AI-water question examined in this report.

Water nevertheless receives substantial attention. TBRPC considers both direct water use at facilities and indirect water associated with electricity generation, along with tradeoffs among cooling approaches, electricity demand and wastewater generation. Its mitigation discussion includes closed-loop cooling, reclaimed and other non-potable supplies, lower-water electricity sources and emerging cooling technologies without presenting any single approach as universally preferable.[56]

TBRPC also identifies access to information as a planning problem. Its white paper recommends standardized and transparent reporting of direct and indirect water use so regulators and communities can evaluate individual facilities and cumulative impacts. It argues that making standardized information public could also help other jurisdictions evaluate future proposals and give operators information useful for reducing water costs.

That recommendation is directly relevant to the information gaps identified in Sections 11.2 and 11.3, where publicly available facility-level data proved insufficient to characterize the water use of Tampa Bay's existing data centers.

The Council's emphasis on cumulative impacts is particularly important. Water availability, heat and noise do not stop at city or county boundaries, and TBRPC warns that strategies developed independently by individual jurisdictions can become disjointed. It therefore

calls for regional collaboration and resource sharing rather than treating every proposed facility solely as an isolated local land-use decision.

That regional perspective is well suited to Tampa Bay's water system. As Section 11.4 established, drinking-water supply already crosses jurisdictional boundaries through Tampa Bay Water, while SWFWMD manages water resources at an even larger regional scale. A large facility physically located in one county can therefore create water, electricity or environmental implications that extend beyond the jurisdiction issuing its development approvals.

TBRPC also considers how communities might capture economic benefits while managing those costs. The white paper discusses Community Benefits Agreements (CBAs) as one possible tool through which developers, governments and communities can establish commitments involving utility use, environmental impacts, workforce development, financial contributions and accountability. It notes that Tampa and St. Petersburg have previously used CBAs for major development projects, although not specifically for data centers.

The report raises a related transparency concern about non-disclosure and confidentiality agreements between governments and data-center developers. TBRPC warns that these arrangements can restrict information sharing, contribute to inaccurate impact estimates, undermine community trust and weaken the ability of communities to negotiate effectively.

TBRPC also recognizes potential economic benefits from data-center development, including investment, property-tax revenue, infrastructure improvements and workforce development.[56] Its analysis considers those potential benefits alongside the infrastructure and environmental costs that local governments may need to assess and mitigate.

The Council also makes clear that its 2026 white paper is a starting point rather than a completed regional data-center policy. TBRPC describes the document as the beginning of a regional conversation and invites member governments to contribute additional research and considerations so that the Council can identify opportunities and facilitate stronger collaboration.

That distinction matters. Tampa Bay does not yet have a single regional answer to hyperscale development. Hillsborough County is developing its own response, Florida has established new statewide water-permitting requirements, and regional water agencies already operate under their own planning and regulatory responsibilities. TBRPC's role is to help connect those conversations across jurisdictions.

For the water question specifically, TBRPC's work suggests that decisions about future large-scale facilities should not be made solely within individual jurisdictional boundaries.

Water supply, cumulative demand and supporting infrastructure can have regional implications that require coordination among local governments, utilities and water-management agencies.

TBRPC's work therefore adds a regional layer to the state and local responses described in Sections 11.7 and 11.8. Tampa Bay has begun developing a framework for evaluating large and hyperscale data centers before those facilities become an established regional industry.[56]

## 11.10 Florida's Forthcoming Independent Study

This report captures Tampa Bay and Florida's data-center landscape at a moment when the state's own comprehensive assessment is not yet complete. Under legislation enacted in 2026, Florida directed the Office of Program Policy Analysis and Government Accountability (OPPAGA) to contract for an independent, interdisciplinary study of large-scale data centers. The completed study must be submitted to the Governor, President of the Senate and Speaker of the House by July 1, 2027.[59]

For purposes of these provisions, Florida law defines a large-scale data center as a single location with an anticipated monthly peak electrical load of 50 MW or more.[59] The forthcoming study will therefore examine an infrastructure category substantially larger than the small- and medium-scale facilities currently identified in Tampa Bay.

The required scope extends beyond water. The Legislature directed the study to examine the construction and operation of large-scale data centers and their effects on state, regional and local economic development and tax revenue; land, water and other natural resources; energy use and associated costs and rates; and public health and safety.[59]

The study must also identify issues that may be unique to large-scale data-center development in Florida and provide recommendations concerning facility siting and mitigation measures that could reduce potential negative impacts. OPPAGA is authorized to contract with one or more nonpartisan academic or nonprofit research organizations with relevant policy and scientific expertise.[59]

The Florida-specific focus is important because much of the existing evidence on data-center impacts comes from national research or development in other states. Those sources establish useful physical and planning principles, but they cannot fully determine how large-scale development would interact with Florida's climate, water resources, reclaimed-water infrastructure, electricity system, growth patterns and regulatory framework.

As of September 2026, OPPAGA's study has not been completed. OPPAGA's contracted-reviews page shows that the office has been soliciting proposals from experienced consultants for the independent interdisciplinary study required by Chapter 2026-65.

This creates an important boundary for the present report. The analysis here is a September 2026 snapshot. It incorporates the evidence publicly available at this time, including peer-reviewed research, federal and state sources, corporate disclosures, Tampa Bay's existing facility review, TBRPC's regional planning work, current water conditions and Florida's newly enacted regulatory framework. However, it cannot incorporate findings from a state study that will not be due until July 2027.

The study is being conducted while state, regional and local responses are already developing. Florida's new water-permitting requirements are in effect, TBRPC has begun regional planning, and Hillsborough County is considering its own regulatory response.

The forthcoming study may provide Florida-specific evidence concerning water requirements, energy demand, siting, economic effects, infrastructure costs and mitigation options that is not currently publicly available. It may also identify constraints or conditions not captured by national research or show that particular impacts differ under Florida conditions.

Conclusions in this report concerning future Florida and Tampa Bay development should therefore be treated as current assessments rather than permanent forecasts. Where Florida-specific evidence remains unavailable, uncertainty should remain explicit rather than being filled with assumptions about what the forthcoming study will conclude.

Accordingly, this report should be read as an assessment based on evidence available as of September 2026, not the final word on data centers and water in Florida. After its scheduled completion in 2027, the OPPAGA study can serve as a benchmark for revisiting these conclusions and incorporating new Florida-specific evidence into future revisions.

## 12. What Could Tampa Bay Do?

The evidence reviewed in this report does not point toward a single policy response such as banning data centers or assuming that technological improvement will resolve every concern. Instead, the available options involve different combinations of water management, infrastructure planning, disclosure and project-specific review, each with its own benefits and tradeoffs.

Tampa Bay has an opportunity to consider these options before large or hyperscale development becomes established in the region. TBRPC reports that the region does not currently contain large or hyperscale data centers, Florida has already adopted new water-permitting requirements for large-scale facilities, and Hillsborough County has begun examining additional local safeguards.[56, 59, 60] This creates an opportunity to establish expectations before major new water, electricity and wastewater demands become embedded in regional infrastructure.

The strongest solutions are unlikely to come from one requirement alone. Water source, cooling technology, electricity supply, siting, drought resilience, disclosure and cumulative regional capacity interact with one another. A project using low-water cooling but requiring extensive new fossil generation may solve one problem while worsening another. A facility using reclaimed water may reduce competition for potable supplies while still requiring substantial treatment and wastewater management. Evaluating these options therefore requires considering the system as a whole rather than optimizing a single metric.

The relevant planning questions are which impacts can be avoided or reduced, what tradeoffs accompany those reductions, and how infrastructure costs associated with new demand would be allocated among developers, utilities and the broader community.

The following options are therefore presented as directions for further policy and infrastructure planning, not as a claim that every measure should be imposed identically on every facility.

## 12.1 Prioritize Reclaimed Water

If large data centers are developed in Tampa Bay and require water for cooling, reclaimed water could reduce reliance on potable supplies where its use is technically, environmentally and economically feasible. Unlike low-water cooling, reclaimed-water use does not necessarily reduce total water consumption; its principal benefit is substituting an appropriate non-potable source for drinking-quality water.[58, 59]

Tampa Bay already has extensive reclaimed-water infrastructure, and SWFWMD identifies reclaimed water as an important alternative supply for reducing pressure on traditional groundwater and surface-water resources.[58] Florida's new data-center legislation also incorporates this principle. Under section 373.262, regulators must require reclaimed water to replace all or part of a proposed surface-water or groundwater allocation for a large-scale data center when the statutory conditions concerning availability, infrastructure, access and feasibility are satisfied.[59]

The regional planning question therefore extends beyond whether reclaimed water exists somewhere in Tampa Bay. A future large facility would require adequate supply at its specific location, and availability depends on treatment capacity, pipeline access, storage, existing customers and competing future uses. Serving a hyperscale campus could also require new transmission, pumping, storage or treatment infrastructure.

Considering those conditions during siting could help identify locations where reclaimed water can be supplied without extensive new infrastructure or displacement of existing uses. Where new infrastructure would primarily serve a major data-center development, local governments and utilities could also determine how associated capital and operating costs should be allocated between the developer and existing utility customers.

Reliability would need to be considered as well. A facility relying on reclaimed water would still require an operating plan for supply interruptions, peak conditions and periods when other reclaimed-water demands are high. Any potable backup supply could therefore be included when evaluating the facility's maximum and drought-period demand rather than considering only its normal reclaimed-water use.

Reclaimed water also has competing beneficial uses, including irrigation, industrial applications, groundwater recharge and environmental projects. Its use by a data center should therefore be evaluated against what would otherwise happen to that supply. Reclaimed water that would otherwise have limited beneficial use presents a different opportunity from water already committed to another customer or purpose.

Water efficiency remains relevant regardless of source. Reducing cooling demand can preserve reclaimed-water capacity while also reducing treatment, pumping and wastewater requirements. Depending on cooling design and reclaimed-water quality, additional treatment and management of concentrated cooling-water blowdown may also be necessary.[26]

For future large-scale development, Tampa Bay could therefore evaluate reclaimed-water availability during siting; consider normal, peak and backup supply requirements; account for competing uses and infrastructure costs; and coordinate those decisions with Florida's new consumptive-use permitting requirements.[59]

## 12.2 Reduce Routine Potable-Water Cooling

Even where reclaimed water is unavailable or unsuitable, Tampa Bay could reduce the pressure that future data centers place on drinking-water supplies by discouraging routine

dependence on potable water for cooling when lower-water alternatives are technically and economically feasible.[56, 59]

The distinction between routine and backup use is important. A facility designed to operate primarily with low-water cooling but capable of using limited potable water under unusual conditions presents a different demand from one that continuously evaporates drinking-quality water during normal operation. Policy could distinguish between those designs rather than treating any connection to the potable system as equivalent.

Closed-loop, dry and hybrid cooling systems provide several ways to reduce routine water consumption, although their performance and electricity requirements vary with climate and facility design.[41, 45] Reducing direct water use can also increase electricity demand in some configurations, so potable-water savings should be evaluated alongside the energy consequences of the alternative.[5, 6]

Hillsborough County is already considering this issue. In August 2026, county commissioners directed staff to examine land-use regulations that could limit or prohibit the use of County potable water for cooling AI data centers.[60] The proposal has not become a final prohibition, but it makes potable-water cooling an active local policy question.

A performance-based approach could focus on the resulting water demand rather than prescribing a particular cooling technology. Expected annual and maximum daily potable-water consumption, cooling configuration, WUE and demand during extreme heat or drought could provide a basis for comparing alternative designs. This would allow reclaimed-water systems, closed-loop cooling, dry cooling, hybrid systems or future technologies to be evaluated according to their actual resource requirements.

Peak and backup demand deserve particular attention. A facility with low annual potable-water consumption could still require substantially more water during extreme heat, supply interruptions or other operating conditions. Those requirements could coincide with periods when Tampa Bay's regional water system is already under greater stress, as demonstrated by the 2026 shortage.[57, 58]

Reducing routine potable-water use could also limit additional demand on Tampa Bay Water's regional supply portfolio. Because drinking water may be supplied from groundwater, surface water or desalinated seawater depending on operating conditions, reducing potable cooling demand avoids adding that requirement regardless of which source is supplying the system at a particular time.[57]

This does not require categorically excluding potable water from data centers. Water quality, reliability, infrastructure and cooling design may make some potable use appropriate. The planning question is whether substantial routine use is necessary when feasible alternatives exist.

For future large-scale development, Tampa Bay could therefore evaluate routine, peak and backup potable-water requirements during project review; compare feasible lower-water alternatives and their energy tradeoffs; and consider whether proposed potable demand remains appropriate during drought and other constrained conditions.

## 12.3 Encourage Low-Water Cooling Technologies

Reducing water consumption does not require eliminating liquid cooling. As discussed earlier, the water implications of a cooling system depend not simply on whether liquid is used near computing equipment, but on how the captured heat is ultimately rejected. Closed-loop direct-to-chip cooling, immersion cooling, dry heat rejection and hybrid systems can all potentially reduce routine consumptive water use, depending on facility design and operating conditions.[41, 45]

For Tampa Bay, a technology-neutral approach could focus on water performance rather than prescribing a particular cooling system. This would allow developers to use existing or future technologies capable of meeting the same resource objectives without requiring regulations or planning standards to favor one engineering solution.

The water-energy tradeoff should remain part of that evaluation. Cooling systems that substantially reduce direct water consumption can require additional electricity under some conditions, particularly in hot climates.[5, 6] A proposed system could therefore be evaluated using both its direct water requirements and the energy consequences of achieving those reductions.

Tampa Bay's climate also makes peak performance relevant. A system with low annual water consumption may operate differently during periods of extreme heat, while hybrid systems may increase evaporative cooling under particular temperature or operating conditions. Expected annual consumption alone may therefore provide an incomplete picture of how a facility would interact with regional water supplies.

Performance-based review could include expected annual and peak water consumption, potable and reclaimed-water demand, WUE, electricity requirements and performance during drought and extreme heat. Where multiple cooling alternatives are feasible, those

measurements would provide a common basis for comparing their water, energy and reliability tradeoffs.

Operational measurements could subsequently be compared with design projections where water-use commitments form part of development approvals, utility agreements or incentives. This would help determine whether anticipated water savings are being achieved after a facility begins operating and would contribute additional facility-level information to the regional evidence base.

Scale should remain part of the assessment. Low-water cooling can substantially reduce the resource intensity of a facility without eliminating its electricity, land or infrastructure requirements, and cumulative development can still create regional impacts even when individual facilities are comparatively efficient.

For future large-scale development, Tampa Bay could therefore encourage cooling systems that minimize routine water consumption while allowing developers flexibility in how that outcome is achieved. Evaluating water use, electricity demand, peak conditions and measured performance together would allow cooling technologies to evolve without treating any single design available in 2026 as the permanent solution.

## 12.4 Require Facility-Level Water Reporting

One of the clearest opportunities identified by this report is improving facility-level water information. Tampa Bay's existing facility review found that data centers are present in the region, but publicly available information generally does not allow their water consumption, sources or cooling systems to be evaluated at the individual-facility level.

That information becomes more important as facilities become larger. Regional totals and corporate sustainability reports cannot necessarily show whether a particular facility represents a significant demand on a specific utility or water source. Standardized facility-level reporting would allow individual projects and cumulative regional demand to be evaluated using comparable measurements.

Reporting should distinguish water withdrawal from water consumption and identify the source of the water used. Potable, reclaimed and other supplies have different implications for regional water planning, while withdrawal and consumption measure different aspects of water use. Absolute quantities should also accompany efficiency metrics such as WUE because improving efficiency does not necessarily mean total consumption is declining.

Reporting periods matter as well. Annual totals can conceal seasonal and peak demand, particularly if cooling requirements increase during hot or dry conditions. Monthly consumption and peak daily demand could provide utilities with a clearer picture of when water is required and whether maximum demand coincides with periods of regional water stress.

Wastewater and cooling information can provide additional context. Where cooling systems produce blowdown or other wastewater, discharge volume and destination affect local infrastructure requirements. Identifying the cooling technology and operating modes can also help explain why water demand changes over time without requiring disclosure of proprietary computing information.

Florida's new large-scale data-center permitting law already establishes part of this framework. For qualifying facilities seeking water allocations averaging at least 100,000 gallons per day, applicants must disclose water sources, amounts and losses associated with cooling and other facility processes and submit a water-conservation plan.[59]

The remaining distinction is between information supplied during permitting and measured operational performance. Projected demand is useful for evaluating a proposed facility, but subsequent measurements can establish whether actual water use matches those projections. TBRPC has similarly identified standardized reporting of direct and indirect water use as a consideration for regional planning.[56]

Standardization would be important for comparison. If one facility reports withdrawal, another consumption, another only WUE and another a corporate-wide total, the resulting figures cannot be meaningfully compared. Wherever practical, facilities could use consistent definitions, units, reporting periods and system boundaries and clearly distinguish measured values from estimates.

A practical facility-level reporting framework could include:

- withdrawal and consumption, reported separately;
- water source, including potable, reclaimed and other supplies;
- monthly and peak daily demand;
- discharge and wastewater volume;
- WUE using a defined methodology; and
- cooling technology and relevant operating modes.

For facilities with substantial electricity demand, indirect electricity-related water use could also be estimated where an appropriate methodology is available, but it should remain clearly separated from measured on-site water consumption.

Public reporting would not require disclosure of model weights, customer workloads, server configurations or other proprietary computing information. The purpose would be to provide enough resource-use information to evaluate individual facilities and, if multiple large facilities eventually develop, their cumulative demand on Tampa Bay's water and wastewater infrastructure.

Establishing standardized reporting before hyperscale development occurs could therefore address one of the principal limitations identified by the Tampa Bay facility review: the region does not currently have evidence of a major data-center water problem, but the available public information is also insufficient to measure the existing facility-level footprint confidently.

## 12.5 Drought Response Plans

Water requirements that appear manageable under normal conditions can become more consequential during drought. For future large-scale data centers in Tampa Bay, water planning should therefore address not only average annual demand but how a facility would operate when regional water supplies are constrained.[57, 58]

The 2026 shortage provides a local example. Prolonged rainfall and streamflow deficits reduced surface-water availability and regional storage, demonstrating that future facilities would need to operate through periods when water-system flexibility is reduced.[57, 58]

Future large-scale facilities could therefore develop drought response plans identifying normal and peak water demand, primary and backup supplies, and operational changes associated with increasingly severe shortage conditions. Where appropriate, those responses could be tied to established regional shortage stages rather than determined only after a drought occurs.

Depending on facility design, responses might include increasing low-water cooling operation, reducing evaporative cooling where technically feasible, maximizing available reclaimed-water use or suspending nonessential water uses. A credible plan should also identify water requirements that cannot safely be reduced because cooling and other essential operations must continue.

Backup supply is particularly important. A facility normally supplied by reclaimed water or operating with low potable demand could require additional potable water during supply interruptions or extreme operating conditions. That contingency should be included when utilities evaluate drought-period capacity rather than considering only normal operating demand.

Drought planning should account for duration as well as severity. Tampa Bay's regional system can shift among groundwater, surface water, reservoir storage and desalination, but those sources operate under different constraints during prolonged shortages.[57, 58] A facility designed around normal-year availability should therefore demonstrate how its water requirements would be met during extended dry conditions.

Florida's large-scale data-center permitting framework provides a foundation for this approach. Qualifying applicants must disclose water sources and losses and develop water-conservation plans, while proposed allocations remain subject to the state's reasonable-beneficial use, existing-user and public-interest requirements.[59] A drought-response plan could extend that planning specifically to declared regional shortages.

The facility-level reporting discussed in Section 12.4 could also support implementation by allowing actual demand during shortage conditions to be compared with approved projections and conservation commitments. Plans could be reassessed when a facility expands substantially or when additional phases materially change its water requirements.

Cumulative development should also be considered. One facility's drought demand may be manageable while several large facilities experiencing the same heat and water-supply conditions could create a substantially different regional requirement.

For future large-scale development, Tampa Bay could therefore require or encourage drought-response planning that answers a straightforward question: When the region enters a serious water shortage, what changes at this facility?

If little or nothing can change because of operational requirements, utilities should know that before committing water capacity. Where facilities can reduce potable demand, alter cooling operation or rely more heavily on alternative supplies, those capabilities should be identified in advance and incorporated into regional drought planning.

The objective is not to attribute Tampa Bay's periodic droughts to future data centers. It is to ensure that major new industrial water demands are planned around the drought conditions the region already experiences.

## 12.6 Strategic Siting

Where a large data center is built can be as important as how the facility itself is designed. The peer-reviewed siting scenarios examined in Section 10 found that geographic location can dramatically alter AI infrastructure's water footprint, while the case studies demonstrated how local water availability, electricity systems and infrastructure can shape facility impacts.[19]

For Tampa Bay, strategic siting could therefore evaluate infrastructure and resource conditions before a developer selects a parcel and utilities are asked to accommodate it. Sites could be assessed partly according to whether the surrounding water, electricity and wastewater systems are appropriate for the facility's expected demands.[56]

Reclaimed-water availability is one consideration. A site near sufficient available reclaimed-water capacity may reduce dependence on potable supplies, while another location could require substantial new infrastructure or reliance on drinking water. Existing pipelines alone would not establish suitability; treatment and distribution capacity, existing customers, future demand and reliability would also need to be considered.[58, 59]

Electrical and wastewater infrastructure present similar constraints. Hyperscale facilities can require substantial electrical capacity, potentially including new substations or transmission infrastructure, while water-based cooling can create additional wastewater-management requirements. Site evaluation could therefore consider whether these systems can accommodate the proposed facility without disproportionate expansion or displacement of existing needs.

Environmental and drought conditions should also influence suitability. A facility requiring substantial freshwater consumption may be a poor match for a location where supplies are constrained, environmental resources are sensitive or peak and drought-period demand cannot be reliably accommodated.

The Nature Sustainability modeling illustrates the potential importance of geography. Under alternative U.S. deployment patterns, water-conscious siting reduced the modeled AI-server water footprint by approximately 52 percent relative to the study's baseline geographical allocation, while unfavorable allocation increased it by as much as 354 percent.[19] These are national modeling scenarios rather than Tampa Bay forecasts, but they demonstrate how strongly location can influence water outcomes.

Strategic siting should also account for cumulative development. Infrastructure capable of supporting one large facility may not necessarily support several facilities concentrated in the same area. Northern Virginia's experience demonstrates why evaluating projects independently can miss the effects of an emerging regional cluster.[33]

Tampa Bay could therefore consider both individual site suitability and cumulative regional capacity. A performance-based approach would not require designation of a single data-center zone or prescribe one type of facility. Instead, large projects could demonstrate that their proposed locations can support expected water, electricity and wastewater demands under normal, peak and drought conditions.

That evaluation is most useful early enough to influence site selection. Once land has been acquired and a campus designed around a particular utility configuration, accommodating the project may require infrastructure extensions or other investments that could have been avoided at a more suitable location.

Relevant siting considerations could include:

- genuinely available reclaimed-water capacity;
- electrical and transmission capacity;
- wastewater treatment or reuse capacity;
- potable-water and drought-period requirements;
- environmental and watershed conditions; and
- cumulative development reasonably foreseeable in the surrounding area.

Sites where those conditions cannot adequately support a proposed facility could receive greater scrutiny or require a different facility design before development proceeds.

Strategic siting therefore does not establish that particular parts of Tampa Bay are universally suitable or unsuitable for data centers. It provides a framework for matching the demands of an individual project with the infrastructure and environmental conditions of its proposed location.

For Tampa Bay, the objective is to consider whether a location can responsibly accommodate a facility's remaining demands before development is approved, rather than attempting to correct an unsuitable infrastructure match afterward.

## 12.7 Data Center & Wastewater Infrastructure

One emerging approach to reducing data-center freshwater demand is to treat data centers and wastewater systems as potentially complementary infrastructure. Municipal wastewater-treatment plants produce treated effluent that may be suitable for industrial reuse, while data centers can require reliable non-potable water supplies for cooling. Recent research suggests that intentionally pairing the two could reduce freshwater demand while creating additional opportunities for resource recovery.[61]

A 2026 peer-reviewed study published in *Environmental Science and Ecotechnology* examined this concept at global scale. Researchers mapped more than 4,775 data centers and 57,547 municipal wastewater-treatment plants across 98 countries and modeled optimized pairings between the two systems. Their framework combines treated

wastewater reuse for data-center cooling with potential recovery of data-center waste heat for wastewater-treatment processes.[61]

The researchers found that approximately 87.8 percent of the beneficial pairings identified in their model occurred within 40 kilometers, or about 25 miles. Across the modeled global system, they estimated potential freshwater conservation of approximately 1.3 billion cubic meters per year, along with greenhouse-gas and economic benefits.[61] These are modeled global potentials, not measured savings from existing facilities.

The water exchange is relatively straightforward: appropriately treated municipal effluent can substitute for higher-quality freshwater in cooling applications. EPA similarly identifies treated municipal wastewater as a potential source for data-center cooling.[62] The research extends that concept by examining whether waste heat from data centers could also support energy-consuming processes at wastewater-treatment facilities, creating a bidirectional exchange rather than water reuse alone.[61]

Whether such pairing is practical depends on local conditions. Distance affects pipeline, pumping and energy requirements; available effluent must be sufficient and reliable; and water quality may require additional treatment for the selected cooling system.[61, 62] Florida already permits reclaimed municipal wastewater for industrial applications including certain cooling uses, subject to applicable treatment and operational requirements.[63]

This concept is particularly relevant to Tampa Bay because the region already has substantial wastewater and reclaimed-water infrastructure. Future siting analysis could consider whether locations near treatment facilities have sufficient reclaimed-water production, distribution capacity, electrical infrastructure and wastewater capacity to support large data-center development without substantial new potable-water demand.

The Quincy case provides an established example of integrating data-center and water-reuse infrastructure. The Quincy Water Reuse Utility treats cooling blowdown from Microsoft's data center and returns treated water for reuse, supplemented primarily with non-potable surface water.[26] Its experience also demonstrates that integration can require substantial specialized infrastructure: the system involved ten treatment systems, more than 30 miles of piping and approximately $31 million in investment.[26]

The emerging research therefore adds another dimension to strategic siting. Rather than evaluating only land, electricity and conventional utility access, communities can consider whether water and wastewater infrastructure make some locations better suited to water-dependent computing than others.[61] EPA's Water Reuse Action Plan similarly includes efforts to advance recycled water for data-center cooling and improve coordination among utilities, regulators and data-center stakeholders.[64]

For Tampa Bay, a useful next step could be mapping potential large-development areas against wastewater-treatment plants, reclaimed-water production and available capacity, transmission infrastructure, electrical capacity and wastewater constraints. Such analysis could identify locations where alternative water can be incorporated into a project from the beginning rather than added after facility design.

Pairing would still require project-specific evaluation. Additional treatment and pumping consume energy, residuals require management, reclaimed water may have competing uses, and heat recovery is practical only under suitable distance and temperature conditions. The presence of a wastewater-treatment plant therefore does not by itself make a location suitable for a data center.

The broader opportunity is to plan large-scale computing and water-reuse infrastructure together. Under appropriate local conditions, one system's output can become another's input, potentially reducing freshwater demand and avoiding more difficult infrastructure retrofits after development occurs.

## 12.8 Developer-Funded Infrastructure

Large data centers can require infrastructure beyond the boundaries of the facility itself, including water and reclaimed-water connections, wastewater capacity, pumping systems, electrical substations and transmission upgrades. When those improvements are primarily required by a particular development, Tampa Bay could consider how their costs should be allocated before utility commitments and development approvals are finalized.

The underlying principle is proportionality. Infrastructure costs created primarily by a new large industrial demand should not automatically be transferred to existing utility customers or taxpayers, but neither does every improvement associated with a project necessarily need to be funded entirely by the developer. Some infrastructure may provide broader benefits or serve additional customers over time.

Water and wastewater infrastructure illustrate the issue. Connecting a hyperscale campus to reclaimed water may require new pipelines, pumping, storage or treatment capacity, while cooling-related wastewater could require additional conveyance or treatment. Alternative-water strategies can therefore reduce pressure on potable supplies while still creating substantial infrastructure costs.

Quincy's experience demonstrates the potential scale. Its data-center water-reuse system ultimately involved specialized treatment infrastructure and more than 30 miles of piping.[26]

The lesson is not that Tampa Bay facilities would require comparable investment, but that water-reuse solutions can themselves require significant infrastructure and financing.

Electrical infrastructure can involve similarly substantial investment. Hyperscale campuses may require new substations, transmission upgrades or other grid improvements. Although responsibility for those costs may fall within utility and state regulatory processes, they remain relevant when evaluating the overall costs and benefits associated with a proposed development.

This is particularly important where governments consider economic-development incentives. Property-tax revenue, construction activity and other potential benefits can be evaluated alongside the public or utility infrastructure necessary to support the project. A development's fiscal impact may look different when both sides of that equation are included.

Cost allocation should also account for infrastructure that benefits users beyond the original project. A developer-funded reclaimed-water extension, for example, could later serve other customers. Cost-sharing, reimbursement, capacity credits or other legally available mechanisms may therefore be appropriate depending on who creates the need and who ultimately benefits.

Existing infrastructure also has finite capacity. Even where no immediate expansion is required, a large facility can consume capacity that would otherwise remain available and potentially accelerate future utility investment. Applicable capacity charges, impact fees or other mechanisms could account for that burden where legally appropriate.

Any specific arrangement would need to comply with Florida and local requirements governing utility rates, development charges and infrastructure financing. The recommendation here is therefore a planning principle rather than a particular fee structure: identify infrastructure attributable to the project, estimate its cost and determine responsibility before major public or utility commitments are made.

TBRPC's regional planning work similarly recommends considering both the benefits and costs of large-scale data-center development and discusses tools such as Community Benefits Agreements for establishing developer commitments.56 Regional coordination may also be necessary where a project located in one jurisdiction depends on water, electricity or other infrastructure extending beyond it.

For Tampa Bay, the guiding principle is straightforward:

> Where new infrastructure is primarily necessary to serve a large data-center development, the developer should bear an appropriate share of the cost rather than automatically transferring that cost to existing residents and utility customers.

Strategic siting and resource-efficient facility design can reduce the infrastructure required in the first place. Cost-allocation policies can then address the additional capacity that remains necessary, allowing projects to be evaluated according to both the benefits they provide and the infrastructure required to support them.

## 12.9 Community Benefit Agreements

Environmental safeguards are only one part of determining whether large-scale data-center development benefits Tampa Bay. If a project requires substantial land, electricity, water or public infrastructure, communities can also consider what commitments should accompany the development so that its economic benefits are shared locally and its potential burdens are addressed.

TBRPC identifies Community Benefit Agreements (CBAs) as one potential tool. Its 2026 white paper describes CBAs as legally binding agreements negotiated among developers, local governments and residents to align major developments with local priorities. The Council identifies utility use, environmental impacts, workforce projections and accountability structures as information communities can use when negotiating developer commitments.[56]

For data centers, a CBA could potentially address issues extending beyond ordinary zoning or water permits, including infrastructure contributions, environmental mitigation, workforce development, community investment, resource-use reporting and enforcement. TBRPC points to a recent agreement in Lancaster, Pennsylvania, containing provisions addressing financial contributions, water and energy use, air quality, noise and enforcement.

CBAs could complement rather than replace regulation. State water permits and local land-use requirements establish applicable regulatory standards, while a CBA could secure additional project-specific commitments involving reclaimed-water infrastructure, potable-water reductions, drought response, facility-level reporting or infrastructure contributions. Basic environmental protections should not depend on whether a community successfully negotiates them with a developer.

Tampa Bay would not be starting from zero. TBRPC notes that both Tampa and St. Petersburg have used CBA frameworks for major development projects, although not

specifically for data centers.[56] Existing local experience could therefore provide institutional knowledge if governments consider CBAs for future hyperscale projects.

CBAs could also address how the benefits and burdens of development are distributed. Large facilities may generate property-tax revenue, investment and construction activity while concentrating infrastructure, noise or other impacts near host communities. Commitments could respond to the circumstances of a particular project rather than follow a generic checklist, whether through workforce programs, infrastructure improvements, environmental mitigation or other locally identified priorities.

Accountability is essential. Where CBAs are used, commitments should identify what must be delivered, who is responsible, when it must occur, how performance will be measured and what happens if the commitment is not met.

TBRPC also raises concerns about non-disclosure and confidentiality agreements between governments and data-center developers, warning that they can restrict information sharing, contribute to inaccurate impact estimates, reduce community trust and interfere with effective CBA negotiations.[56] Legitimate proprietary information may require protection, but communities need sufficient information about environmental and public-infrastructure impacts to evaluate proposed commitments meaningfully.

TBRPC points local governments toward existing resources, including the Sabin Center for Climate Change Law's database of publicly available CBAs, data-center examples and a model developed by the NAACP.[56]

For Tampa Bay, CBAs could provide a project-specific mechanism for connecting economic development with measurable community commitments. Where governments provide infrastructure, approvals or incentives for major development, agreements can help clarify what benefits, mitigation measures and accountability the project will provide in return.

The recommendation is therefore not that every future data center require an identical CBA. Local governments could consider CBAs for developments large enough to create significant public infrastructure or community impacts, with commitments tailored to documented project effects and local priorities.

## 12.10 Regional Coordination

Large-scale data centers may be approved within individual cities or counties, but many of the resources required to support them operate at a regional scale. Water supplies,

watersheds, electricity systems and transmission infrastructure do not necessarily follow political boundaries. A project located in one jurisdiction can therefore create demands or impacts extending beyond the government responsible for approving its land use.[56]

Tampa Bay's drinking-water system demonstrates this clearly. Tampa Bay Water supplies multiple member governments across Hillsborough, Pasco and Pinellas counties using a regional portfolio of groundwater, surface water and desalinated seawater.[57] SWFWMD similarly manages water resources across jurisdictional boundaries through regional water-supply planning.[58] A substantial new demand in one jurisdiction therefore cannot always be understood solely within that jurisdiction's boundaries.

Electricity creates a similar issue. A hyperscale campus can require substantial generation, substation and transmission capacity located beyond the facility itself. This can geographically separate development benefits from infrastructure and environmental impacts: one jurisdiction may receive property-tax revenue and investment while some supporting costs or resource demands occur across a larger service territory.

TBRPC specifically emphasizes this concern. Its 2026 white paper warns that strategies developed independently by individual local governments can become disjointed and identifies regional collaboration and resource sharing as important considerations for potential large and hyperscale data-center development.[56]

Regional coordination is particularly important for cumulative development. Several projects that appear manageable when evaluated individually could collectively create substantially greater demand on the same water supply, electrical grid or wastewater infrastructure. Northern Virginia demonstrates how data-center development can become a regional infrastructure question once facilities begin clustering.[33]

A regional approach would not require every jurisdiction to adopt identical zoning or development rules. Hillsborough, Pinellas and Pasco counties and their municipalities have different infrastructure and local priorities. Coordination could instead establish shared information and planning assumptions for resources and impacts that cross jurisdictional boundaries.

Standardized facility-level information, as discussed in Section 12.4, could support that process. Tampa Bay could maintain a regional picture of major operating, approved and proposed data-center capacity, including expected electrical load, water source and demand, reclaimed-water requirements, wastewater capacity and development status. For phased campuses, reasonably anticipated full buildout could also be considered rather than only the capacity operating at the time of review.

This information could help planners evaluate cumulative infrastructure commitments and drought-period demand before they become constraints. It could also reveal situations in which several facilities depend on the same regional supply, electrical infrastructure or potable backup capacity even though they were approved by different jurisdictions.

TBRPC is well positioned to help facilitate this coordination because its data-center white paper was developed specifically as a regional resource for Tampa Bay governments.[56] Its role would not replace Tampa Bay Water, SWFWMD, local utilities, governments or electricity providers, but could help connect information and planning occurring across those institutions.

For Tampa Bay, regional coordination could therefore focus on several shared questions: How much large-scale data-center capacity is operating, approved or proposed? How much potable and reclaimed water has been committed? What electrical and wastewater infrastructure would be required at full buildout? How would those demands behave during drought and peak conditions? And when would additional development require major regional infrastructure expansion?

Tampa Bay does not need a single regional verdict on whether data centers are good or bad. It needs a way for governments, water agencies, utilities and regional planners to see the same infrastructure picture before individual local decisions collectively create a regional outcome.

## 12.11 Transparency and NDAs

Large data-center projects can involve legitimate confidential information. Developers may need to protect details about customers, computing hardware, security systems, proprietary technology, financing and business strategy. Commercial confidentiality, however, should not automatically extend to information describing a project's demand on public resources and infrastructure.[56]

This distinction is particularly important for water. If a proposed facility expects to withdraw or consume substantial quantities from a public utility or regional resource, communities have an interest in understanding how much water is required, where it will come from, when demand will peak and how the project could affect infrastructure capacity. None of those questions requires disclosure of customer identities, proprietary computing systems or security details.

TBRPC specifically raises concerns about non-disclosure and confidentiality agreements in data-center development, noting that they can restrict information sharing, contribute to

inaccurate impact estimates, undermine community trust and weaken communities' ability to negotiate effectively with developers.[56]

The Dalles provides an example of why water information can become controversial when treated as confidential. Google's water use became the subject of litigation after the City of The Dalles sought to prevent facility-specific consumption records from being disclosed. The eventual release of those records allowed the facility's demand to be evaluated using actual municipal information.[35, 37]

The lesson for Tampa Bay is not that governments should reject all confidentiality requests, but that proprietary business information should be distinguished from information necessary to understand public-resource and infrastructure demands before confidentiality agreements are established.

The same principle extends beyond water. Major electrical capacity and infrastructure requirements, publicly supported utility improvements, project incentives and infrastructure financing may be relevant to understanding the public implications of a development even when its computing operations remain confidential.

Transparency should also extend beyond project approval. Pre-development estimates describe expected resource demand, while operational measurements establish what actually occurred. The facility-level reporting discussed in Section 12.4 could allow projected water performance to be compared with measured results without requiring disclosure of proprietary computing information.

This does not require every operational measurement to be published in real time. Reporting intervals and aggregation can preserve meaningful oversight while protecting legitimately sensitive operating information. Security-sensitive details likewise may warrant protection without automatically making annual water consumption, water source or other resource-use information confidential.

Florida's new permitting requirements provide part of this foundation. Large-scale data centers seeking qualifying water allocations must provide regulators with information concerning water sources, amounts and losses associated with cooling and other processes.[59] The remaining question is what information should remain publicly accessible and how subsequent operational performance should be reported.

Local governments could consider that distinction before entering project-specific NDAs. Once broad confidentiality commitments are established, information necessary for public discussion may become more difficult to recover. Any approach would need to comply with Florida public-records law and applicable statutory exemptions; this report does not assume that every resource-use metric is legally required to be public under current law.

A useful principle is:

> Protect legitimate proprietary and security information, but preserve public access to information necessary to understand significant demands on public resources, infrastructure and finances.

Regional consistency could strengthen that approach. If jurisdictions apply substantially different confidentiality standards to facility-level resource information, cumulative regional planning becomes more difficult. TBRPC could provide a forum for considering common expectations across Tampa Bay governments.[56]

Transparency can also verify good performance. Standardized measurements could demonstrate when a facility minimizes potable-water use, meets reclaimed-water commitments or performs better than projected rather than serving only to identify poor performance.

For Tampa Bay, the objective is to retain enough information to answer a basic question after a major project is approved:

> What resources did this facility require, what did the public commit to support it, and did its actual performance match what was promised?

If confidentiality prevents those questions from being answered, it has extended beyond protecting the data center's business into obscuring the public-resource implications of the project.

## 13. What We Still Don't Know

After reviewing the available research, corporate disclosures, government data, Tampa Bay's existing facilities and the emerging state and regional policy response, substantial uncertainty remains. Some important questions about AI and water cannot currently be answered with the precision that public discussion often implies.

That does not mean that nothing is known. The evidence establishes that data centers can consume water directly for cooling, electricity production can create substantial indirect water demand, semiconductor manufacturing adds an upstream footprint, and the significance of those demands depends heavily on technology, location and scale. What remains difficult is connecting those components into a complete, current and independently verifiable accounting of AI's water footprint.

At the workload level, there is no universal answer to how much water an AI prompt consumes. Models, hardware, utilization, cooling, electricity, geography and workload

complexity can substantially change the result. Estimates for one generation of technology may also become outdated as later systems are deployed.

At the industry level, AI remains difficult to separate from the broader data-center sector. Facilities increasingly support mixtures of cloud computing, storage, search, enterprise services and AI workloads. National data-center statistics therefore cannot simply be labeled AI statistics, while corporate totals rarely reveal resource use attributable to particular models or workloads.

Indirect water use introduces additional uncertainty because electricity may come from changing mixtures of generation technologies across regional grids. Calculating the associated water footprint requires choices about geographic and temporal boundaries, generation mix and whether withdrawal or consumption is measured. Different methodologies can therefore produce different estimates.

Upstream manufacturing is similarly difficult to allocate. Semiconductor fabrication requires substantial quantities of ultrapure water, but assigning that footprint to an individual AI task depends on assumptions about hardware manufacturing, utilization, lifetime and the enormous number of computations performed over that equipment's useful life.

Corporate reporting does not yet resolve these problems. Companies publish increasingly detailed water, energy and replenishment information, but reporting boundaries and metrics are not standardized enough to support straightforward comparisons among companies or facilities.

The same limitation exists locally. Publicly available information generally does not reveal enough about Tampa Bay's existing individual facilities to calculate their actual water footprints. There is no reliable regional total for data-center water consumption, much less a defensible estimate of the portion attributable specifically to AI.

Future development introduces additional uncertainty. TBRPC reports that Tampa Bay currently has no large or hyperscale facilities, so there is no existing Tampa Bay hyperscale water footprint to measure.[56] Whether such facilities arrive, how large they would be, which cooling technologies they would use, where they would locate and what water sources they would require remain prospective questions.

Technological change further complicates long-range projections. Hardware and models are becoming more efficient while reasoning, multimodal generation, video and agentic systems can increase computational demand. Future resource use therefore depends not simply on efficiency improvements but on whether those improvements outpace growth in deployment and workload intensity.

Water volume alone also cannot fully describe environmental significance. A gallon of potable water consumed during drought in a constrained watershed is not equivalent to a gallon of reclaimed water used where adequate supply exists. Timing, source, local scarcity, competing demand and what happens to the water afterward all matter.

Several Tampa Bay questions consequently remain open: How quickly will large-scale AI development expand in Florida? How much water will future cooling systems consume under Florida conditions? How much reclaimed water could support that growth without displacing other beneficial uses? What indirect water demand will accompany the required electricity? And how many large facilities could the region accommodate before individually manageable projects become a cumulative infrastructure issue?

Florida itself is still investigating many of these questions. The independent OPPAGA study required by the state's 2026 legislation is due July 1, 2027.[59] Its findings may provide Florida-specific evidence unavailable when this report was prepared and could confirm, refine or challenge conclusions reached here.

This report should therefore be understood as a September 2026 evidence snapshot rather than a permanent accounting of AI's water footprint. Estimates should be updated as measurements improve, Tampa Bay's development landscape changes, technologies evolve and Florida-specific research becomes available.

Uncertainty should not be used in either direction. It does not justify assuming that future AI infrastructure will create a water crisis, but neither does it justify assuming that efficiency improvements will make large-scale development inconsequential. Where the evidence does not currently support a number or conclusion, the appropriate answer is that we do not yet know.

That is particularly important in a debate frequently reduced to simple numbers, such as water per prompt, gallons per data center or AI’s total share of water use. The evidence throughout this report shows that boundaries, location, technology and scale determine what those numbers actually mean.

These information gaps also identify what should be measured next. Better facility-level reporting, standardized accounting, operational measurements, Florida-specific research and transparent regional planning could turn many of today's unknowns into questions that can eventually be answered with evidence.

The objective is not to eliminate uncertainty before Tampa Bay makes decisions. It is to identify what is known, preserve what is genuinely unknown, and build the reporting and planning systems necessary to know more before the stakes become larger.

## 13.1 How Much Water Individual Data Centers Actually Use

One of the most basic questions in the AI-water debate remains surprisingly difficult to answer: how much water does an individual data center actually use? Corporate sustainability reports and national studies provide valuable aggregate information, but facility-level withdrawal and consumption are often not publicly available in a standardized form.

Even when a water-use figure is available, it may not describe the complete water balance. Withdrawal and consumption are different measurements, and facilities can differ substantially in how much withdrawn water is ultimately consumed through evaporation or returned as wastewater.

Water source is another important variable. Potable, reclaimed, groundwater, surface-water and other supplies have different implications for resource planning, yet consistent facility-level information identifying those sources is not generally available across the industry.

Cooling technology further complicates interpretation. Water demand can vary substantially among evaporative, dry, closed-loop and hybrid systems, meaning that electrical capacity or computing load alone cannot reliably establish facility water consumption.

WUE can provide useful efficiency information, but it does not resolve the disclosure problem. It may be reported at company or fleet level rather than for individual facilities, methodologies can differ, and efficiency does not reveal the absolute water consumption of a very large facility.

Annual figures can also conceal seasonal and peak demand, while wastewater information is necessary to understand how much water leaves a facility after use and what treatment or disposal it requires.

A useful facility-level water balance would therefore identify:

- withdrawal and consumption separately;
- potable, reclaimed and other water sources;
- cooling technology and facility-level WUE;
- monthly, seasonal and peak demand; and
- wastewater and discharge volumes.

The absence of these data also limits broader research because national and global estimates depend partly on assumptions about how representative facilities operate. Better standardized facility-level measurements would therefore improve both local oversight and aggregate modeling.

Tampa Bay provides a local example. As documented in Sections 11.2 and 11.3, our facility review identified existing data centers but generally could not locate sufficient public information to reconstruct their individual water balances. That prevents this report from producing a defensible estimate of current regional data-center water consumption.

The public information reviewed does not establish whether existing Tampa Bay facilities have a large or negligible water footprint. Resolving that question requires sufficient evidence about their actual operations, rather than treating missing information as evidence of either outcome.

The unanswered question is therefore more specific than simply “How much water do data centers use?” For many facilities, the missing information is:

> How much water entered this particular facility, where did it come from, how much was consumed, when was it used, and what happened to the remainder?

Until those questions can be answered consistently at the facility level, estimates of individual data-center water use will continue to carry substantial uncertainty.

## 13.2 How Much of the Data-Center Footprint Actually Belongs to AI

Even if the total water and electricity use of the data-center industry could be measured perfectly, another major uncertainty would remain: how much of that footprint should actually be attributed to artificial intelligence?

Data centers existed long before the current generative-AI boom and continue to support cloud storage, websites, enterprise software, databases, streaming, telecommunications, search and other digital services. As established in Section 5, total data-center resource use is therefore not the same thing as AI resource use.

The distinction becomes difficult at the facility level because data centers can support AI inference, conventional cloud computing, storage and other services simultaneously. Cooling systems, networking equipment, power conversion and other supporting infrastructure may also serve multiple workloads.

This creates an allocation problem. Electricity consumed directly by AI accelerators may be associated with AI workloads, but shared facility overhead and water consumption

cannot necessarily be divided as cleanly. A cooling system does not produce separate quantities of water consumption labeled “AI” and “non-AI,” so attribution requires assumptions about workload composition and utilization.

National estimates face the same problem at a larger scale. Government and industry datasets can describe total data-center electricity consumption and, in some cases, water use without identifying which kilowatt-hours or gallons supported AI. The broader data-center baseline established in Section 5.1 therefore cannot simply be relabeled as an AI footprint.

At the same time, AI is contributing substantially to new data-center growth. Accelerated computing, increasingly powerful models and construction of new hyperscale capacity are increasing electricity and infrastructure demand. But identifying AI as a driver of growth is different from determining its exact share of the industry's total existing footprint.

The boundary is also becoming less distinct as AI is incorporated into conventional digital services. Search, social media, cloud applications and other services increasingly combine traditional computing with AI systems. A single interaction can therefore involve both, making attribution partly an accounting problem rather than a clean physical distinction.

Training and inference introduce additional allocation challenges. Training can involve concentrated AI-specific computation, but assigning its energy, water and hardware footprint across later uses requires assumptions about model lifetime and utilization. Inference may be distributed across multiple facilities and regions, making location-specific water attribution difficult without provider-level information.

Estimates of AI's total water footprint therefore depend partly on modeling and allocation assumptions, particularly where companies do not disclose workload-specific resource use. Such estimates can be useful, but they should not be presented as though every gallon attributable to AI has been directly measured.

This distinction matters in public discussion. Growth in total U.S. data-center electricity or water demand is evidence of expanding digital infrastructure, but it should not automatically be described as an equivalent increase in AI water use unless the underlying evidence separates AI from other workloads.

Conversely, difficulty isolating AI does not establish that its contribution is insignificant. The uncertainty concerns the size of AI's share, not whether AI contributes to data-center demand.

The most defensible conclusion is therefore:

> We know AI is driving a substantial portion of new data-center growth, but we still cannot cleanly divide the industry's total water footprint into “AI” and “non-AI” gallons.

Resolving that uncertainty would require better information about workload composition, AI-specific electricity consumption, shared facility overhead and cooling allocation. Until those data become available, this report should continue to distinguish among AI-specific measurements, modeled AI estimates and statistics describing the data-center industry as a whole.

## 13.3 How Different AI Services and Digital Activities Really Compare

Published estimates exist for ChatGPT, Gemini, Le Chat, older GPT models, image and video generation, web browsing, social media and streaming. The difficulty is that many were not produced using the same methodology, system boundary, workload definition or unit of activity.

Among AI providers, for example, Google has published a measured operational estimate for the median Gemini Apps text prompt, OpenAI has reported an average ChatGPT query estimate without the same level of methodological detail, and Mistral has published a broader lifecycle assessment for Le Chat.[10, 17, 18] These figures provide useful information, but they do not measure precisely the same thing.

As of September 2026, provider-specific water data are not standardized enough to calculate a defensible water-efficiency ranking across ChatGPT, Claude, Gemini, Le Chat and other widely used systems. A smaller published number may reflect a narrower accounting boundary rather than a genuinely more water-efficient service.

The problem becomes greater when AI is compared with conventional internet activity. Research has quantified lifecycle impacts of web browsing, social media, video streaming, music streaming and video conferencing using consistent methodologies within individual studies.[20] Those comparisons can be meaningful within a common framework, but lifecycle results cannot automatically be placed beside operational water estimates for individual AI prompts.

An operational AI estimate may include data-center computing and cooling while excluding telecommunications networks, user devices or hardware manufacturing. A lifecycle estimate for another digital activity may include those components. Directly comparing the

resulting numbers would therefore assign environmental impacts to one activity that were excluded from the other.

This report consequently cannot responsibly establish universal comparisons such as how many searches equal one AI prompt, whether an hour of social media uses more water than an hour of AI, or whether one commercial AI service is more water-efficient than another when the underlying measurements are incompatible. Published figures can be combined mathematically, but arithmetic does not correct incompatible assumptions.

The unit of activity creates another problem. One prompt, one search, one hour of streaming and one gigabyte of transmitted data are not equivalent functional units. AI prompts also vary substantially in length and complexity, while social media and other digital services increasingly combine text, images, video and AI-enabled systems.

Multimedia AI introduces similar uncertainty within AI itself. Controlled research supports the conclusion that image generation can require substantially more energy than text generation under comparable experimental conditions, while emerging evidence indicates that video generation can be more computationally intensive still.[12, 13] However, standardized provider-independent water measurements for ordinary commercial image, audio and video generation remain limited. Greater energy demand also does not translate into a universal water multiplier because cooling technology, climate and electricity source still matter.

A rigorous comparison would ideally evaluate activities using the same system boundary and treatment of data-center electricity and cooling, electricity-related water, networks, end-user devices, hardware manufacturing, geography, withdrawal versus consumption and functional unit. Very little current research provides that level of consistency across both AI and conventional digital activities.

The categories themselves are also beginning to overlap. Search, social media, recommendation systems and other digital services increasingly incorporate AI components.[21] A supposedly conventional internet interaction may therefore already involve AI computation, making the boundary between the categories less distinct.

The absence of standardized comparisons does not mean AI and conventional digital activities have identical footprints. It means their quantitative relationship cannot yet be expressed confidently using many of the simple conversions circulating in public discussion.

Accordingly, as of September 2026, this report does not identify a defensible water-efficiency leaderboard for ChatGPT, Claude, Gemini, Le Chat or other major AI services,

nor a defensible universal conversion between AI prompts and searches, social media or streaming.

That is not because no numbers exist. It is because the available numbers frequently answer different questions.

Future standardized reporting and comparable lifecycle research could make stronger comparisons possible. Until then, numbers should be compared directly only when their measurement boundaries are sufficiently compatible; otherwise, they should remain separate pieces of evidence rather than being presented as though they measure the same thing.

## 13.4 The Full Supply-Chain Footprint

Operational water use is only part of AI's total footprint. Before an AI model is trained or used, the physical infrastructure supporting it must first be manufactured and built, including semiconductors, servers, memory, networking and power equipment, cooling systems and construction materials.[2, 27-29]

Semiconductor fabrication is the best-documented part of that upstream footprint. Chip manufacturing can require large quantities of ultrapure water for wafer cleaning and other fabrication processes, while producing that ultrapure water itself requires additional source water.[27] Industry-level research also shows substantial water consumption across semiconductor manufacturing.[28]

But semiconductor water use is not equivalent to AI water use. Fabrication plants produce chips for many products besides AI accelerators, while AI servers contain components from multiple manufacturers and supply chains. Once installed, the same hardware may support multiple models, AI and non-AI workloads, and changing utilization over its useful life.

Attributing manufacturing water to a particular model or inference therefore requires assumptions about hardware lifetime, utilization, workload mix and allocation methodology. The same problem applies to servers, memory, networking equipment and other supporting hardware whose embedded water, energy and material footprints are not consistently available at the level needed for AI-specific accounting.

Construction adds another upstream component. Hyperscale campuses require buildings, foundations, electrical equipment and cooling infrastructure whose steel, concrete and other materials carry embedded water and energy impacts.[2] Hardware replacement can

add further impacts when equipment is replaced as newer accelerators become available, even if the newer hardware reduces operational resource use per computation.

Geography also separates these impacts from operational water use. Semiconductor fabrication, hardware assembly and materials production may occur in different states or countries from the final data center, and the electricity used during manufacturing can carry additional water impacts. A model served in Florida may therefore depend on embedded water consumed months or years earlier in entirely different watersheds.

A full lifecycle calculation would consequently need to determine how much water was required to manufacture the relevant infrastructure, how much of that infrastructure should be attributed to AI, how shared hardware should be allocated among models and workloads, and how those impacts should be distributed across the equipment's useful life.

Different reasonable assumptions can produce substantially different results. This is why operational estimates such as “water per prompt” generally do not represent the complete physical water footprint of the infrastructure making the prompt possible. Expanding the boundary to include supply-chain water would provide a more complete accounting, but doing so without a defensible allocation method could introduce more apparent precision than real knowledge.

Provider-level comparisons face the same limitation. Operational water may be measured at a company's data centers while manufacturing impacts occur elsewhere in its supply chain. Corporate environmental reporting increasingly addresses supply-chain impacts, but available information is not yet standardized enough to reconstruct a complete AI-specific water footprint across providers.

What is known is therefore clearer than what remains unknown. Semiconductor manufacturing, hardware production and data-center construction have real upstream water footprints. What remains much harder to determine is:

> How much of that embedded water should reasonably be attributed to one AI model, one provider or one unit of useful computation?

As of September 2026, the evidence is not strong enough to assign a reliable universal supply-chain water value to an AI prompt or provider. The embedded footprint is real, but its allocation remains methodologically uncertain.

## 13.5 What Future AI Infrastructure Will Actually Look Like

One reason long-range AI water projections remain uncertain is that the infrastructure being projected is itself changing rapidly. The uncertainty is not only how much AI use will exist in 2030, but what a representative AI workload, model and data-center architecture will look like by then.

Some trends could reduce resource intensity. Smaller specialized models, mixture-of-experts architectures, quantization, improved utilization and more efficient accelerators can reduce the energy required for comparable tasks.[11, 33, 44] At the same time, long reasoning, tool use, multimodal processing and agentic workflows can increase the amount of computation performed for a single user request.[11, 33]

Reasoning and agentic systems illustrate the problem. Long reasoning workloads can require substantially more energy than ordinary text inference, while an agent may perform multiple model calls, tool interactions and intermediate steps in response to one user request.[11] The relevant unit of activity may therefore become more complex than “one prompt.”

Multimodal systems add further uncertainty. Future AI may routinely process and generate combinations of text, images, audio and video, with video among the emerging workloads capable of requiring substantially more energy than simple text generation.[13] Resolution, duration, model architecture and generation methods can also change rapidly.

Hardware and cooling may move in the opposite direction. New accelerators can increase useful computation per unit of electricity, while direct-to-chip liquid cooling, closed-loop systems and dry heat rejection can potentially reduce routine water consumption.[11, 13, 41, 45] Future facilities may therefore support substantially more AI activity per megawatt or gallon than current systems.

The electricity system introduces another variable because AI's indirect water footprint depends on the generation supplying future data-center loads.[48] Geography and workload scheduling may matter as well if computing can increasingly be shifted among regions or times with different electricity, cooling or water conditions.

Efficiency alone cannot resolve the uncertainty because of the rebound effect. More efficient models and hardware reduce the resource cost of individual tasks but can also make AI cheaper and easier to deploy, potentially increasing total activity enough to offset some or all of those savings.[13, 54]

Future projections consequently depend on assumptions about workload intensity, model architecture, hardware efficiency, utilization, cooling technology, electricity generation,

geographic deployment and total demand growth. Changing several of those assumptions can produce substantially different outcomes.[19]

The central uncertainty is therefore not simply "How many prompts will people send?" That assumes the unit being counted remains technologically stable. A more useful question is:

> What will one unit of useful AI work actually require five years from now, and how many of those units will society demand?

We do not yet know either part with confidence.

This limits extrapolation from current per-query measurements. Present-day text-prompt, reasoning or multimedia benchmarks describe the systems tested, but they do not establish the representative workload mix of AI infrastructure in 2030.

As of September 2026, the evidence supports scenario planning more strongly than point prediction. Future AI resource intensity may improve while workload intensity and total demand also increase. The interaction among those trends, rather than any one of them alone, will determine the resulting water footprint.

## 13.6 What Will Actually Be Built in Tampa Bay

The largest local uncertainty is also the most fundamental: Tampa Bay does not yet know what large-scale data-center development, if any, will actually be built in the region. As of September 2026, TBRPC reports that Tampa Bay has small and medium data centers but no current large or hyperscale facilities.[56]

That means there is no existing Tampa Bay hyperscale water footprint to measure. This report can evaluate how large facilities operate elsewhere, identify regional water constraints and assess potential mitigation strategies, but it cannot determine the impact of facilities that have not yet been proposed or constructed.

Future proposals could produce substantially different outcomes. A 100 MW facility and a 500 MW campus present different infrastructure questions, while facilities of similar electrical size could have very different water requirements depending on cooling technology, water source and location.

A facility relying heavily on evaporative cooling could create substantial consumptive demand, while closed-loop liquid cooling with dry heat rejection could require considerably less routine water. Similarly, a facility using available reclaimed water presents a different regional supply question from one requiring the same volume of potable water.

Location connects these variables. A project near adequate electrical infrastructure, reclaimed-water capacity and appropriate wastewater systems could present a substantially different infrastructure proposition from an otherwise similar facility requiring those systems to be expanded specifically to accommodate it.

The scale of development beyond the first project also matters. One large facility could remain isolated, or it could contribute to a regional cluster as electrical, fiber and other infrastructure attracts additional development. Several individually manageable projects could collectively create much greater demand on regional water, wastewater and electrical systems.

The absence of publicly identified hyperscale projects should also be understood as the absence of confirmed public projects, not proof that no developer is considering Tampa Bay. Economic-development negotiations, site acquisition and utility planning can precede formal development applications.

Florida's regulatory environment is also changing. New large-scale data-center water requirements became effective in 2026, while Hillsborough County is considering additional local policies and TBRPC has begun regional planning.[56, 59, 60] The independent OPPAGA study due July 1, 2027, may provide additional Florida-specific evidence after this report's cutoff.[59]

This report is therefore a September 2026 snapshot. It describes the facilities, policies, water conditions and publicly available development information that can currently be established. It should not be read as a prediction that Tampa Bay will, or will not, become a major data-center market.

If a large proposal becomes public, several currently hypothetical questions would become answerable:

- What is its proposed electrical capacity and ultimate buildout?
- What cooling technology will it use?
- How much water will it withdraw and consume, and from what sources?
- What will its monthly and peak demand be?
- What wastewater and electrical infrastructure will it require?
- How will it operate during drought?
- Who will pay for necessary infrastructure?
- Is it an isolated project or part of a larger development pipeline?

Those project-specific facts would be more useful for assessing Tampa Bay's actual future water impact than applying a generic national “gallons per data center” figure.

The uncertainty cuts in both directions. Tampa Bay's lack of hyperscale facilities means there is currently no evidence that they are creating a major regional water burden, but there is also no local operational evidence demonstrating that future hyperscale development would be inconsequential.

The correct local conclusion is therefore not that Tampa Bay has a future data-center water problem, nor that it does not. The impact has not yet been determined because the infrastructure that would create it has not yet been built.

That is why the planning work underway in 2026 matters. Florida, Hillsborough County, TBRPC, regional water agencies and local governments have an opportunity to establish expectations about water sources, cooling, reporting, siting, drought resilience and infrastructure while those decisions can still influence what ultimately gets built.

## 13.7 What Uncertainty Does and Does Not Mean

This report distinguishes measured conditions, modeled scenarios, company-reported figures and unresolved questions. Each supports different conclusions: measurements describe the conditions observed; models explore outcomes under stated assumptions; and company disclosures provide information whose independent verification may be limited.

For Tampa Bay, the absence of established large or hyperscale facilities describes the region’s current development, while the lack of sufficient facility-level water data limits assessment of existing operations.[56] Neither resolves the effects of future proposals.

Planning can proceed while those questions are investigated. Decisions affecting shared resources should identify the evidence available, disclose assumptions and require information sufficient to assess facility operations and cumulative demand. Monitoring and transparent reporting should then support revisions to conclusions and operating requirements as conditions change.

Florida’s forthcoming OPPAGA study may provide additional evidence for that process.[59] Future Tampa Bay proposals should likewise be evaluated using their actual capacity, cooling systems, water sources, discharges and locations.

# 14. Beyond Consumption: How Data Centers Can Affect Water and the Environment

The previous sections have focused primarily on the amount of water associated with AI and data-center infrastructure, where that water comes from and how differences in cooling, electricity, manufacturing, geography and scale can change the result. Those questions are central to understanding the industry's water footprint, but water consumption alone does not describe its full environmental relationship.

Large computing facilities are physical infrastructure. They occupy land, produce heat, require electricity and utility connections, generate wastewater under some cooling configurations, and can place new demands on shared water and wastewater systems. The significance of those effects varies considerably by facility and location. A large campus using evaporative cooling in a water-stressed region presents a different set of concerns from a facility using dry cooling and low-water electricity, just as a project built on previously developed industrial land may interact differently with the surrounding environment than one requiring substantial new land disturbance.

Artificial intelligence adds another distinction. Many of these effects are not unique to AI. Conventional data centers, industrial facilities and other large buildings can also require cooling, electricity, wastewater service and supporting infrastructure. AI matters because its rapid growth is contributing to increasing computing demand and the development of larger and more power-dense facilities. Separating the impacts of data centers generally from the additional demands associated with AI is therefore necessary to understand what is actually changing.

## 14.1 Water Quantity vs. Water Quality

Sections 2 and 3 established that withdrawal and consumption measure different parts of a facility's water footprint, while Section 3.4 examined another distinction: the amount of water used does not by itself describe the condition of that water after use.

For facilities using evaporative cooling towers, evaporation concentrates dissolved minerals and other constituents in the circulating water. A portion of that increasingly concentrated water must eventually be removed as blowdown, and cooling systems may also use chemical treatment to control scaling, corrosion and biological growth.[5, 7] As discussed earlier, the resulting wastewater may be treated, reused, sent to a wastewater-treatment system or discharged under applicable requirements. The existence of

blowdown therefore does not establish pollution, but it creates a water-quality question that gallons of withdrawal or consumption alone cannot answer.

This distinction becomes especially important when comparing environmental strategies. Reclaimed water can reduce demand for drinking-quality supplies without eliminating the need to manage wastewater. Recycling blowdown can reduce both freshwater demand and wastewater volumes, but may require additional treatment. Dry or closed-loop cooling can substantially reduce direct water consumption while creating different energy, equipment or chemical-management tradeoffs. The environmental outcome therefore depends on the system as a whole rather than any single water metric.

A complete water assessment should consequently follow water in both directions. Where did it come from, how much was withdrawn and consumed, what happened to it during use, what left the facility, and where did that water ultimately go? The first questions describe water quantity. The latter questions help determine whether water use also creates water-quality or wastewater-management concerns.

## 14.2 Cooling Water, Blowdown and Wastewater

The water leaving a cooling system can matter as much as the water entering it. As discussed in Section 3.4, evaporative cooling systems periodically remove concentrated water as blowdown, which can contain elevated levels of dissolved minerals and residual treatment chemicals.[5, 7] What happens next depends on the facility and local infrastructure. Blowdown may enter a municipal wastewater system, receive additional treatment, be reused within the facility or, where permitted, follow another regulated discharge pathway.

This creates a connection between data-center development and wastewater infrastructure that is easy to overlook when attention is focused only on water supply. A large facility using evaporative cooling may require not only a reliable source of makeup water but also sufficient treatment and conveyance capacity for the wastewater it generates. The quantity, composition and timing of that wastewater can therefore matter to utilities evaluating whether existing systems can accommodate a new facility.

A 2026 incident in Cheyenne, Wyoming illustrates why the discharge side of the water balance deserves attention. During commissioning of cooling infrastructure associated with Meta's new data-center development, an industrial user discharged water containing the bacterium Cupriavidus gilardii into Cheyenne's wastewater system. The bacterium subsequently affected the city's wastewater-treatment and recycled-water systems, and

the Cheyenne Board of Public Utilities classified the industrial user as being in significant noncompliance with its discharge requirements.[67]

The incident did not affect Cheyenne's drinking-water system, but it disrupted the city's separate recycled-water service. The utility reported that its wastewater-treatment facilities could not remove the bacterium and suspended production of recycled water for irrigation as a precaution. Irrigation service subsequently resumed with continued monitoring. The utility also prohibited discharges of fill-and-flush water and closed-loop cooling water until further notice, while continuing to allow domestic sewage and evaporative-cooling blowdown. According to the utility's July 2026 release, Goat Systems, LLC had appealed the significant-noncompliance notice.[67]

Cheyenne should not be interpreted as evidence that data centers routinely contaminate municipal water systems or that C. gilardii is a normal component of data-center wastewater. Its significance is narrower but still important: water management does not end when water reaches a cooling system. Commissioning, treatment, discharge and wastewater-system compatibility can all create environmental and operational risks that are invisible in a simple calculation of gallons withdrawn or consumed.

The same principle applies when reclaimed water is used for cooling. Reclaimed water can substantially reduce competition for potable supplies, but its use still requires attention to incoming water quality, treatment requirements, cooling-system chemistry and the management of resulting wastewater or residuals. The environmental benefit therefore depends on the entire water cycle surrounding the facility, not simply whether the original source was potable or reclaimed.

## 14.3 Construction, Groundwater and Stormwater

The environmental footprint of a data center begins before its servers are switched on. Large campuses can require land clearing, grading, excavation, roads, buildings, electrical infrastructure and extensive areas of impervious surface. These activities can change how rainfall moves across a site and, depending on local geology and hydrology, can affect stormwater, wetlands, surface waters and groundwater.[56, 68]

These concerns are not unique to data centers. They are familiar issues in large-scale development generally. In Florida, the Environmental Resource Permitting program regulates many alterations to the land surface, including stormwater-management systems, activities that alter surface-water flows, and dredging or filling in wetlands and other surface waters.[68, 69] The relevance to data centers comes primarily from their

potential scale: a large campus may combine multiple buildings with substations, backup-generation equipment, cooling infrastructure, roads and other supporting development.[56]

Stormwater provides one example of why land use belongs in a water assessment. Replacing vegetation or permeable ground with roofs, pavement and other impervious surfaces changes how rainfall infiltrates, collects and runs off a property. The resulting effects depend heavily on site design, soils, drainage, stormwater infrastructure and surrounding water resources. The presence of a data center therefore does not establish that nearby waters will be harmed, but its development footprint can create water-management questions separate from the water consumed during operation.[68, 69]

In 2026, four families living near Meta's Stanton Springs data-center campus in Newton County, Georgia filed a lawsuit alleging that clearing, blasting and excavation disturbed groundwater and introduced silt and sediment into their private wells. The complaint describes clogged water lines, damaged appliances and additional concerns involving dust, lighting and noise.[70]

Reporting by The Atlanta Journal-Constitution documented the practical burden on one household. A reporter observed the Morrises' water-hauling tank, muddy runoff from their water filter and jugs supplying water for cooking, cleaning and drinking. These observations document the household's circumstances but do not independently establish what caused the well-water problems. Meta disputes the allegations and said a study it commissioned found that groundwater flows away from the plaintiffs' properties. The company also said water used during construction and operation comes from a local utility.[70]

The dispute highlights the value of documenting groundwater conditions, nearby private wells and site hydrology before construction begins. Baseline measurements and continued monitoring can help investigators assess later complaints, identify possible causes and determine appropriate corrective action.

For Tampa Bay and surrounding areas, this broader perspective matters because evaluating future large-scale data centers solely by their operational water demand could overlook how the site interacts with rainfall, stormwater, wetlands and groundwater. Water planning begins with the facility's location and development footprint, not only with the cooling system it eventually operates.

## 14.4 Heat, Electricity and Air

Water is not the only environmental pathway created by cooling. Nearly all of the electricity consumed by computing equipment ultimately becomes heat, which must be transferred

away from servers and released somewhere else. Cooling therefore does not eliminate heat; it moves it from computing equipment to another medium and ultimately to the surrounding environment.

The field measurements and emerging studies reviewed in Section 7.8 identify localized warming as another potential impact of data-center development. Its extent depends on facility design, land cover, vegetation, weather and the concentration of development. Local assessments should consider these conditions when evaluating heat exposure in surrounding communities.[30-32]

Electricity creates a broader environmental connection. As discussed earlier in this report, data centers can require very large and relatively continuous electrical loads.[15] The environmental effects associated with that demand depend heavily on how the electricity is generated. Where additional demand is supplied by fossil-fuel generation, it can contribute to greenhouse-gas emissions and conventional air pollutants; where lower-emission generation supplies the load, that profile can be substantially different. The same principle applies to backup generation located at data-center campuses, which can create localized emissions when operated.

On-site power generation can create community impacts beyond those associated with occasional backup-generator use. A July 2026 Reuters investigation documented natural-gas turbines installed to supply xAI's Colossus 2 project, including 57 turbines in Southaven, Mississippi. Residents reported persistent noise and sleep disruption. Using manufacturer specifications and assumed operating conditions, Reuters estimated substantial potential annual air-pollutant emissions. Those estimates describe potential emissions, rather than measured annual releases or demonstrated cases of illness caused by the facility.[71]

The permitting dispute illustrates the importance of evaluating how supporting power infrastructure actually operates. Mississippi regulators treated the temporary turbines as mobile sources exempt from stationary-source permitting, an interpretation challenged by civil-rights and environmental groups. A July 2026 state order documented 69 temporary turbines, established retirement deadlines and imposed operating and reporting conditions while permitted permanent turbines were being developed. The case shows why environmental review should distinguish equipment used during outages from generation supplying ongoing computing demand.[71, 73]

This reinforces a recurring theme throughout this report: reducing one environmental impact does not necessarily eliminate the others. A cooling design that greatly reduces direct water consumption may require more electricity, while an energy-efficient

evaporative system may consume more water locally. Heat, electricity, air emissions and water should therefore be considered as connected parts of the infrastructure rather than evaluated in isolation.

## 14.5 Cumulative and Community Effects

The environmental significance of a data center can look different when facilities are considered collectively rather than one at a time. A single project may represent a manageable addition to local water and electrical demand, while several large facilities drawing on the same watershed, utility or electrical system can create pressures that are less apparent during individual project reviews. This makes cumulative impact particularly important in regions where data-center development becomes concentrated.

Water availability provides a clear example. A facility's annual demand may appear relatively small compared with total regional supply, but annual averages can conceal periods when demand and water stress are highest. Drought, seasonal variation, competing municipal and agricultural needs, and the simultaneous operation of multiple facilities can all change the significance of the same volume of water. Research examining AI infrastructure has similarly found that geographic distribution can substantially increase or decrease projected water impacts depending on where future computing capacity is located.[19]

A 2026 analysis by Ceres illustrates this issue using the Phoenix metropolitan area. Rather than evaluating individual facilities in isolation, the study modeled how existing and proposed data centers could collectively affect water stress across the region. It found that impacts varied substantially by location and season, with some already water-stressed areas experiencing modeled increases in annual water stress of as much as 17 percent and larger increases during certain months.[72] These results are specific to the Phoenix case study and should not be treated as predictions for Tampa Bay. Their broader relevance is methodological: regional effects can emerge from the combined demands of facilities that may appear less significant when evaluated separately.

Community impacts can also be unevenly distributed. The benefits of data-center development, including investment, tax revenue and infrastructure improvements, may occur across a broader jurisdiction while water demand, construction, noise, heat, air emissions or new utility infrastructure are concentrated near particular communities. Research examining data-center development and environmental justice has therefore begun asking not only how large an environmental footprint is, but where that footprint occurs and who experiences it.[35, 36]

Noise is another neighborhood impact that water-use totals do not capture. Virginia's Joint Legislative Audit and Review Commission found that some residents near data centers report that constant low-frequency noise affects their well-being. The review noted that this noise rarely violates existing noise ordinances, which typically address loud, short-term disturbances. It also found that a large majority of data centers do not generate noise complaints because of their location or design. These findings highlight the importance of proximity to homes, acoustic design and noise standards that account for continuous low-frequency sound.[64]

Regional totals can hide differences in exposure and resource pressure among communities. Evaluating cumulative effects requires asking whether multiple facilities depend on the same water source, wastewater system or electrical infrastructure; whether their demands overlap during drought or peak conditions; and how environmental burdens and economic benefits are distributed among affected communities.

For Tampa Bay, where large-scale data-center development is still emerging, cumulative planning offers an opportunity that regions with established clusters may not have had. Individual projects can be evaluated not only against available capacity at the time of application, but also against the amount and type of development the region could support if additional facilities follow. That shifts the question from "Can this project be supplied?" to the broader question: "What happens if this project becomes the first of several?"

## 14.6 What Is Actually AI-Specific?

Many of the environmental effects discussed in this section are data-center impacts, not impacts unique to artificial intelligence. Cooling towers can consume water and produce blowdown regardless of the software running on the servers. Construction can alter land and stormwater patterns. Computing equipment produces heat. Electricity generation can carry water, carbon and air-quality consequences. These issues existed before the current expansion of generative AI.

AI changes the picture primarily through scale, computing intensity and the pace of infrastructure growth. Modern AI training and inference increasingly rely on specialized accelerators deployed in high-density computing environments. As AI use expands, it can contribute to greater electricity demand, more concentrated computing capacity and the construction or expansion of facilities designed to support those workloads.[15, 19]

That distinction also explains why national statistics for the entire data-center industry cannot simply be relabeled as AI statistics. Many facilities support mixtures of cloud computing, storage, search, streaming, business applications and AI, and public reporting

often does not separate their resource use by workload.[1, 15] A data center that supports AI is therefore not necessarily an exclusively "AI data center," and its total water or electricity use cannot automatically be attributed to AI.

At the same time, separating AI perfectly from the infrastructure supporting it may become increasingly difficult. A facility can host multiple workloads, electrical and cooling systems serve the facility as a whole, and the same computing hardware may perform different tasks over time. The most defensible approach is therefore to distinguish between measured facility impacts and claims about what portion of those impacts is specifically attributable to AI.

For environmental planning, that distinction does not make the infrastructure irrelevant. Communities still have to supply electricity, manage water and wastewater, evaluate land development and determine whether supporting infrastructure can accommodate proposed facilities regardless of the exact percentage of computing devoted to AI. What AI changes is the possibility that demand for that infrastructure may grow faster, become more concentrated and reach larger scales than previous patterns of data-center development.

This is why the environmental question is broader than whether AI itself "uses" or "pollutes" water. A more useful question is: How is AI changing the amount, type and location of computing infrastructure being built, and what environmental demands follow from those changes?

## 15. Key Takeaway

AI really does use water, but no single scientifically meaningful number describes its footprint across all applications and locations. That footprint depends on the work being performed, the hardware running it, how facilities are cooled and powered, and the water required to manufacture their equipment.

Water quantity is only part of the environmental question. Assessment must also consider wastewater, construction effects, electricity demand, air emissions, heat, noise and utility costs. The cases reviewed include measured neighborhood temperature increases, residents' reports of persistent noise and sleep disruption, a documented interruption of recycled-water service, disputed allegations of private-well damage, and contested permitting for on-site power generation. These findings differ in what they establish, but each identifies an issue requiring scrutiny. Most concern data-center infrastructure

generally; AI can increase demand for that infrastructure even when its precise share of a facility's resource use cannot be isolated.

Siting, cooling, electricity supply, water sources, treatment systems and public policy shape these impacts. Efficiency improvements can reduce resource requirements for equivalent computing, while total demand can rise as computing expands. Regional planning must also account for the combined pressures of facilities sharing water, wastewater and electrical systems.

For the broader Tampa Bay region, the most consequential finding is that many choices remain open. The regional assessment reviewed in this report identifies no established large or hyperscale data-center development as of September 2026. That gives the region an opportunity to set expectations for where large facilities belong, what they must disclose, how they should operate during drought, how their discharges and development footprints should be managed, and who should pay for the infrastructure they require.

The practical question is therefore larger than "How many gallons does AI use?" It is whether proposed infrastructure can operate within the region's water, environmental and utility constraints while accounting for the combined demands of projects that may follow.

## 16. Sources & Further Reading

Sources are numbered according to their citations throughout the report.


1. Resources, Conservation & Recycling — The Water Use of Data Center Workloads: A Review and Assessment of Key Determinants
2. ICEF Sustainable Data Centers Roadmap (2025)
3. U.S. Geological Survey — Water-Use Terminology
4. World Resources Institute — Aqueduct 4.0: Updated Decision-Relevant Global Water Risk Indicators
5. U.S. Department of Energy, Federal Energy Management Program — Cooling Water Efficiency Opportunities for Federal Data Centers
6. Resources, Conservation and Recycling — Water-Energy Tradeoffs in Data Centers: A Case Study in Hot-Arid Climates (2022)
7. Journal of Environmental Chemical Engineering — Recovery and Treatment of Cooling Tower Blowdown Water
8. Energy — Comprehensive Assessment of Coolants Used Single-Phase Immersion Cooling Data Center (2025)

9. Communications of the ACM — Making AI Less “Thirsty”: Uncovering and Addressing the Secret Water Footprint of AI Models
10. Google — Measuring the Environmental Impact of Delivering AI at Google Scale
11. Joule — Energy Use of AI Inference, Efficiency Pathways, and Test-Time Scaling
12. ACM FAccT — Power Hungry Processing: Watts Driving the Cost of AI Deployment?
13. International Energy Agency — Key Questions on Energy and AI
14. Riccardo Passoni and colleagues — Diffused Responsibility: Analyzing the Energy Consumption of Generative Text-to-Audio Diffusion Models (2025; accepted at WASPAA 2025; arXiv version)
15. Lawrence Berkeley National Laboratory — 2024 United States Data Center Energy Usage Report
16. Lawrence Berkeley National Laboratory — United States Data Center Energy Usage Report: 2025 Update (June 2026)
17. Sam Altman — The Gentle Singularity
18. Mistral AI — Our Contribution to a Global Environmental Standard for AI
19. Nature Sustainability — Environmental Impact and Net-Zero Pathways for Sustainable Artificial Intelligence Servers in the USA
20. Nature Communications — The Environmental Sustainability of Digital Content Consumption
21. Google — AI Overviews: About AI Overviews and More
22. AGU Advances — Data Centers Water Footprint: The Need for More Transparency (2026)
23. ACM e-Energy — Balancing Bits and Drops: Stress-Adjusted Water Management for Data Centers (2026)
24. Yuelin Han, Pengfei Li, Adam Wierman and Shaolei Ren — Small Bottle, Big Pipe: Quantifying and Addressing the Impact of Data Centers on Public Water Systems (2026) — arXiv preprint
25. U.S. Geological Survey — Colocating Artificial Intelligence Data Centers with Energy Infrastructure on Federal Public Lands: A Science Synthesis and Spatial Analysis to Inform Decision Making (2026)
26. U.S. Environmental Protection Agency — Water Reuse Case Study: Quincy, Washington
27. U.S. Department of Commerce / NIST — Final Programmatic Environmental Assessment for Modernization and Expansion of Existing Semiconductor Fabrication Facilities under the CHIPS Incentives Program (2024)
28. Resources, Environment and Sustainability — Environmental Data and Facts in the Semiconductor Manufacturing Industry: An Unexpected High Water and Energy Consumption Situation (2023)

29. U.S. Environmental Protection Agency — Electrical & Electronic Components (40 CFR Part 469) Detailed Study Report (2022)
30. Journal of Engineering for Sustainable Buildings and Cities — Data Center Waste Heat as an Emerging Urban Thermal Hazard: First Field Measurements of Neighborhood-Scale Air Temperature Impacts (2026)
31. Local Heat Islands and Vegetation Losses Are Microclimatic Consequences of Global Data Centers (2026) — Preprint
32. Spatially Concentrated Heat Emissions from Data Centers Lead to Localized Urban Warming and Heat Exposure (2026) — Preprint
33. Interstate Commission on the Potomac River Basin — Data Centers and Water Use in the Potomac River Basin (2026)
34. ACM COMPASS — The Cloud Next Door: Investigating the Environmental and Socioeconomic Strain of Datacenters on Local Communities (2025)
35. PLOS Water — Four Water Insecurity Concerns About Datacenters Driving the AI Revolution (2026)
36. Next 10 / Santa Clara University — The Intersection of Data Center Development, Water Availability, and Environmental Justice in California (2026)
37. City of The Dalles v. Advance Local Media LLC — Settlement Agreement (2022)
38. PJM Interconnection — 2025 Year in Review: Planning Prepares for Burgeoning Electricity Demand (January 8, 2026)
39. Arizona Department of Water Resources — Drought Management and Colorado River Water Supply
40. Microsoft — Datacenters in Arizona
41. Microsoft — Sustainable by Design: Next-Generation Datacenters Consume Zero Water for Cooling
42. Google — Assessing Watershed Health in Data Center Host Communities
43. Microsoft — Datacenters in Washington
44. Array — Green AI Techniques for Reducing Energy Consumption in AI Systems (2026)
45. U.S. Department of Energy — COOLERCHIPS 1.5 Program: Categorical Exclusion Determination CX-271071 (August 26, 2026)
46. U.S. Environmental Protection Agency — Basic Information about Water Reuse
47. Google — Our Commitment to Climate-Conscious Data Center Cooling (November 21, 2022)
48. National Renewable Energy Laboratory — Electricity Generation Baseline Report, NREL/TP-6A20-67645 (January 2017)
49. International Energy Agency — Energy and Water
50. Microsoft — 2025 Environmental Sustainability Report

51. Meta — 2025 Sustainability Report and 2024 Volumetric Water Benefits
52. Amazon — Water Stewardship / AWS Water Positive
53. World Resources Institute — Volumetric Water Benefit Accounting 2.0: Guidance for Implementing, Evaluating, and Claiming Volumetric Water Benefits of Water Stewardship Projects (2025)
54. International Energy Agency — Efficiency Should Always Be the First Answer (November 19, 2018)
55. Julien Delavande, Sasha Luccioni and Régis Pierrard — How Much Power Does a SOTA Open Video Model Use? Hugging Face community benchmark (July 2, 2025)
56. Tampa Bay Regional Planning Council — Data Center Debrief: Information & Considerations for the Tampa Bay Region (2026)
57. Tampa Bay Water — Regional Drinking Water Supply System
58. Southwest Florida Water Management District — 2025 Regional Water Supply Plan: Tampa Bay Planning Region
59. Florida — Large-Scale Data Center Permitting, Florida Statutes § 373.262; Chapter 2026-65, § 7, independent OPPAGA study requirement (2026)
60. Hillsborough County — Board of County Commissioners Meeting Recap, August 5, 2026
61. Environmental Science and Ecotechnology — Global Data–Water Symbiosis Reduces AI Infrastructure's Carbon and Water Footprint (2026)
62. U.S. Environmental Protection Agency — Water Reuse for Industrial Applications Resources
63. U.S. Environmental Protection Agency — Florida Water Reuse Requirements for Industrial Applications
64. Virginia Joint Legislative Audit and Review Commission (JLARC) — Data Centers in Virginia (December 9, 2024)
65. Berkeley Lab — Electricity Rate Designs for Large Loads: Evolving Practices and Opportunities, 2026 Update
66. UNEP/U4E — Sustainable Procurement Guidelines for Data Centres and Servers (2025)
67. Cheyenne Board of Public Utilities — Timeline, FAQ & Public Records Regarding Significant Noncompliance
68. Florida Department of Environmental Protection — Submerged Lands and Environmental Resources Coordination Program
69. Florida Department of Environmental Protection — Environmental Resource Permit e-Permitting
70. Atlanta Journal-Constitution — Georgia Families Sue Meta, Accuse Its Data Center of Tainting Their Water (2026)

71. Reuters — Pollution from Musk's unpermitted xAI power project hits hardest in Black communities (July 14, 2026)
72. Ceres — Drained by Data: The Cumulative Impact of Data Centers on Regional Water Stress (2026)
73. Mississippi Commission on Environmental Quality — Agreed Order: MZX Tech LLC, Southaven (July 2026)
74. Southwest Florida Water Management District — District Extends Modified Phase III Water Shortage (September 22, 2026)
75. Tampa Bay Water — August Rainfall Boosts Regional Water Supplies, Long-Term Deficits Remain (September 4, 2026)
76. Tampa Bay Water — Tampa Bay Region Enters Extreme Regional Water Supply Shortage: Tampa Bay Water Asking Residents to Step Up and Save Water (news release dated March 5, 2026; posted March 6)
77. Alana Lackner, Columbia Gorge News — The Dalles settles Oregonian lawsuit, releases Google water records (December 14, 2022)
78. Mike Rogoway, The Oregonian/OregonLive — Google's water use is soaring in The Dalles, records show, with two more data centers to come (December 17, 2022; reprinted by WaterWatch)
79. Google — 2025 Environmental Report, pp. 110 and 114, water-use tables and methodology.
80. April Ehrlich, Oregon Public Broadcasting — The Dalles' mayor called OPB's data center story inaccurate. Here are the facts (January 23, 2026)
81. PJM — PJM Proposes Reliability Standards to Manage Large Load Disconnection Events (September 10, 2026)
82. Hillsborough County — Hillsborough County Water Resources to Implement New Water and Wastewater Rates on Oct. 1, 2026 (September 15, 2026)