The Potential for Geothermal Energy to Meet Growing Data Center Electricity Demand

March 11, 2025
Ben King (bking@rhg.com), Wilson Ricks, Nathan Pastorek (npastorek@rhg.com), John Larsen (jwlarsen@rhg.com)

Growing electricity demand from new data centers, particularly driven by the explosive growth of artificial intelligence (AI), has quickly become an important topic in energy and technology circles. Though there is uncertainty as to just how big this boom in demand could be, what's clear is that electric power utilities and policymakers are taking the possibility of this growth seriously and are contemplating a range of solutions to meet this growing need. Continuing American leadership in AI—a major priority of the federal government across administrations—relies in part on ensuring there is sufficient electricity to meet load growth from new data centers. Major technology companies have ambitious greenhouse gas and clean electricity targets that mean this new load growth should come from clean sources.

Next-generation geothermal has a number of advantages in meeting this new load growth, including high capacity factor output, wide geographic dispersion, and the sheer amount of subsurface energy available to harness. In this note, we estimate how much electric demand at data centers could potentially be served by geothermal over the next decade. There are multiple promising next-generation geothermal technologies; we focus on the use of behind-the-meter enhanced geothermal systems (EGS) because of its broad geographic availability as well as the public availability of modeling data. We find that if the growth patterns of data centers follow historical clustering trends, geothermal could economically meet up to 64% of expected demand growth by the early 2030s under our baseline assumptions. If data centers locate in areas with the best geothermal resource, geothermal has the potential to meet all projected data center demand growth at prices 31-45% lower than in a clustered approach. Policymakers, technology companies, and geothermal developers need to act quickly to achieve the speed and scale required to meet this opportunity, and we outline policy changes like

improvements to permitting processes that would be necessary to do so. Geothermal could be a key solution to meeting the growing electricity needs of data centers.

Current growth in data center demand and expectations for the future

Rapid growth of electricity demand from data centers has quickly come to the forefront as a potential major challenge facing the US power sector today. This increase in demand is largely coming from the construction and use of data centers for training artificial intelligence (AI) models and providing AI services, particularly in the form of generative AI like ChatGPT, xAI, Meta AI, Google Gemini, and Microsoft Copilot. Combined with other sources of electricity load growth, this surge in demand has the potential to push the limits of the power grid, with utilities and grid operators contending with the need to connect a lot of new generating capacity and keep the system in balance after decades of effectively flat load growth.

Over the past few years, electricity use at data centers has grown rapidly but from a relatively small base. As reported in a December 2024 Lawrence Berkeley Lab (LBL) publication on data center energy use, several organizations have estimated recent historic growth in data center electricity demand and found compound annual growth rates (CAGR) on the order of 20-25% in the early 2020s. Put another way, the share of all US electricity consumption coming from data centers has increased from around 2% in 2020 to around 4.5% in 2024—though exact figures here are difficult to come by given a lack of consistently reported data (Figure 1).

FIGURE 1
Historical and expected electricity demand growth from data centers, 2014-2030
Billion kilowatt-hours
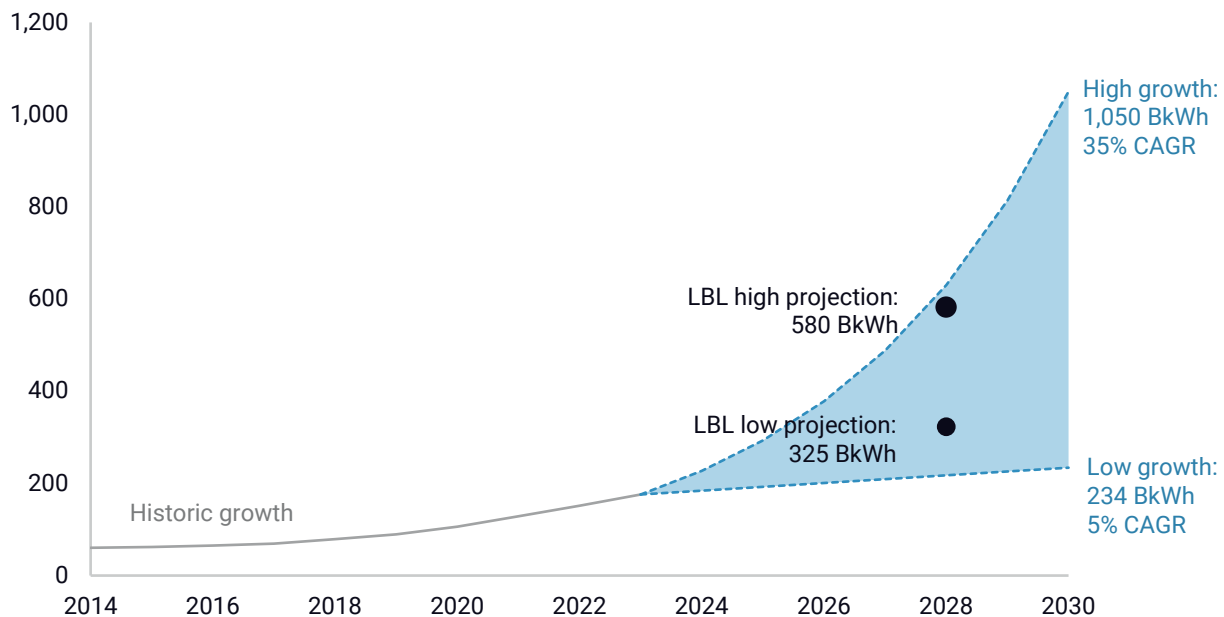
Source: Rhodium Group, Lawrence Berkeley Lab

There's even greater uncertainty about how much data center demand will grow in the future. The same LBL data center report projects that total data center electricity demand could increase to represent 7-12% of total electricity demand by 2028, representing roughly a 15-35% CAGR from today. A host of groups including research organizations, consulting firms, and investment banks have produced a wider range of estimates for growth through 2030 ranging from a relatively modest 5% CAGR to a sustained level of growth of 35% or more, aligned with the upper end of the LBL projection.

Major sources of uncertainty in estimating future data center electricity consumption include the ability for data centers to actually get access to the electricity they would need to run—the topic of this note—as well as availability of AI chips, the potential for efficiency gains at data centers (especially for energy-intensive end uses like cooling), the potential for model efficiency gains in training and inference (the use of models to provide AI services), the amount of speculative or double-counted projects in demand forecast calculations, the degree to which fast follower model developers provide lower-cost AI services that undercut the business models of frontier model developers, and the level of overall sustained investment in the sector.

The recent emergence of DeepSeek and o3-mini, which claim considerably lower training costs and electricity use than past models, sheds light on another part of the ongoing conversation: does a more efficient model necessarily mean less total electricity use, or does it just mean more affordable and widespread access to AI services leading to much more demand for AI inference—an example of the Jevons paradox.

Another related factor in the future composition of future data centers is the balance between AI training versus inference. Current model development approaches require a model to be trained in a single data center—so a bigger, more capable model requires a bigger data center and a higher concentrated increase in load growth. Once a model is trained, inference calls are flexible and need not be geographically concentrated. If future approaches to AI are more weighted toward inference than training, as is the case with DeepSeek, the level of electricity demand could be roughly the same, but the flexibility and relative dispersion of that load could make it easier to meet.

One commonly cited source for very high load growth expectations is utility and grid operator forecasts. While these projections are important data to consider and can be particularly helpful in shedding light on regional pinchpoints, caution is also warranted. Utility forecasts for future demand increases from data centers may overstate the level of growth in the aggregate as data center developers may explore siting their facilities in multiple locations, leading to double counting at a national level.

Despite uncertainty about the absolute level of growth in data center electricity demand, a few things are clear. First, sustaining American leadership in AI is a major, bipartisan policy priority—both the Biden and Trump administrations issued executive orders with that goal in mind.

Second, major technology companies that are driving this boom in AI, including Amazon, NVIDIA and Meta, have greenhouse gas (GHG) and clean electricity commitments. A few companies, including Microsoft and Google, have more ambitious goals of running on 100% clean electricity matched on an hourly basis by 2030. Large electricity buyers benefit from having more options available for procuring clean generation to meet their commitments, particularly as some companies have seen recent increases in emissions from their electricity consumption. Continuing to procure zero-carbon electricity from all

sources can help these companies make progress toward these goals, and sourcing electricity from clean sources that have high availability, like geothermal, can make it much easier to meet these 24/7 targets as well.

Finally, demand from data centers is far from the only source of growth for electricity. Accelerated electrification in the transportation and buildings sectors, adoption of electrified technologies in industry, and overall housing stock and economic growth are all driving higher demand for electricity. Under current policy, we previously estimated that total electricity demand in the US would grow by 24-29% in 2035 relative to 2023 levels, and that data centers only accounted for about a quarter of this growth. And beyond 2035, nascent but electricity-hungry technologies like direct air capture and electrolytic hydrogen could drive even faster demand growth. With this background, figuring out how to meet growing electricity demand from data centers is useful in and of itself, but it's also an important near-term test run for sorting out geothermal's potential for meeting longer-term demand growth while continuing to drive down power sector GHG emissions.

Solutions to data center demand growth

There are a host of potential solutions to meet or mitigate potential future large growth in electricity needs from data centers, each with associated advantages and challenges. A non-exhaustive list of these solutions includes:

▪ **Build new generating resources:** The most common response to meeting growing demand in the power sector is constructing new grid-connected generators. The US has added an average of just under 40 GW of utility-scale generation and energy storage each year since 2019, with the vast majority of that new generation coming from zero-emitting generation like solar, wind, and batteries. But new generation faces challenges including lengthy interconnection queues, long waits in sourcing equipment (including solar panels and gas turbines), and policy uncertainty. Zero-emitting resources like wind and solar are inherently variable, and while well-planned combinations of generators and energy storage can operate at relatively high capacity factors, they still don't achieve the 24/7 availability required by many data centers. In light of the long wait times to connect solar and batteries to the grid, some tech companies are turning to building new gas generation to meet data center demand, pulling in the opposite direction of their decarbonization commitments. Also, while new gas capacity can usually operate when called upon to do so, there are GHG and conventional pollutant emissions associated with their operation. As we discuss in this note, next-generation geothermal can provide clean, firm electricity, resolving challenges facing other types of new generating resources.

▪ **Keep existing fossil generation online:** Some existing fossil-generating assets that are slated for retirement, especially coal plants, could instead be kept on the grid and continuing generating. Because this generation is already constructed and connected to the grid, there would be little or no delay in using these assets to meet data center demand in the near term. But coal plants produce high levels of greenhouse gas and conventional pollution emissions, and their continued operation is inherently at odds with decarbonization.

- **Restart shuttered nuclear plants:** Microsoft made headlines when it inked a 20-year power purchase agreement to take electricity from a reactor at Three Mile Island. Around the same time, DOE finalized a loan guarantee to aid with the restarting of the Palisades nuclear plant. Depending on the nuclear plant, restarting shuttered plants may take some time—TMI is due to come back online in 2028, about four years after the Microsoft deal—but they may be available faster than some new generating resources. Nuclear is an also attractive solution because of its high capacity factor and the lack of GHG emissions at the point of generation. This approach is also not cheap, with TMI's owner Constellation indicating a $1.6 billion cost to get the plant operational again, and there are only around 8 GW of recently shuttered facilities that could feasibly be restarted.

- **Build off-grid resources:** To avoid lengthy delays associated with connecting both new generators and new data centers to the grid, data center developers can also consider behind-the-meter energy solutions to meeting their own electricity demand. Interesting new work proposes that off-grid solar microgrids could quickly meet new electricity demand at data centers with relatively low levels of GHG emissions at relatively competitive costs.

- **Demand shifting:** Recent analysis found that the US grid as it currently exists could accommodate tens of gigawatts of new data center load if these facilities were to be operated flexibly, reducing electricity use during periods of peak demand on the grid. AI data centers generally operate at very high capacity factors as the economic incentive is to use the very expensive computing equipment as much as possible, so this integration would require a shift in data center operations. As discussed above, a shift toward more AI inference at these data centers could enable greater demand flexibility.

In all likelihood, a combination of many of these solutions will be needed to meet growing electricity demand from data centers as well as other new sources of load. In the rest of this note, we discuss the use of behind-the-meter geothermal power generation to provide electricity to hyperscale data centers—a promising approach that has the potential to play a significant role in meeting this load growth.

Unpacking the role and promise of geothermal

Geothermal energy leverages huge amounts of heat under the Earth's surface to spin a turbine that generates electricity or to directly provide heating and cooling. Geothermal is particularly promising as a source of electricity generation given its renewable nature, low or zero greenhouse gas emissions, and high availability—geothermal power plants generally have a capacity factor of 90% or higher, meaning they can generate electricity on a near-constant basis. This is a useful complement to other renewable technologies like wind and solar that produce electricity on a variable basis.

Today, there is around 4 gigawatts (GW) of geothermal nameplate capacity on the US power grid. This capacity is all in western states and is highly concentrated in California and Nevada. Effectively all operating geothermal plants in the US today are based on conventional hydrothermal technology, in which wells are drilled to produce hot water or steam from naturally permeable reservoirs up to a few kilometers underground. That produced fluid is used to spin a turbine to generate electricity and is typically then injected

back underground in whole or in part. In dry-steam plants, high-temperature fluid is produced in the form of pure steam and passed directly into a turbine to generate electricity and afterwards is vented directly into the atmosphere or cooled and condensed into a liquid for reinjection into the reservoir. In flash-steam plants, the produced fluid is fully or partially liquid and is depressurized after being brought to the surface, causing it to turn partially to steam, which is then utilized similarly to a dry-steam plant. In binary-cycle power plants, the fluid typically remains a liquid throughout the process as it transfers heat energy to another working fluid before being reinjected underground as part of a closed loop. Conventional hydrothermal power generation technologies are well-demonstrated and commercially available but limited by the prevalence of suitable geological conditions.

By contrast, next-generation geothermal technologies can be economically deployed in a much wider swath of the country because they are not reliant on these natural subsurface conditions. While surface facilities are often similar to those used at conventional geothermal power plants, next-generation technologies utilize advanced drilling and reservoir engineering techniques to enable extraction of geothermal heat from rock formations that do not naturally host hydrothermal reservoirs.

This note focuses on the use of enhanced geothermal systems (EGS), one type of next-generation geothermal, in which hydraulic fracturing and horizontal drilling techniques are used to create fractures through which fluid can be injected to be warmed by the Earth's heated rock formations. Because EGS does not require a naturally permeable hydrothermal reservoir like conventional geothermal, it is able to tap into heat across broader swaths of the globe. As with conventional geothermal, this fluid is produced to the surface to generate electricity before being reinjected. Promising developments at the US Department of Energy's [Frontier Observatory for Research in Geothermal Energy](#) (FORGE) and [at wells drilled by geothermal company Fervo](#) have shown rapid improvements in drilling times for EGS, thereby reducing the cost of drilling these wells. Data from these sources also enables a solid techno-economic analysis of EGS potential.

EGS isn't the only new geothermal technology with promise for meeting future power needs. Though out-of-scope for this note, closed-loop geothermal systems (CLSG), also sometimes called advanced geothermal systems (AGS), are another type of next-generation geothermal. In CLGS, the working fluid is circulated through a series of subsurface loops before being produced at the surface rather than circulating openly through subsurface fractures. [Recent analysis](#) from the US Department of Energy found that next-generation geothermal can scale to 90-300 GW of installed capacity by 2050, representing a substantial expansion over today's level. This represents a small portion of [the country's technical potential](#), with 7 TW of capacity accessible at depths less than 5 km and more than 70 TW accessible across all depths—about one-eighth of total global geothermal potential.

Quantifying the potential role of geothermal to meet data center energy demand

In this analysis, we quantify the amount of behind-the-meter EGS electricity generating capacity that could be used to meet growing electricity demand from hyperscale data centers. Though this represents only a portion of data center demand and one geothermal

solution to meeting this demand, we believe this approach demonstrates geothermal's potential for a couple of reasons.

First, most of the future growth in data center electricity demand is projected to come from AI training and inference at hyperscale data centers—the largest data centers operated by major tech companies like Amazon, Google, and Microsoft—so focusing on meeting this need is the most critical to achieving continued American leadership in the space. Second, our focus on behind-the-meter deployments is driven by the fact that geothermal installations are subject to the same delays in interconnecting to the grid that we discuss above for other types of resources, with 1.7 GW of new capacity waiting to secure an interconnection agreement—representing a potential 40% increase in total installed geothermal capacity in the US. Bypassing this lengthy wait and instead plugging directly into data centers that need electricity can meaningfully accelerate how quickly new generating capacity can serve this load. Geothermal power is uniquely well-positioned for behind-the-meter applications due to its clean, firm nature and low surface land requirements.

There are two key components to our analysis: determining where data centers will be constructed and, thus, where electricity needs to be delivered; and determining the extent to which geothermal electricity can meet that need.

ESTIMATING DATA CENTER ELECTRICITY DEMAND

Beginning with Taking Stock 2024, Rhodium Group has been tracking growth expectations for data centers and developing geographically granular projections for incorporation into our modeling. For this analysis, we use a revised version of the high growth scenario in Taking Stock that results in about 80 GW of total data center peak demand online in 2030, representing a 22% CAGR over 2022 levels. We assume that roughly 27 GW of this demand growth comes from hyperscale facilities. As discussed above, there is a high degree of uncertainty in projecting future demand from data centers, and we don't pretend to have a crystal ball. Still, we believe this level of demand provides us a useful analytical starting point for this project, and we estimate how much of this demand can be met by new geothermal capacity in the early 2030s.

To test the value of geothermal in a range of potential data center developer decision-making environments, we construct three different approaches to the geographic distribution of these facilities:

▪ **Clustered:** Until recently, the location of data centers was driven by a few key factors, including proximity to a connection to the fiber optic network and proximity to other data centers to minimize latency and availability of land on which to build. This has led to significant clustering of data centers in places like Northern Virginia, New York and its suburbs, the San Francisco Bay area, Atlanta, Phoenix, and several Texas cities. Under our historic demand growth approach, we assume this trend continues and calculate regional data center growth as a function of current installed data center capacity, commercial projections of regional market growth, specific announced hyperscale projects, and proximity to a fiber node. Using this approach, we find 28 markets with at least 100 MW of projected data center demand growth, with the highest concentrations in many of the locations noted above. The top five biggest growth markets using this approach are Northern Virginia, Phoenix, Dallas, Atlanta, and Las Vegas.
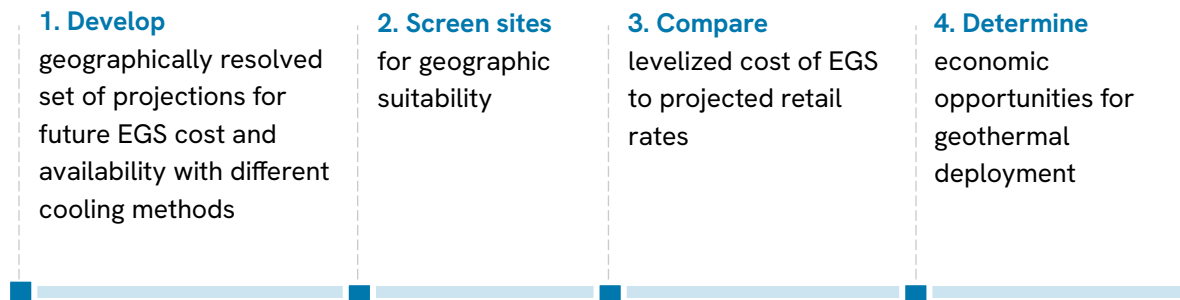
- **Geothermal-driven**: Increasingly over the past couple of years, a major factor in data center siting decisions is simply access to sufficient amounts of electricity. A behind-the-meter geothermal approach to providing electricity inherently solves this problem. In our geothermal-driven siting approach, we assume data center developers elect to locate their facilities wherever they have access to the most affordable geothermal energy.

- **Geothermal-plus-fiber driven**: Our fiber-constrained siting approach starts with the geothermal-driven approach described above, but limits the locations in which data centers can be built to within a 25-kilometer radius of a fiber node, as sourced from the International Telecommunications Union Infrastructure Connectivity Map.

ESTIMATING GEOTHERMAL OPPORTUNITY

To determine the extent to which geothermal can meet the electricity needs of these data centers, we first estimate the cost and availability of EGS across the continental US. Then, we compare the levelized cost of behind-the-meter EGS with the retail rate of electricity that we estimate the data center would otherwise pay. We find economic opportunity where the cost of EGS is equal to or lower than the expected retail rate plus an expected green premium. We do not run a full capacity expansion model to compete behind-the-meter EGS with other grid-connected resources; rather, we find places where behind-the-meter EGS would compete with expected grid prices.

FIGURE 2
Analytical approach to assessing geothermal opportunities

1. Develop geographically resolved set of projections for future EGS cost and availability with different cooling methods	**2. Screen sites** for geographic suitability	**3. Compare** levelized cost of EGS to projected retail rates	**4. Determine** economic opportunities for geothermal deployment

To develop a highly geographically resolved set of projections for future EGS cost and availability, we use a peer-reviewed EGS project cost model that incorporates cost and performance data from recent EGS demonstration projects to produce empirically grounded near-term technology cost projections under specified reservoir design assumptions. We use this cost model to build a national EGS supply curve representing near-term (circa 2030-2035) project costs under a baseline technology development scenario. Estimated costs include the impact of Inflation Reduction Act investment tax credits.

Highly spatially resolved temperature-at-depth data is a critical input to this cost model. Throughout this note we report results using two different sources for this data. One data set we use comes from Project InnerSpace's Geothermal Exploration Opportunities Map (GeoMap), which uses surface temperatures and both corrected and non-corrected bottom hole temperature datasets, combined with a newly developed lithospheric thermal model to define temperature gradients and predict temperatures at depths of 1-

5 km within the continental US. For further details on the methodologies employed, refer to Fullea et al. (2021), the GeoMap user manual, and the GeoMap tool. We also use temperature-at-depth data from a recently published article by Stanford researchers Aljubran and Horne, which utilizes similar (though not identical) inputs and a graph neural network approach to predict temperatures at depths of 0-7 km within the continental US. We restrict both sets of temperature maps with geothermal development exclusion areas developed for the DOE's GeoVision report and other land use constraints, including the exclusion of areas where land surface gradients are too steep to allow co-located data center construction.

Because near-field resources are promising prospects for near-term EGS development, we also calculate supply curves for these resources at depths of 2-5 km based on a recent estimate of near-field resources at nine large known hydrothermal sites across the western US. Both the Project InnerSpace and Stanford temperature-at-depth datasets are produced using methodologies that assume heat flow occurs by conduction through rock. The Project InnerSpace dataset also incorporates advection into its modeling, which overrides the conductive model in areas of volcanism. Because of these differences in approach, we used Project InnerSpace's conduction-only model for non-near-field resources and used the approach described above for all near-field EGS development.

The output profile and cost of a geothermal plant is sensitive to assumptions on what cooling method is used, but we assume data center demand is flat. As such, we use a version of the GenX capacity expansion model to develop the capacity of EGS and other on-site resources needed to supply a flat data center load curve in different regions of the country under three approaches:

- **Wet-cooled EGS**, where the electric output is assumed to be flat and no flow rate optimization or additional resources are needed.

- **Dry-cooled EGS with optional on-site batteries**, where overbuilding the geothermal plant, flexible operation, and batteries installed on-site flatten the electric production profile, which would otherwise be variable due to changes in air temperature.

- **Dry-cooled EGS with optional on-site batteries and solar**, which adds the potential of on-site solar installations of up to 300 MW to complement the EGS production profile.

The final EGS supply curves for each temperature-at-depth dataset include both cost and developable EGS capacity for each surface grid node and depth in the temperature-at-depth dataset. Where appropriate, the cost in these supply curves also includes any additional costs associated with on-site resources (i.e., batteries, solar) beyond EGS.

Finally, for each EGS resource and data center demand scenario, we screen the EGS supply curves against data center demand regions to identify the deployment options that can support the projected load at least cost. We compare these costs to projected future electricity rates that would be paid by these data centers were they to be developed and grid connected. To do so, we use future electric rates from Taking Stock 2024, our current policy energy system projection, to which we add a green premium of 0-50% depending on the scenario.

Geothermal can play a major role in meeting growing electricity demand from data centers
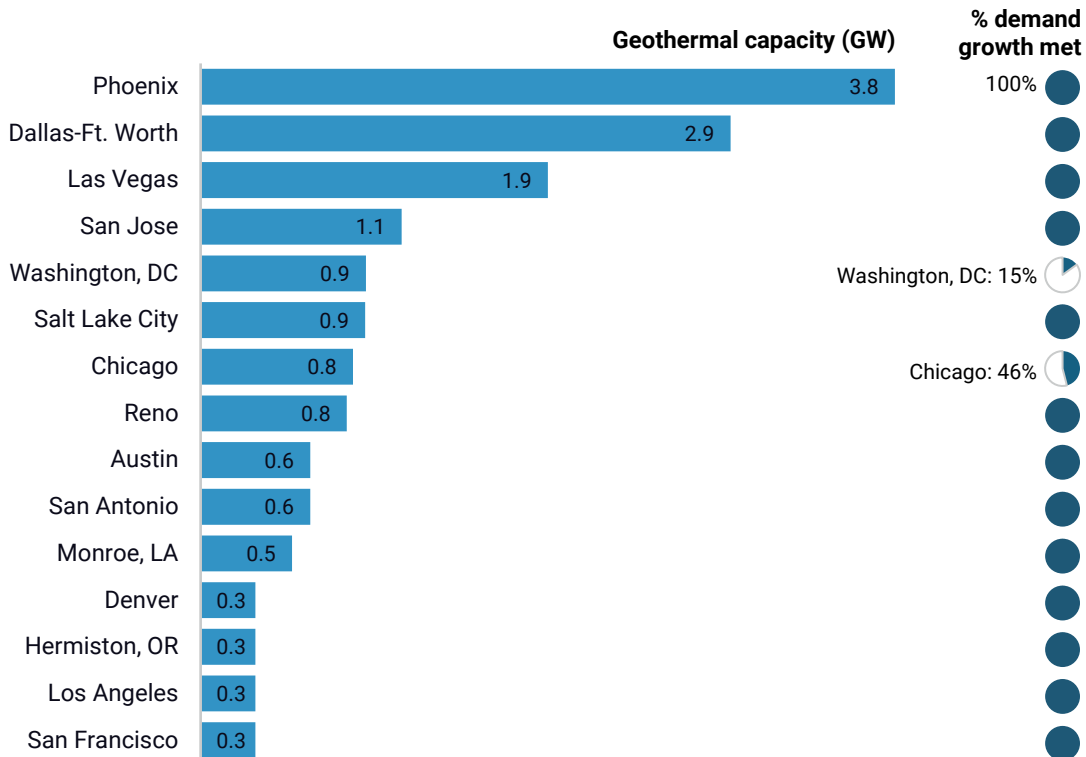
Across the scenarios we consider, we find that behind-the-meter geothermal could meet a large share of demand when co-located with new hyperscale data centers. The extent to which this is the case varies depending on the siting strategy used for building data centers as well as attributes of the geothermal installations, which we unpack below.

Promising opportunities in most regions in the clustered approach, mostly in the West

When we assume that data center siting follows a similar pattern as it has in the past, in which large clusters of data centers develop based on proximity to population, available land, and the fiber optic network, we estimate that 55-64% of our projected growth at hyperscale facilities could be met with behind-the-meter geothermal, representing 15-17 GW of new data center and geothermal capacity.

Geothermal can meet 100% of anticipated data center demand growth in 13 of the 15 largest markets using the most promising cooling approach (Figure 3), and it can meet at least 15% of power needs in 20 of 28 markets nationally. Much of the opportunity we find is in the West, but geothermal can provide power for some or all growth even in growth markets in the central and eastern US including the Washington, DC/Northern Virginia cluster, Chicago, Columbus, OH, and Memphis. Among major growth markets, only Atlanta and New York City don't show meaningful promise for behind-the-meter geothermal.

FIGURE 3

Behind-the-meter geothermal potential at data centers in 15 largest markets, clustered siting approach

GW, % of total projected demand growth

	Geothermal capacity (GW)	% demand growth met
Phoenix	3.8	100%
Dallas-Ft. Worth	2.9	
Las Vegas	1.9	
San Jose	1.1	
Washington, DC	0.9	Washington, DC: 15%
Salt Lake City	0.9	
Chicago	0.8	Chicago: 46%
Reno	0.8	
Austin	0.6	
San Antonio	0.6	
Monroe, LA	0.5	
Denver	0.3	
Hermiston, OR	0.3	
Los Angeles	0.3	
San Francisco	0.3	

Source: Rhodium Group. Note: The chart represent capacities using dry cooling plus optional solar and battery cooling approach with Project InnerSpace temperature-at-depth maps.

The extent to which geothermal can meet growing data center demand in the clustered approach is sensitive to assumptions around what green premium a data center developer is willing to pay for clean, firm electricity and which cooling method is used. In our baseline scenario, which we discuss above, we assume that data centers pay a 20% green premium above the regional retail rate for electricity they'd otherwise face. This results in a national weighted average levelized cost of electricity (LCOE) for the new geothermal capacity of $78-85/MWh, though there is considerable regional variation. Regional average prices range from $49-111/MWh. These prices are generally within the range of prices in recent geothermal power purchase agreements.

FIGURE 4

Behind-the-meter geothermal potential at data centers, by green premium scenario and cooling method
GW
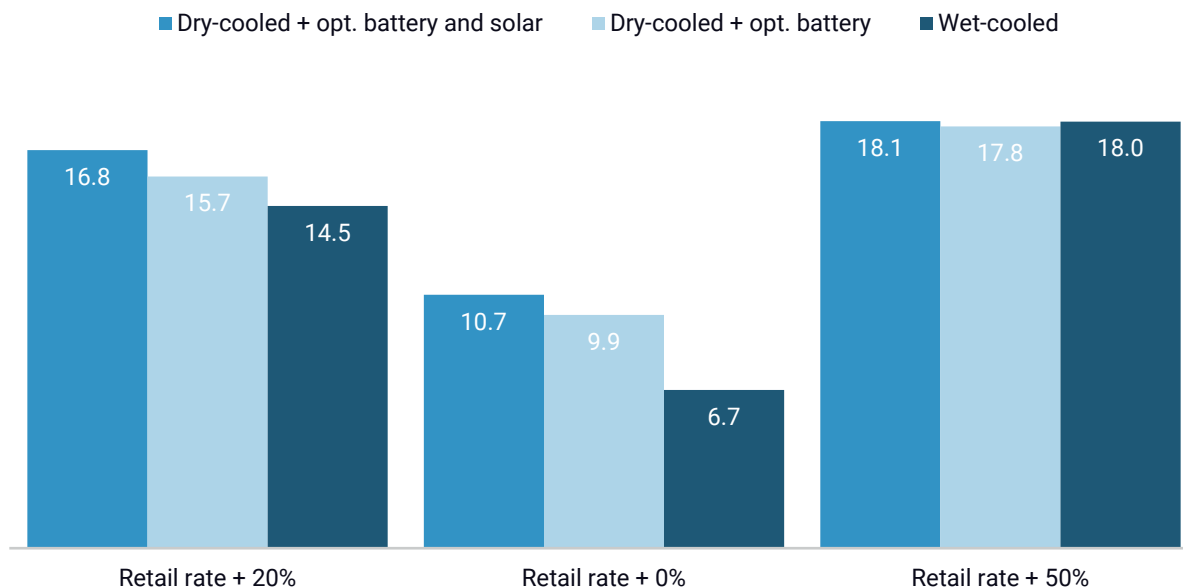
Source: Rhodium Group. Note: The chart represents capacities using Project InnerSpace temperature-at-depth maps.

Within the 20% green premium scenario, the difference in national deployment levels is mostly driven by slightly higher prices for the wet-cooled approach in the Phoenix market pushing some potential geothermal installations to no longer be economically competitive (Figure 4). A higher willingness to pay, in this case represented as a 50% green premium, yields consistent increases in economic deployment of geothermal across all cooling scenarios. At a 50% cost premium, all available geothermal sites that provide capacity up to the expected level of data center demand are used. Conversely, with 0% green premium, the economics of geothermal deployment no longer make sense in several key markets, especially Phoenix and the Washington DC/Northern Virginia region, leading to 38-55% lower geothermal deployment.[1]

Using geothermal potential to inform data center siting leads to higher levels of geothermal deployment

When data center developers make siting decisions based on using the best-available geothermal resources, as described above in the geothermal-driven and geothermal-plus-fiber-driven siting approaches, we find geothermal could easily meet all projected data center load growth in the early 2030s.

By opting to build data centers in regions where they can leverage the highest quality geothermal resources in a self-supply approach, data center developers and operators

---

[1] Using Stanford temperature-at-depth data yields directionally similar results. In the 20% green premium baseline, we find 56-62% the national deployment we report for the Project InnerSpace temperature-at-depth maps. At a 50% green premium, this increases to 76-89% of capacity reported above. Regionally, 13 of 15 top markets for deployment are the same between the two cases.

face even lower costs for geothermal electricity. The national capacity-weighted average levelized cost of electricity (LCOE) of new geothermal capacity built at data centers in the geothermal-driven and geothermal-plus-fiber-driven siting approaches are 31-45% lower than LCOEs in the clustered siting approach (Figure 5). Incorporating near-field EGS sites results in a further 1-6% reduction in LCOEs. As discussed above, the geothermal LCOEs in the clustered siting approach are already on the low end of average regional retail rates; following the best geothermal resource in the geothermal-driven and geothermal-plus-fiber-driven approaches are clear economic winners.

FIGURE 5
National average levelized cost of geothermal electricity by deployment approach
2022 $/MWh
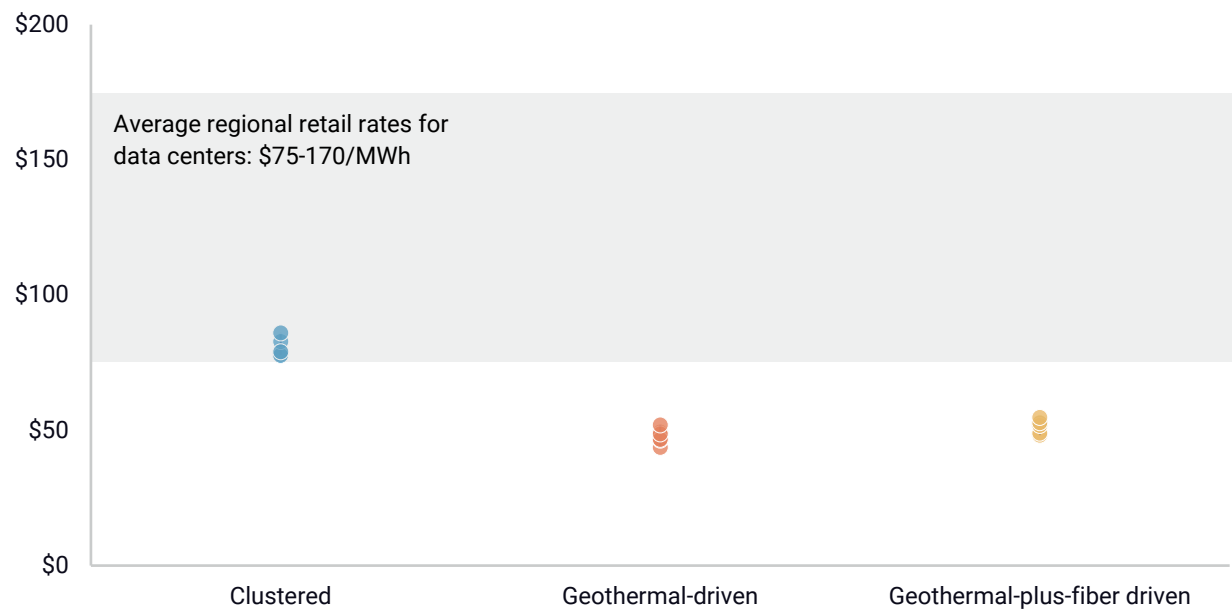
Source: Rhodium Group. Note: individual points represent national average LCOE under comparable scenarios between the three siting approaches.

As in the clustered approach results, much of the deployment under the geothermal-driven and geothermal-plus-fiber-driven siting approaches focuses on key states in the West including California, Colorado, New Mexico, Nevada and Oregon (Figure 6). These states are consistently among the top targets for geothermal deployment regardless of which siting scenario or temperature-at-depth map we used. Differences in assumptions around cooling type, use of near-field EGS resources, and temperatures at depth drive the variance within states between cases. Depending on these assumptions, some parts of the country that have been less of a focus for geothermal development also appear promising, including midwestern states like Indiana and Ohio, as well as Texas.

FIGURE 6

Behind-the-meter geothermal potential at data centers, geothermal-driven and geothermal-plus-fiber-driven siting approaches
GW
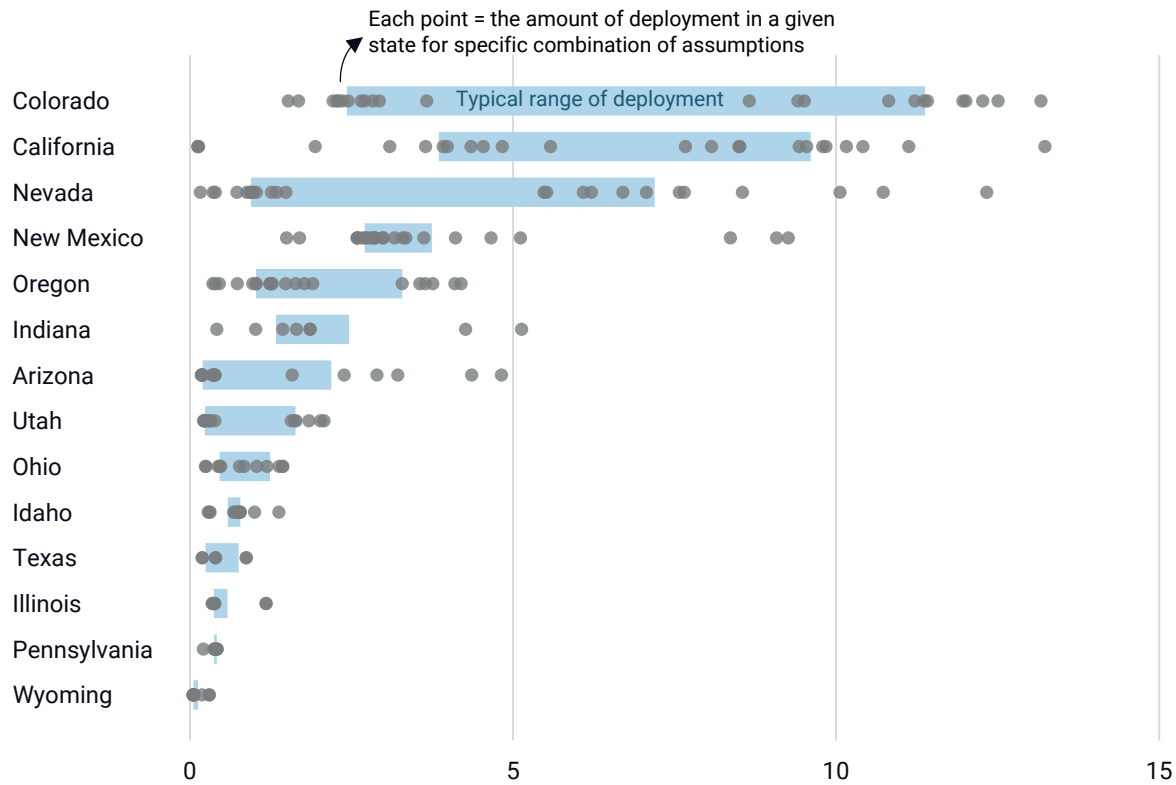
Source: Rhodium Group. Note Each point on the chart represents new behind-the-meter geothermal capacity for a given combination of assumptions (e.g., cooling approach, temperature-at-depth map). The bars represent the 25th through 75th percentile of deployment among all considered cases.

Direct geothermal cooling can complement geothermal power generation

Geothermal energy can be used to do more than generate electricity. The use of geothermal to provide direct cooling for data centers can cut overall data center demand, reducing the challenge of securing electric supply. Currently, most large data centers use adiabatic or liquid cooling systems along with air- or water-cooled chillers for space cooling. These systems, which largely run on electricity, have large-scale cooling capabilities but also can have high energy demand—as much as 40% of total data center energy consumption. AI data centers in particular generate more heat compared to traditional data centers as the GPUs used in them require more energy per rack and therefore release more heat. This increase in heat requires more efficient cooling systems to keep energy consumption as low as possible while protecting sensitive equipment.

Directly using the geothermal resource, rather than converting it into electricity to power traditional cooling approaches, holds promise, and a range of innovative geothermal approaches have the ability to cool data center systems. Geothermal heat pumps, which leverage the consistently cool temperatures found between a few and few hundred feet

below the surface, use a network of underground pipes to efficiently cool off water or a working fluid and provide cooling for the IT components of a data center. The presence of shallow aquifers in some regions can provide access to similar temperatures at shallower depths. Geothermal absorption chillers use low-grade geothermal heat found at shallower depths than those needed for power generation to drive a chemical reaction that produces water vapor. This water vapor cools as it is run through a condenser and cools the IT components of a data center using evaporation. Absorption chillers could also use some of the residual heat from the EGS process rather than purpose-drilled wells. These approaches could also integrate other emerging approaches to efficiently cooling data centers, particularly the use of liquid cooling for chips.

Employing direct geothermal cooling techniques reduces the overall electric load of a data center, leading to cost savings by shrinking the size of the geothermal system that needs to be built to generate electricity. This reduces the need for the deep geothermal wells and can have a substantial effect on overall price as drilling costs comprise a major portion of the all-in price of EGS. Geothermal heat pumps also have relatively low water consumption as they are often closed-loop systems. Many large data centers using water-cooled chillers today require millions of gallons of water annually, which could present a challenge in some areas of the country that have promising geothermal resources but are water-constrained.

Geothermal heat pumps are not widely installed but have a foothold in the US in residential and commercial buildings. Geothermal energy has also been used in wider heating and cooling settings at large facilities like university campuses and research facilities. Part of the reason for the lack of wider acceptance is the relatively high capital cost of such installations relative to other cooling methods, but tax credits and utility programs can help close this gap—particularly for data centers seeking to reduce electricity consumption and minimize emissions. The use of direct geothermal cooling at hyperscale data centers is still nascent—though companies are beginning to explore such opportunities. Further work can explore the economic case as well as possible limitations, including the potential for heat saturation and limits on efficiency improvements in some climates.

Enabling the growth of geothermal to meet demand from data centers

This analysis has identified the extent to which geothermal could play a major role in helping meet growing electricity demand coming from data centers given the clean, firm nature of geothermal power and an appetite and willingness to pay for such power from data center companies. This presents a sizable and lucrative demand-side pull for geothermal developers: the levels we've discussed represent a 2-7-times increase in the amount of installed geothermal power generating capacity beyond what's on the grid today in less than 10 years. Moreover, access to electricity is already constraining the growth of data centers in some regions of the US today, and that constraint will only grow—especially if data centers expand at the fastest expected rates. Put another way, this is an issue that needs to start to be addressed today. As such, taking advantage of this opportunity requires the geothermal industry to dramatically increase both the *speed* at which it can build new facilities as well as the *scale* at which these new facilities are built. Changes to policy at all levels of government can help enable this growth.

Streamline permitting

This biggest obstacle to increasing the speed of deployment is navigating the challenging permitting process. Geothermal projects on federal land can take 7-10 years to build, with a large portion of that time spent on securing permits and approvals from eight or more federal agencies. Projects on state or private land can take nearly as long to develop and require engaging with a multitude of state and local agencies in addition to the potential for federal interactions. In addition to slowing down the process of bringing a project online, these lengthy approval chains can also add to the overall cost of a given project. And as fewer projects are deployed, opportunities for cost reductions from learning by doing are also reduced, also contributing to higher costs than would otherwise be faced.

Next-generation geothermal technologies may be somewhat less affected by permitting constraints, as they enable deployment of much more generating capacity at a single development site. Still, streamlining state and federal permitting processes to minimize overlap and facilitate concurrent review wherever possible will be important steps to accelerating the speed at which geothermal projects can be deployed, as will adding more institutional capacity to the relevant approving agencies. There are some encouraging early signs of progress here. The Railroad Commission of Texas recently approved its first geothermal well permit in about six months. At the federal level, the Trump administration has ordered a review of permitting processes with a goal of removing barriers and reducing delays. Chris Wright, the new Secretary of Energy, is prioritizing geothermal innovation and permitting and building energy infrastructure.

Clarify resource rights and behind-the-meter regulations

Several other potential permitting-related roadblocks affect geothermal's ability to meet demand with speed and at scale. As geothermal developers look to expand beyond states in which they've built projects in the past, they may encounter state laws that don't—or don't clearly—identify ownership of the subsurface geothermal resource. These definitional issues may also impact a developer's access to water rights. In 2023, Texas passed legislation clarifying heat ownership and the regulating entity, providing a model for other states attempting to scale the industry.

Other state regulations may also come into play. Specifically, state utility regulators generally have oversight of in-state electricity sales, and the approach to this oversight varies considerably from state to state. States may need to clarify laws relating to the ability of data centers to use behind-the-meter generation so that such arrangements don't trigger regulation of the transaction as a utility-customer relationship.

Improved data

Rapidly scaling geothermal, especially in geographies that haven't been considered key locales for deployment in the past, requires access to high-quality, granular data characterizing subsurface conditions. Maps with less geographic granularity may miss important hot spots that could support meaningful levels of geothermal capacity. Efforts to collect this type of data and make it widely available, like the development of Project InnerSpace's GeoMap, can help open up new areas for consideration, reduce the length of the early stages of project development timelines, and enable more successful exploration.

Scale up supply chains

Building out a robust supply chain will also be necessary to build at the scale we find in our analysis. Limitations could arise at multiple points across the supply chain, including the availability of drilling equipment and pipes, shortages of turbines and other componentry for the surface power plant, and even a potential shortage of skilled labor. On the materials side, though the equipment market is currently dominated by foreign producers, federal and state policymakers could expand current policies like the federal advanced manufacturing tax credit to cover geothermal equipment or enact state-level policies to attract local production of these components. On the labor side, the geothermal industry can look to leverage considerable US expertise in oil and gas production.

Invest in further innovation

Continued federal leadership across research, development, and demonstration (RD&D) will also play a critical role in enabling the rapid scale-up of geothermal. This includes a sustained focus on early-stage R&D breakthroughs that drive innovations in drilling, materials, and site design optimization. On average, over fiscal years 2021 through 2023, the US invested $116 million per year on enhanced geothermal RD&D. Maintaining and expanding annual investment levels has the potential to accelerate geothermal innovation. The federal government also has a role to play in supporting early demonstration-scale facilities to provide important data and learning as well as improved confidence for investors. These demonstration projects would ideally be situated across a range of potential types of geothermal sites, and some could be focused explicitly on behind-the-meter provision of power including to data centers.

Maintain federal incentives

Finally, continued federal support for deployment of these technologies will also be critical. Today, geothermal facilities qualify for tax credits that play an important role, especially in improving the economics of early projects. These early installations matter a lot, as that's where much of the learning takes place that helps drive down costs across the industry. If Congress or the Treasury were to restrict access to these credits, or repeal them altogether, early projects may struggle to secure financing and price output competitively.

A promising path forward

In this note, we've identified the compelling economic opportunity that exists for behind-the-meter geothermal power generation to meet growing electricity demand from hyperscale data centers. In parts of the West traditionally associated with geothermal, but also in other pockets where data center clusters may develop, behind-the-meter geothermal can cost-competitively provide clean, firm power while helping data centers avoid challenges of plugging into the grid. The opportunity can be further enhanced by considering direct geothermal cooling to efficiently meet the substantial cooling needs at these facilities.

But this path is far from guaranteed, and data center developers, tech companies, and geothermal developers need to start working today to realize this potential. Supportive enabling policy from all levels of government will also be critical to achieving this outcome.

As we said earlier in this note, predicting the long-term growth trajectory of data centers is exceedingly uncertain. But working to pair data centers and geothermal represents a no-regrets strategy regardless of that trajectory. If data center growth exceeds our wildest expectations, geothermal is available to meet a large portion of that growth—and there's plenty of headroom to meet even more load than we've anticipated with favorable economics. On the other hand, if data center load growth sputters, more experience building geothermal facilities and scaling the industry can help the resource position itself for a critical role in the larger energy transition. In either case, the rich geothermal resource in the US, which accounts for as much as one-eighth of total global technical potential, alongside early investments in next-generation geothermal to help drive down costs can help the US maintain its global leadership in geothermal.

ABOUT THE AUTHORS

Ben King is an Associate Director with Rhodium Group's Energy & Climate practice, where he researches the effects of policy and economic changes to the US energy system. Prior to joining Rhodium, Ben was an analyst in the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE), where he worked on demand-side efficiency analysis and electricity market policy. Ben also served as Deputy Chief of Staff in EERE and developed clean energy policy resources for cities and states at the U.S. Environmental Protection Agency.

Wilson Ricks is an energy systems researcher at Princeton University and an expert in next-generation geothermal power. He has authored multiple peer-reviewed studies on the role of geothermal power in future energy systems and was a contributor to the US Department of Energy's *Pathways to Commercial Liftoff: Next-Generation Geothermal Power* report.

Nathan Pastorek is a Research Analyst with the Energy & Climate practice at Rhodium Group, working on research projects relating to emerging clean technologies and decarbonization. He previously worked as an intern with the National Renewable Energy Laboratory (NREL) focused on projects relating to geothermal energy resources and for KeyLogic Systems focused on economic and geologic modeling for carbon capture and storage scenarios.

John Larsen is a Partner at Rhodium Group and leads the firm's US energy system and climate policy research. John specializes in the analysis of national and state energy and climate policy, market trends, and emerging clean technologies. He manages a multidisciplinary team of energy modelers, policy specialists, and systems analysts focused on accelerating America's transition to a net-zero economy. Previously, John worked for the US Department of Energy's Office of Energy Policy and Systems Analysis, where he served as an electric power policy advisor. Before working in government, John led federal and congressional policy analysis in the World Resources Institute's Climate and Energy Program.

ABOUT RHODIUM GROUP

Rhodium Group is an independent research provider with deep expertise in policy and economic analysis. We help decision-makers in both the public and private sectors navigate global challenges through objective, original, and data-driven research and insights. Our key areas of expertise are China's economy and policy dynamics, and global climate change and energy systems. More information is available at www.rhg.com.

New York | California | Washington, DC | Paris

Website: **www.rhg.com**

Rh^g **Rhodium Group**