

Data Centers in the News, 2019-2026: Attention, Geography and Tone of Global Data-Center Coverage, and How It Compares with Research and Invention

Fliar BI · Prism Research Layer · Working paper, October 2026 · Dataset and code: github.com/fliarbi/prism-news-api · Licence CC BY-NC 4.0

Abstract

We measure how data centers were covered in a global news index from January 2019 to July 2026, using 160,171 de-duplicated data-center articles, and set that coverage beside 13,096 research papers (2015-2025) and 15,994 patent families (2015-2024) on data centers. Six results stand out:

1. **Attention.** Data centers took 15.14 of every 10,000 news articles in 2025, 3.15 times their 2019 share (4.8). Most of the rise came after 2023 (7.26).
2. **The 2026 level.** In the partial year January-July 2026 the share was 30.5 per 10,000, and May 2026 was the peak month at 52.16 per 10,000.
3. **Geography.** The United States holds 47.88% of located data-center news at a lift of 0.99. Among the 20 countries with the most articles, Finland (9.64), Taiwan (6.17), Malaysia (5.25) and Singapore (4.63) are among the most over-represented. Within the United States, Virginia (5.01), Wisconsin (3.03), Wyoming (2.87) and Texas (2.48) lead; within India, Andhra Pradesh (5.65), Tamil Nadu (2.91) and Telangana (2.17).
4. **Tone and urgency.** 84.8% of data-center articles are positive in tone (all news: 56.8%) and 14.2% are negative (all news: 33.6%). 69.1% are rated critical priority (all news: 42.6%).
5. **Measurement.** 15.1% of raw matches for data-center vocabulary were repeated page furniture, job listings or scientific archive senses. Counting them would inflate coverage.
6. **Research and invention.** Data-center papers fell 16% as a share of all papers from 2019 to 2022, then rose 1.43-fold to 2.25 per 10,000 in 2025. Patent families grew 2.75-fold from 2015 to 2024, and China's office held 66.6% of office filings.

Findings stated for citation

- **F1.** Data centers took 15.14 of every 10,000 news articles in 2025, 3.15 times the 2019 share (4.8).
- **F2.** The partial year January-July 2026 reached 30.5 per 10,000; May 2026 was the peak month at 52.16.
- **F3.** The United States held 47.88% of located data-center news (lift 0.99); India 9.86% (lift 0.75).
- **F4.** Among the 20 leading countries, Finland (9.64), Taiwan (6.17), Malaysia (5.25) and Singapore (4.63) were among the most over-represented.
- **F5.** 84.8% of data-center articles were positive in tone, 1.49 times the all-news share; 14.2% were negative, 0.42 times.

- **F6.** 69.1% of data-center articles were rated critical priority, 1.62 times the all-news share.
- **F7.** Green cloud storage and energy-efficiency initiatives appeared in 7.09% of themed articles, 592.1 times their frequency in all news.
- **F8.** 15.1% of raw data-center matches were page furniture, job listings or archive senses.
- **F9.** Data-center papers fell 16% in share from 2019 to 2022 and then rose 1.43-fold to 2025.
- **F10.** China's office held 66.6% of office filings of data-center patent families (2015–2024).
- **F11.** Data-center patent families published per year grew 2.75-fold between 2015 and 2024.
- **F12.** Within the United States, Virginia (5.01), Wisconsin (3.03), Wyoming (2.87) and Texas (2.48) had the most over-represented coverage.
- **F13.** Within India, Andhra Pradesh (5.65), Tamil Nadu (2.91) and Telangana (2.17) had the most over-represented coverage.
- **F14.** From 2023 to 2025 coverage as a share of national news rose in 24 of the 25 leading countries (United States 2.76-fold); Ireland alone fell.
- **F15.** Relative to national baselines, the negative share came closest to national news in the United Arab Emirates (0.79) and Ireland (0.67), and was furthest below it in India (0.12) and Nigeria (0.13).

1. Introduction

A data center is a building that houses computer servers, storage and network equipment, together with the power and cooling they need. Data centers carry cloud services, the internet's large platforms and, since about 2023, the training and use of large artificial-intelligence models. They draw large amounts of electricity and water, and their siting has become a local political issue in several countries.

News is one public record of this build-out: who builds where, how the facilities are powered and how communities respond. Coverage is rarely measured as data in its own right. Industry surveys count facilities and capacity. Bibliometric studies count research. Few sources count attention, meaning how much of the news data centers take and when, where and in what tone.

This paper and its dataset do that. They combine three records for one topic: a global news index, research papers and patents. Each comparison uses a denominator, so that growth in the index or in publishing is not mistaken for growth in interest. The companion dataset publishes an article-level sample (40,165 rows) and full coverage tables, so every figure here can be re-derived. The [Prism News API](#) serves the full history.

2. Data and methods

News. The Prism news index holds articles from outlets worldwide; non-English articles are indexed in English translation. Each article carries its opening text (the first few sentences), a primary location, a graded tone label, an editorial priority rating (1–9), themes, keywords and named organisations. We count unique article identifiers; repeated rows for the same article are removed. When this release was built, 92.88% of the index was readable, so all counts are lower bounds.

Matching. An article is about data centers when its opening text contains *data center*, *data centers*, *data centre*, *data centres*, *datacenter(s)* or *datacentre(s)* as whole words. Three families of matches are excluded:

- repeated page furniture: sidebars of headlines, site navigation and vendor pages that place the phrase in front of unrelated articles (one opinion-column teaser alone sat in front of tens of thousands of articles before it was excluded);
- job listings;
- scientific archive senses, such as climate, seismological or snow and ice data centers.

Together these removed 28,416 of 188,587 raw matches (15.1%). One reviewer read 40 randomly drawn matched openings. 36 referred to data centers as computing facilities that were part of the story (precision 90%), and in 24 (60%) data centers were the main subject. Fifteen openings were re-judged by a second reviewer. The four misses were a corporate board notice that names data centers only as a market in its company description, a supply-chain feature where the phrase sat outside the visible opening, a graphics-card rumour that uses the term only for a product segment, and an executive appointment where it appears only in a past job title.

Measures. Coverage over time is expressed per 10,000 articles in the index that month or year, because the index grew from 11,807,177 articles in 2019 to 26,990,006 in 2025. Year comparisons use 2019–2025; 2026 (January–July) is partial. Geography, tone, priority and themes use January 2023 – July 2026. Geography is expressed as lift: a place's share of data-center articles divided by its share of all located articles. 44.0% of articles in that window carry a primary country. Place tags derived from ordinary words (714 words such as "Congress", "Earth", "Winter", "Cloud", "Mobile" and "Hurricane") were treated as unlocated in every location field, which removed 15.3% of city-tagged articles.

Research and patents. Research papers are de-duplicated articles and reviews (2000–2025) with a data-center term in the title. Repository deposits and scientific archive senses (genomics, soil, nuclear reaction and climate data centers) are excluded. Patent families (2006–2024) use the same title rule and are counted once per filing office.

3. Results

3.1 How much and when

The index holds 160,171 data-center articles from January 2019 to July 2026. Raw counts rose from 5,665 in 2019 to 40,867 in 2025. The share of all news rose too, from 4.8 per 10,000 in 2019 to 15.14 in 2025, a 3.15-fold increase (Table 1). The share was nearly flat from 2021 to 2023 (7.4, 7.02 and 7.26) and then rose 2.09-fold in two years. In January–July 2026 it reached 30.5 per 10,000. The highest monthly share was May 2026, at 52.16 per 10,000 (Figure 1). Because the index thins after May 2026, the 2026 values rest on a smaller denominator and are not used in year comparisons.

Year	Data-center articles	All articles in index	Per 10,000
2019	5,665	11,807,177	4.8
2020	9,001	16,230,014	5.55
2021	12,786	17,281,220	7.4
2022	13,425	19,115,515	7.02
2023	14,863	20,484,506	7.26
2024	23,541	23,311,896	10.1
2025	40,867	26,990,006	15.14

Table 1. Data-center articles per year and as a share of the index. 2026 (January–July) is partial and omitted.

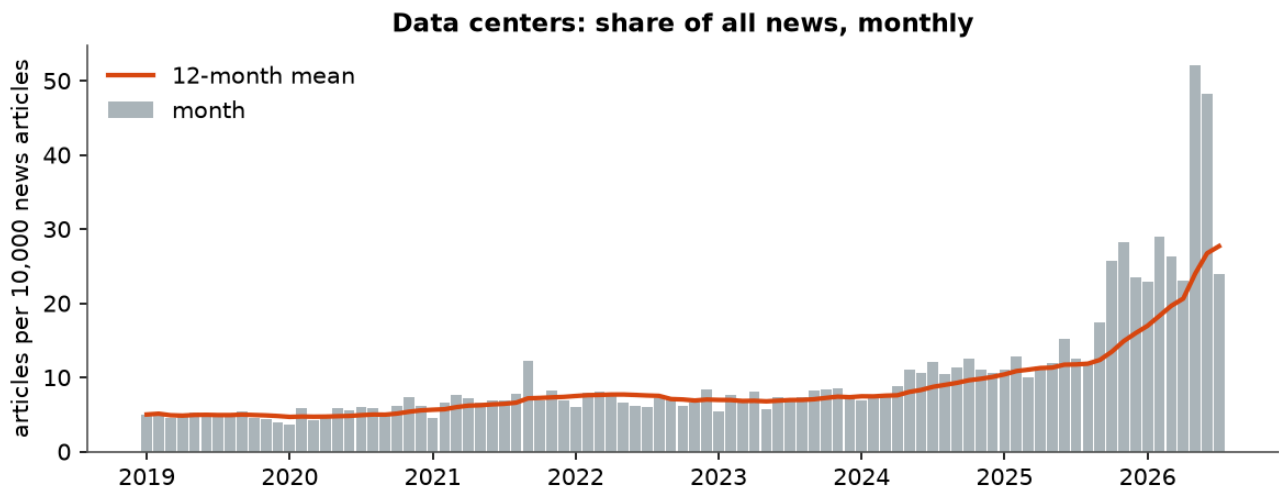


Figure 1. Data-center coverage as articles per 10,000 news articles, monthly, January 2019 – July 2026. Bars are months; the line is the 12-month mean.

Unlike weather topics, data-center coverage has little seasonality (Figure 2). September (10.26%), October (9.3%) and November (9.51%) together hold 29.07% of a typical year's articles, against 25% for an even spread. January is the quietest month (6.98%).

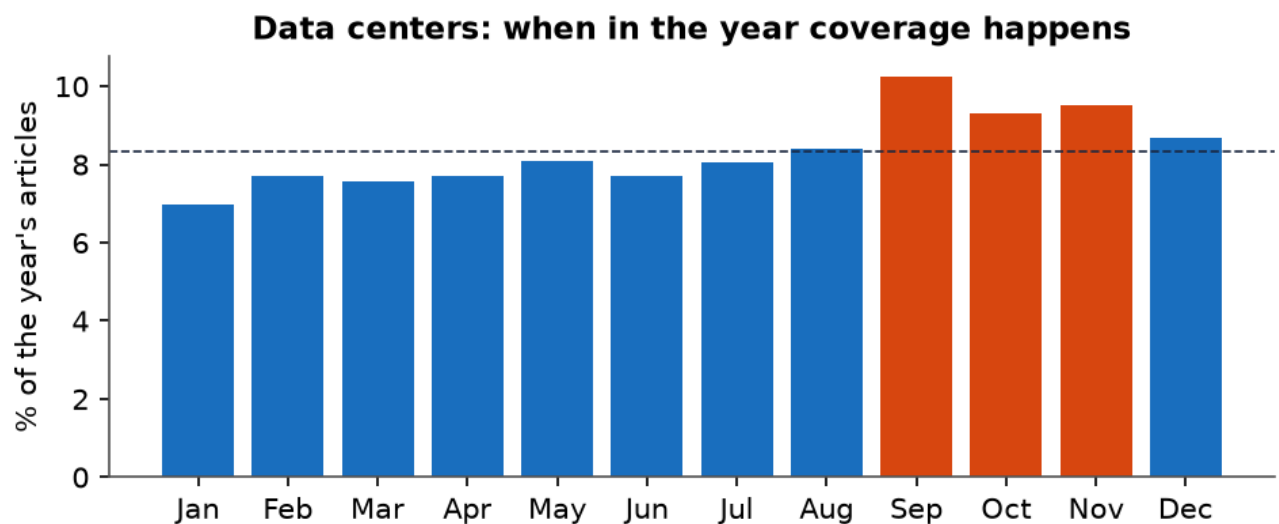


Figure 2. Share of each year's data-center articles published in each month, averaged over 2019–2025. Dashed line = an even spread (8.3% per month).

3.2 Where

Between January 2023 and July 2026, 52,492 data-center articles carried a primary country. The United States holds almost half of them (47.88%), close to its weight in located news

(lift 0.99). India's share (9.86%) comes with a lift of 0.75. Among the 20 countries with the most articles, the most over-represented are Finland (9.64), Taiwan (6.17), Malaysia (5.25) and Singapore (4.63), followed by the United Arab Emirates (2.43) (Table 2, Figure 3). Part of Finland's lift is a tagging artefact: close to half of its sample rows are tagged to the town of Nokia, a reading of the company name, so its lift is overstated, though its remaining rows (Helsinki, Hamina, Espoo, Kajaani) still point to a real hub. South Korea's lift (2.0) is not interpreted, because 110 of its 210 sample rows are tagged to the city of Samsung, a reading of the company name. Saudi Arabia, with fewer articles, also has a high lift (2.77). Colombia also shows a high lift (4.7), but nearly all of its sample rows are tagged to the Colombian department of Meta, a misreading of the company name, so it is not interpreted.

Country	Located data-center articles	Share %	Lift vs all news
United States	25,133	47.88	0.99
India	5,177	9.86	0.75
United Kingdom	2,934	5.59	0.96
Singapore	2,163	4.12	4.63
China	1,820	3.47	1.32
Australia	1,111	2.12	0.99
Japan	841	1.6	1.51
United Arab Emirates	832	1.59	2.43
Germany	786	1.5	1.27
Canada	733	1.4	0.65
France	719	1.37	0.69
Colombia*	640	1.22	4.7
South Korea*	630	1.2	2.0
Ireland	545	1.04	1.44
Nigeria	531	1.01	0.74
Spain	503	0.96	0.98
Malaysia	496	0.94	5.25
Finland*	496	0.94	9.64
Taiwan	430	0.82	6.17
South Africa	385	0.73	1.62

Table 2. The 20 countries with the most located data-center articles, January 2023 – July 2026, with lift against all located news. * Lift affected by company names read as places (Colombia: Meta; South Korea: Samsung; Finland, in part: Nokia).

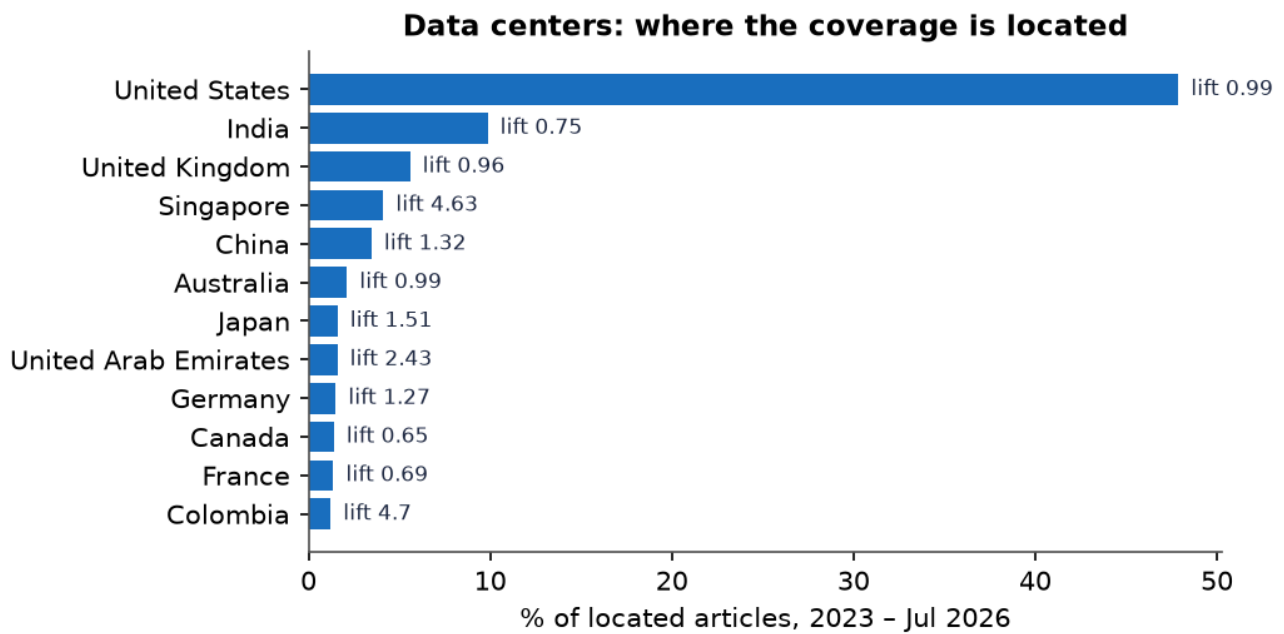


Figure 3. Leading countries by share of located data-center articles, 2023 – Jul 2026, labelled with lift against all located news.

3.3 Geo intelligence: continents, states and shifting attention

The world map (Figure 4) shows the pattern at a glance. Data centers are over-represented in the news of Northern Europe, East and Southeast Asia and the Gulf states, and under-represented in Israel, Brazil, Canada, Italy, France and India. Most African countries have too few located articles to measure. By continent the differences are small, because each continent mixes hubs and non-hubs: Asia (1.11) is slightly over-represented, Europe (0.99), North America (0.96) and Oceania (0.95) are close to parity, and Africa (0.79) is under-represented (Figure 7, left). South America's lift (1.31) is raised by the Colombia tagging artefact noted above.

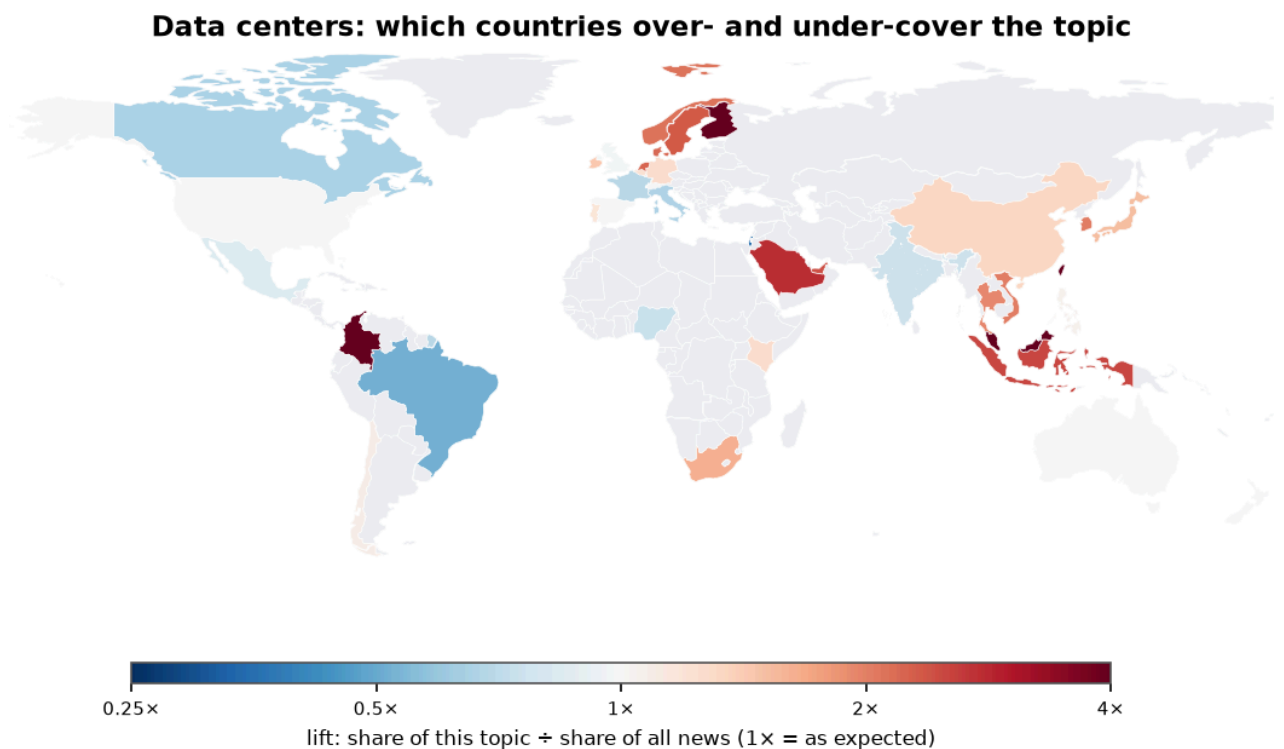


Figure 4. Lift of each country's share of located data-center articles, January 2023 – July 2026 (red = over-represented, blue = under-represented, grey = too few articles). India is drawn with its official outline.

Within countries, the picture is sharper (Table 3). In the United States, 19,236 located articles carry a state. Virginia has the highest lift, 5.01. Wisconsin (3.03), Wyoming (2.87), Texas (2.48), Arizona (2.39) and Ohio (2.11) follow. Oklahoma's lift (2.18) is not interpreted, because 81 of its 155 sample rows are tagged to the town of Blackwell, a reading of the name of a chip product line. New Jersey (0.1) and Florida (0.4) are far below their weight in American news, and New York (0.67) is under-represented despite 1,507 articles (Figure 6). Iowa, which ranked second before place tags from ordinary words were removed, now sits near parity (0.96).

In India, 3,388 located articles concentrate in the southern and western states. Measured against each state's share of Indian news, Andhra Pradesh (5.65), Tamil Nadu (2.91), Telangana (2.17) and Maharashtra (1.98) are the most over-represented. Maharashtra holds the largest raw share (31.46%). Delhi (0.4) and Uttar Pradesh (0.41) are under-represented (Figure 5). States with fewer than 50 articles are not interpreted.

India: state	Articles	Lift	United States: state	Articles	Lift
Andhra Pradesh	436	5.65	Virginia	943	5.01
Tamil Nadu	494	2.91	Wisconsin	778	3.03
Telangana	256	2.17	Wyoming	132	2.87
Maharashtra	1,066	1.98	Texas	2,809	2.48
Karnataka	300	1.6	Arizona	275	2.39
Gujarat	148	1.32	Ohio	865	2.11
Uttar Pradesh	176	0.41	Florida	417	0.4
Delhi	177	0.4	New Jersey	371	0.1

Table 3. States with the highest and lowest lift in data-center coverage, January 2023 – July 2026. Lift = the state's share of the country's located data-center articles ÷ its share of all located news from that country. The last two rows of each half are the two lowest-lift states with at least 50 articles. Oklahoma (2.18) is left out of the US half because most of its tags come from the chip name Blackwell.

Data centers: states of India by lift

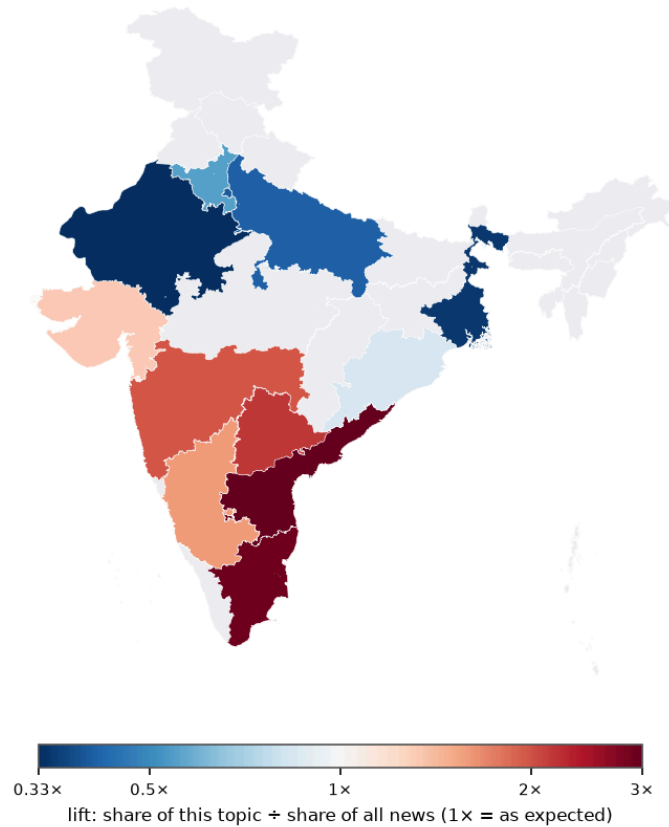


Figure 5. Lift of each Indian state's share of located data-center articles against its share of all located Indian news, January 2023 – July 2026. Official outline.

Data centers: states of United States by lift

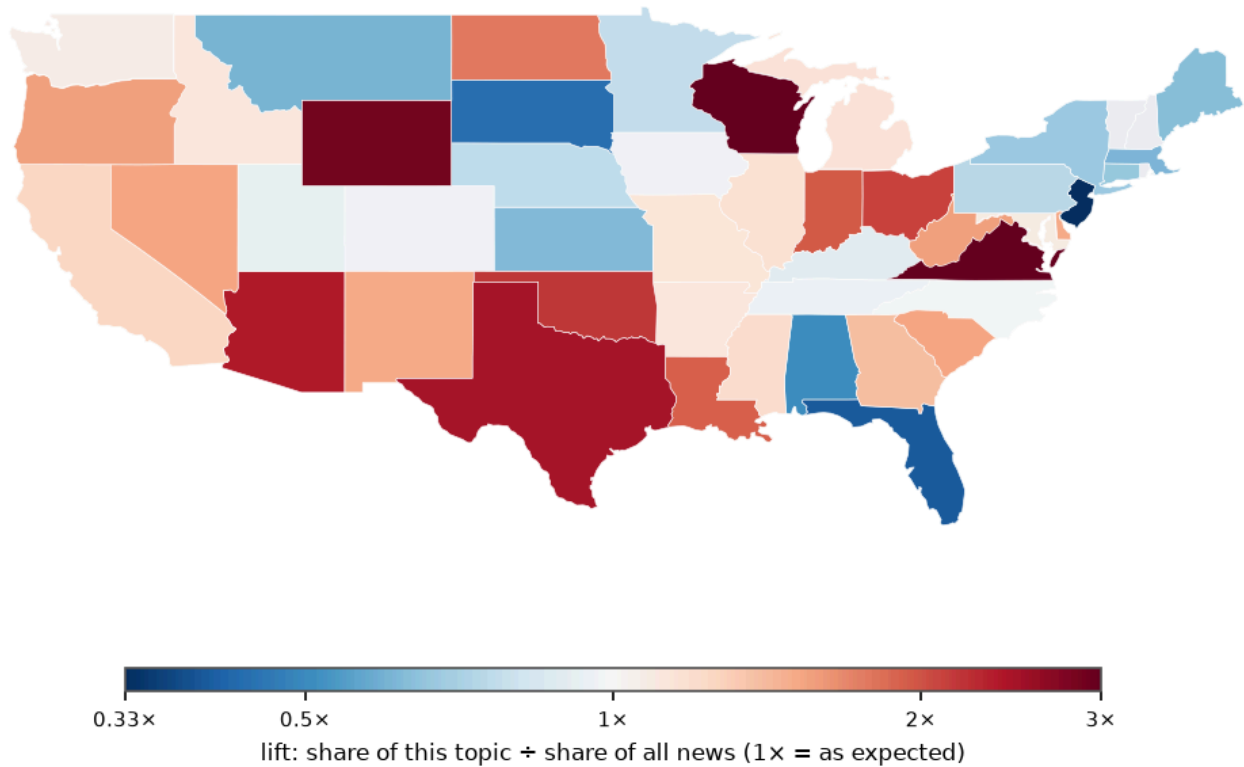


Figure 6. Lift of each US state's share of located data-center articles against its share of all located US news, January 2023 – July 2026.

Attention also moves between countries from year to year (Figure 7, right). Measured as data-center articles per 10,000 of each country's located news, the United States rose from 5.98 in 2023 to 16.51 in 2025 (2.76 times), the United Kingdom 2.56 times, Saudi Arabia 2.52 times, Finland 2.44 times, the United Arab Emirates 2.28 times and Australia 2.13 times. India rose 1.67 times, from 6.96 to 11.62. Colombia's apparent rise (4.92 times) reflects the Meta tagging artefact. Of the 25 countries with the most articles, only Ireland fell, from 17.88 to 10.98 (0.61 times), after 33.25 in 2024. China (1.14) and Japan (1.16) grew slowly, and South Africa barely changed (17.8 to 18.28). Among the five countries with the most articles, Singapore has the highest level throughout, at 61.6 per 10,000 in 2025; Finland's level (159.16) is higher still.

Tone varies by country even after allowing for each country's own news. Data-center coverage is less negative than national news in every country listed. Among the 15 countries with the most labelled articles, it comes closest to the national baseline in the United Arab Emirates (19.4% negative against 24.6%, a ratio of 0.79) and Ireland (22.8% against 34.0%, 0.67). In the United States the ratio is 0.57 (19.0% against 33.1%). It is lowest in India, where 5.0% of data-center articles are negative against 41.2% of Indian news (0.12), and Nigeria (5.6% against 44.4%, 0.13).

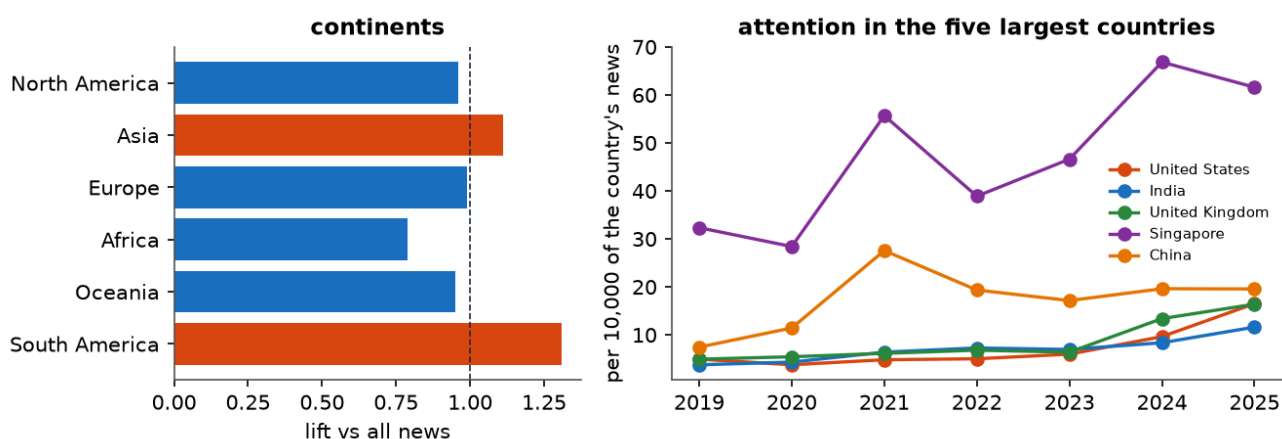


Figure 7. Left: lift by continent, January 2023 – July 2026. Right: data-center articles per 10,000 of each country's located news, 2019–2025, for the five countries with the most articles.

3.4 Tone, urgency and themes

Data-center coverage is more positive and more urgent than news in general (Figure 8). Among articles with a tone label, 84.8% are positive, against 56.8% across the index (1.49 times), and 14.2% are negative, against 33.6% (0.42 times). The machine-assigned priority rating puts 69.1% of data-center articles at critical level (7 or higher), against 42.6% of all news (1.62 times).

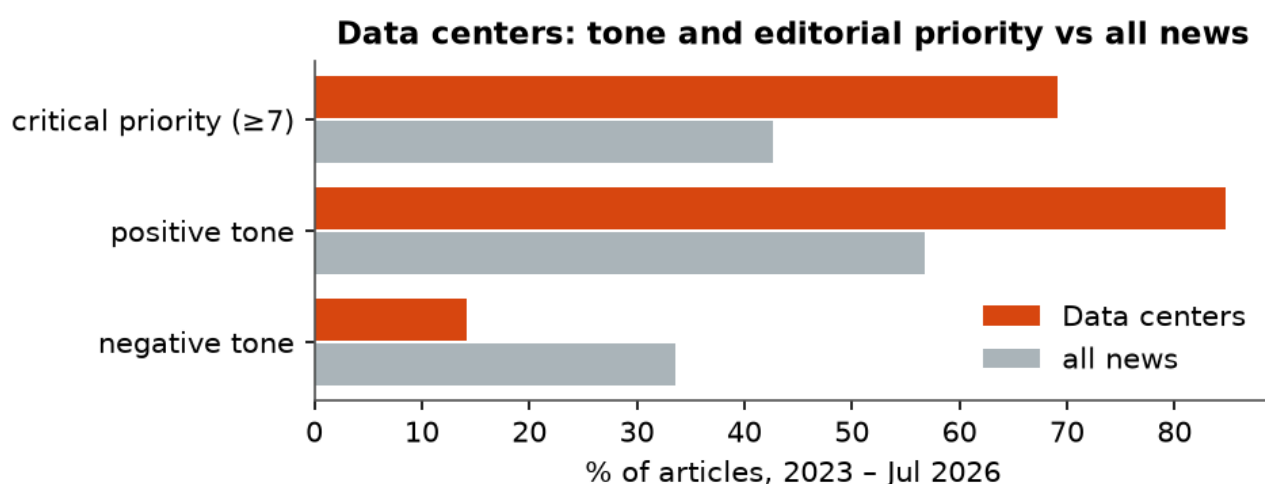


Figure 8. Tone and the share rated critical priority, data-center articles vs all news, 2023 – Jul 2026.

Themes show what data-center stories are about (Table 4). The most common themes are general business ones: product launches (15.24% of themed articles) and partnerships (13.75%). The most specific themes are about energy. Green cloud storage and energy-efficiency initiatives appear in 7.09% of themed articles, 592.1 times their rate in all news. Renewable energy (213.18) and the link between nuclear energy and artificial intelligence (177.38) are also highly specific to the topic.

Theme	Articles	Share %	Lift
green cloud storage and energy efficiency initiatives	4,313	7.09	592.1
renewable energy tech advancements	3,287	5.41	213.18
cloud storage for ai/ml workloads	675	1.11	207.97
green technology and sustainability	4,250	6.99	199.71
nuclear energy & artificial intelligence integration	161	0.26	177.38
server components (rack-mounted hardware)	1,427	2.35	174.66
cloud computing technologies	6,383	10.5	165.14

Table 4. Themes most specific to data-center coverage (lift against all themed news), January 2023 – July 2026.

The organisations named most often are technology companies: Microsoft (11,405 articles), Google (9,638), Nvidia (7,592), OpenAI (5,011), Meta (4,869) and Amazon Web Services (3,266). Specialist operators such as Equinix (2,331) and Digital Realty (1,664) follow. Generic labels also rank high, so the list describes who is named rather than who builds.

3.5 Research attention and invention

Research on data centers did not grow steadily. Papers with a data-center term in the title held about 1.82 per 10,000 papers in 2015 and 1.87 in 2019, fell to 1.57 in 2022 (16% below 2019), and then rose to 2.25 in 2025, 1.43 times the 2022 level. Over 2015–2025 the share grew 1.24-fold, while the total literature in the corpus grew by a factor of 1.0. Between 2015 and 2025, 13,096 such papers were published. The most-cited include work on energy-aware resource management in cloud data centers [1], data-center network traffic [2, 3] and global data-center energy use [4–6].

News attention and research attention diverged after 2023 (Figure 9). News share doubled between 2023 and 2025, while research share rose more slowly.

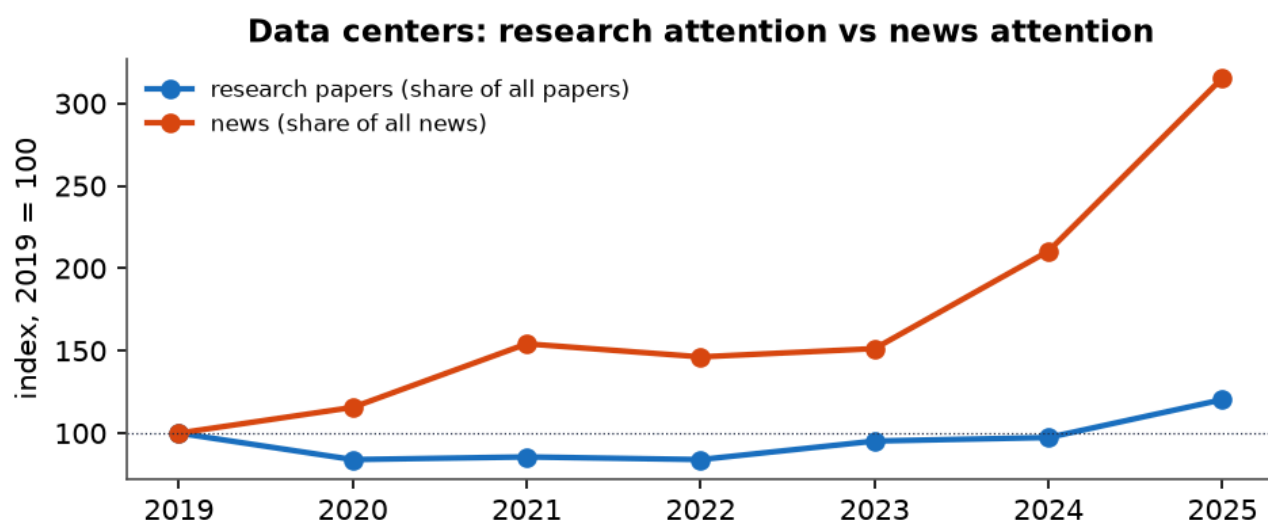


Figure 9. Research papers with data center in the title (share of all papers) and data-center news (share of all news), each indexed to 2019 = 100.

Patenting grew steadily: 852 data-center patent families were published in 2015 and 2,342 in 2024 (2.75 times), after 2,348 in 2023. China's office holds 66.6% of families from 2015 to 2024, followed by the United States (17.8%) and the international route (5.1%) (Figure 10). The leading technology classes are electric digital data processing (G06F, 2,006 families), electrical apparatus housings and cooling (H05K, 1,801) and digital transmission (H04L, 1,759).

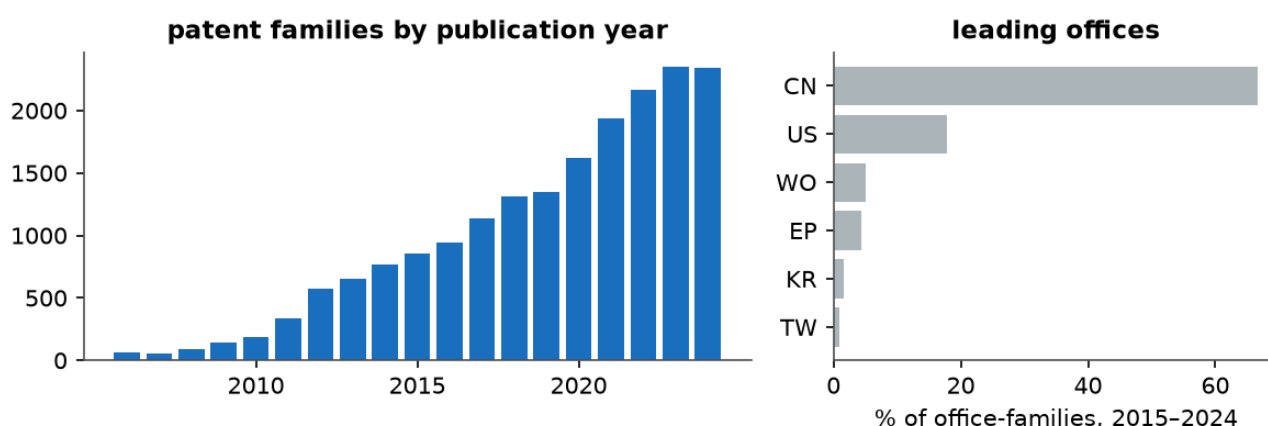


Figure 10. Data-center patent families by publication year (left) and the leading filing offices, 2015-2024 (right).

4. Discussion

Three points matter for anyone who uses news as a record of data centers. First, coverage measures attention. The share of news taken by data centers roughly doubled between 2023 and 2025, and the partial-year level for January-July 2026 was higher again. This is a change in what the press writes about. It is not a count of facilities, and the dataset does not test what caused it. Second, raw shares mislead. The United States dominates in raw terms but sits at its weight in the index. The places where data centers are most over-represented relative to all news are smaller hubs (Finland, Taiwan, Malaysia, Singapore) and, within countries, the states where facilities cluster: Virginia, Wisconsin, Wyoming and Texas in the United States, and Andhra Pradesh, Tamil Nadu and Telangana in India. Third, a keyword count needs cleaning. One site teaser alone repeated the phrase in front of tens of thousands of unrelated articles, and 15.1% of raw matches were removed.

Tone is a second lesson. Coverage is mostly positive, because much of it is corporate: launches, partnerships and investments. Negative coverage is a minority everywhere, but it is closest to national norms in the United Arab Emirates and Ireland. Ireland is also the only one of the 25 leading countries where attention fell between 2023 and 2025.

Research, invention and news run on different clocks. Research share dipped and recovered. Patent filings grew every year from 2015 to 2023 and are concentrated in one office. News share rose fastest of the three. The dataset lets researchers link them, for example to test whether heavy local coverage precedes planning decisions in the states where facilities are proposed.

5. Limitations

- **Index availability.** 92.88% of the index was readable at build time, so counts are lower bounds. Shares and lifts are less affected, because the missing part is spread across topics.
- **Coverage after May 2026.** The index thins after May 2026, so 2026 is partial and is excluded from year comparisons. The May 2026 peak rests on a smaller denominator than earlier months.
- **Matching.** Matching uses the article opening. Articles that mention data centers only later are missed, and about 10% of matches are not about data centers. Some repeated page furniture with unusual wording may remain.
- **Machine labels.** Location, tone, priority and themes are machine-labelled. The city field is noisy and is not analysed. Place tags derived from ordinary words were treated as unlocated in every location field, which removed 15.3% of city-tagged articles; this lowered some state lifts sharply (Iowa fell from second place to parity). Company and product names are still read as places: Meta as the Colombian department of Meta, Samsung as a South Korean city, Blackwell as a town in Oklahoma and, for part of Finland's rows, Nokia as a Finnish town. Colombia, South Korea and Oklahoma are therefore not interpreted, and Finland's lift is overstated.
- **Outlet mix.** The index over-represents English-language outlets. This shapes geography and tone.
- **Research and patent counts.** These match titles only. Office shares reflect filing practice.

Data availability

The dataset (article sample, monthly, country and state tables, research and patent tables, all metrics) is released under CC BY-NC 4.0 with this paper. The sample holds 40,165 of 160,159 eligible articles to 31 July 2026. Full article text, headlines, publishers, later articles and other topics are served by the Prism News API: <https://github.com/fliarbi/prism-news-api> (base URL <https://radar-api.fliarbi.com/news>). A worked query: `prism-news trends "data center" --since 2019-01-01 --until 2026-07-31 --granularity monthly`.

References

1. Anton Beloglazov et al. (2011). Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2011.04.017>

2. Theophilus Benson et al. (2010). Network traffic characteristics of data centers in the wild. <https://doi.org/10.1145/1879141.1879175>
3. Mohammad Alizadeh et al. (2010). Data center TCP (DCTCP). *ACM SIGCOMM Computer Communication Review*. <https://doi.org/10.1145/1851275.1851192>
4. Eric Masanet et al. (2020). Recalibrating global data center energy-use estimates. *Science*. <https://doi.org/10.1126/science.aba3758>
5. Miyuru Dayarathna et al. (2015). Data Center Energy Consumption Modeling: A Survey. *IEEE Communications Surveys & Tutorials*. <https://doi.org/10.1109/comst.2015.2481183>
6. N. Jones (2018). How to stop data centres from gobbling up the world's electricity. *Nature*. <https://doi.org/10.1038/d41586-018-06610-y>
7. Khosrow Ebrahimi et al. (2014). A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. *Renewable and Sustainable Energy Reviews*. <https://doi.org/10.1016/j.rser.2013.12.007>
8. Qixiang Cheng et al. (2018). Recent advances in optical technologies for data centers: a review. *Optica*. <https://doi.org/10.1364/optica.5.001354>
9. Fliar BI (2026). *Data Centers in the News 2019-2026: article-level dataset and coverage tables*. Prism News API. <https://github.com/fliarbi/prism-news-api>