INSIGHT REPORT

What an answer costs.

Using AI responsibly. The energy, the people, and who pays.

BespokeWorks

Insight report / September 2026

Why this report.

Most businesses now use AI in some form, and the same doubts surface in almost every first conversation about it. People want to know what it does to the environment, whether client information is safe inside it, what it means for the people they employ, and whether using it for trivial things is a waste. Those questions get asked constantly and examined rarely, partly because the loudest voices on the subject are either selling the technology or opposing it.

What follows examines them against the source documents rather than the coverage of them, from the perspective of someone who installs these systems for a living. The figures all belong to other people; what is offered here is a reading of them, which in several places runs against the consensus. It is written for business owners, because they are the ones deciding, though the same questions apply to anyone who opens ChatGPT at the kitchen table.

What this is not

A complete account of AI and the environment. The well-covered ground is left to others. These are the parts that get less attention: what a data centre is actually for, where the new clean power went, the gap between two audited emissions figures for the same company, and who pays for the delay.

An evidence review of four common beliefs about AI, and what the published figures actually show.

The short version.

The usual worry runs like this. AI needs enormous amounts of electricity, the world cannot build clean power fast enough to keep up, and so every question typed into a chatbot is another small push towards coal and gas. Very little of it holds up. What does hold up sits somewhere most coverage never looks, which is the reason for writing any of this down.

Clean electricity is being built faster than data centres can use it. In 2025 new wind, solar and other clean generation outgrew new data-centre demand by roughly twelve to one, and global power-sector emissions stayed flat. The difficulty is what happened to that clean power. Nearly all of it was absorbed by other new demand, cars and cooling and industry, before data centres were counted at all, and the grids that carry it cannot connect new supply or new load anywhere near quickly enough. Meanwhile the way the largest companies count their own emissions has drifted a long way from what their power stations actually emit. Google publishes two audited figures for the same year and one is more than five times the other.

For an individual, none of this turns on how often they use ChatGPT. A year of heavy use costs about as much electricity as twelve miles in a petrol car. For a business it does turn on something, because a business chooses at volume, and the difference between the cheapest and the dearest thing a model can be asked to produce is around sixty thousand to one.

A teaspoon of water.

On 1 September 2026 Sam Altman said on a podcast that a ChatGPT query uses about a fifteenth of a teaspoon of water. Set beside the three litres that go into growing a single almond, the worry looks small, which was the point of saying it.

PolitiFact checked the comparison three days later and found it did not hold, for a reason worth understanding. The almond figure is a total water footprint. It counts rain that fell on the orchard and would have fallen regardless. The query figure counts something much narrower. It appears to cover only the water evaporated inside a data centre's own cooling system, and not the water consumed at the power station generating the electricity to run it, though that was never spelled out. When the International Energy Agency traced where a data centre's water actually goes, in its 2025 report on energy and AI, about two thirds of it turned out to be at the power station rather than in the cooling towers. The smaller of the two numbers is the one being quoted.

What the comparison hides is the shape of the problem. Water is not drawn a teaspoon at a time across the whole planet. It is drawn in volume in particular places, wherever a cluster of data centres and the power stations behind them happen to sit, and it is drawn there whether the local river is high or low that summer. A total divided by a trillion queries will always come out as a teaspoon. The figure that would tell a town whether to worry is the draw at the site, and no company publishes it.

What makes that figure matter is that cooling of this kind consumes water rather than borrowing it. The water leaves as vapour and does not return to the river it came from, and the power station does the same thing again further upstream. Where a cluster of data centres lands in a dry region, that consumption competes directly with everything else drawing on the same supply, during exactly the hot weeks when the cooling load is highest and the river is lowest. Neither the operator nor the utility is obliged to say how much.

What one answer takes.

A sense of the scale of the machinery behind one answer is worth having, if only because nobody outside the companies knows it exactly. OpenAI's documentation for its current top model gives no parameter count, no compute figure, no energy figure and no water figure; the only numbers it publishes are prices. So every estimate in circulation is a guess, and this one is no different, but the working is shown so the guess can be checked.

The arithmetic is simple enough to follow. A model performs roughly two operations for each of its active parameters, for every word it reads and every word it writes. A frontier model is thought to switch on something like two hundred billion parameters for any given word, which puts one word at about four hundred billion operations. An ordinary exchange, five hundred words in and three hundred words back, is eight hundred words. Multiplied out, that is around three hundred and twenty trillion operations for a single answer.

The assumptions

Two operations per active parameter per word, the rule Epoch AI used in its 2025 estimate for ChatGPT. Two hundred billion active parameters, five hundred words in, three hundred out. Halve or double any of those and the answer moves accordingly; the plausible range runs from forty trillion to several thousand trillion. World population 8.21 billion.

The number is hard to picture, so here are three ways to hold it. Shared between everyone alive, it is about thirty-nine thousand sums each. Set the whole planet doing one sum a second and it takes eleven hours. One person alone, never stopping, would need ten million years. The “thinking” mode that ChatGPT, Claude and Gemini now offer multiplies all of that by roughly seven.

Two cautions belong beside it. These are multiply-and-add operations on very small numbers, run side by side, which is why an answer arrives in seconds. And the total says nothing about whether the answer was any good.

Training or use.

If the machinery behind a single answer is that large, the next question is where the electricity for it actually goes. The common assumption is that it goes on training: the enormous one-off runs that produce a model, with everyday use a rounding error afterwards. The assumption is a comfortable one, because it keeps the whole problem inside four or five companies, and it is the assumption underneath the teaspoon. If use really were a rounding error, a teaspoon would be the right unit for it.

The figures point the other way. The International Energy Agency’s estimate for every frontier training run since 2020, added together, is about 1.7 terawatt-hours, a tenth of one per cent of what data centres use in a year. Almost everything else is use. Google’s 2025 fleet-wide measurement was of answering questions, not of training, because that is where its AI electricity goes, and the Microsoft study published this year in the journal Joule models the same thing: the cost of an answer, multiplied by the number of people asking.

The number most often misquoted

Google’s 0.24 watt-hours per text prompt, published in August 2025, is a careful measurement. It is also a median across Google’s models rather than across prompts, and Google does not publish the mean, because a few very large prompts pull it up. Microsoft’s aggregate figures imply a mean about one and a half times its median. A flagship answer costs several times a typical one, and a thinking answer many times that again.

Training gets the attention because it is large and happens once. The weight sits on use, which never stops, grows with every new user, and is the part a business can actually choose.

What a data centre is for.

Before any of this can be weighed, one distinction is worth drawing that most coverage skips entirely. A data centre is not an AI facility. The buildings that took roughly 415 terawatt-hours worldwide in 2024, about one and a half per cent of global electricity, are mostly doing what they were doing five years ago: holding files, running company systems, serving search results, and above all delivering video and social feeds to phones. Streaming and social platforms were the reason data-centre demand grew through the 2010s, and they remain the larger share of the base.

AI is the fastest-growing slice of that total rather than the whole of it. The distinction matters because almost every headline about data-centre growth attributes the entire figure to AI, and because the two have different remedies. A recommendation engine feeding an infinite scroll draws power from the same building as a model drafting a contract, though the coverage tends to name only the second.

How the split falls is not published in any useful form. Companies report data-centre energy in total, if at all, and not by what the servers are doing. Anyone quoting a precise share of data-centre electricity for AI is estimating, whichever side of the argument they are on.

Where the new clean power went.

With that distinction in place, the claim heard most often can be tested: that AI's demand for power is outrunning clean energy. On the 2025 figures it is false, and the way it is false is the more interesting finding.

Renewable generation grew about 8.5 per cent in 2025. Solar alone added roughly 600 terawatt-hours, the largest single-source increase on record. All clean sources together added 887. Data centres added 70, a ratio of about twelve to one. Renewables overtook coal in the global mix for the first time in a century, and global power-sector emissions were flat in a year when data-centre demand grew seventeen per cent. Taken on its own, that reads as a closed question.

WHERE THE NEW CLEAN GENERATION WENT IN 2025 / TERAWATT-HOURS

Ember's Global Electricity Review 2026 gives 887 terawatt-hours of new clean generation and 849 of new demand for 2025, which leaves 38 that displaced fossil fuel. The data-centre and electric-vehicle figures are the IEA's, from its Global Energy Review 2026; the rest of new demand is what remains. Sizes are to scale.

The bar shows why it is not. Total demand grew 849 terawatt-hours in the same year. Clean supply grew 887. Fossil generation is whatever is left over, so the entire global fall in fossil electricity in 2025 was the difference between those two: 38 terawatt-hours. Data centres alone added 70. The new clean power had been absorbed by other new demand before data centres were counted at all.

Netting one favoured use against the whole clean build is the error, and an easy one to make because the ratio it produces is so flattering. Electric vehicles added 66 terawatt-hours in 2025, nearly the same as data centres, and pass the same test just as easily. So does air conditioning. Together they take the lot. The IEA puts data centres at about a tenth of demand growth to 2030, behind industrial motors, air conditioning and electric vehicles.

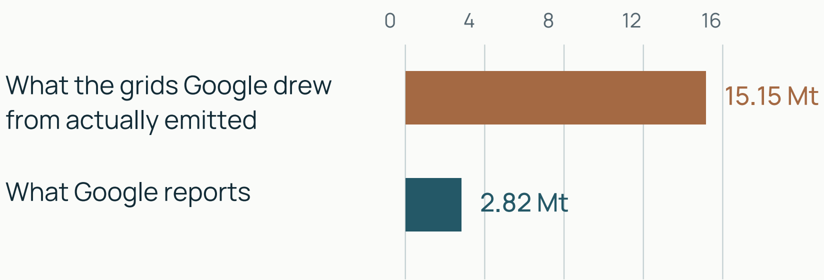
The condition for all of this is worth stating plainly, because it is the one the coverage tends to skip. Data centres cost the climate nothing for exactly as long as clean generation grows faster than all new demand put together, not faster than data centres alone. In 2025 it did, by 38 terawatt-hours. Through July 2026 it did not: demand grew 547 terawatt-hours against 427 of clean growth, and fossil generation rose. That rise is in China and India rather than in the countries building data centres fastest, and the data shows where emissions moved without showing why. Looking ahead, the IEA expects clean supply to add about 1,000 terawatt-hours a year to 2030 against demand growth of 1,100, which is close enough to a dead heat that every extra terawatt-hour, from a data centre or a car, is met by something burning.

There is a geography hidden inside the global figure as well. China built more than half of the world's new solar and two thirds of its new wind in 2025, and Chinese data centres run on a grid that is about seventy per cent coal. The renewables and the data centres are, to a large extent, in different countries, and the global average smooths that difference away.

Two ways of counting emissions.

That geography is also why the emissions a company reports can drift so far from the emissions its electricity causes. Google's electricity use rose 37 per cent in 2025. The emissions of the grids it actually drew from rose 37 per cent as well, to 15.1 million tonnes, and have nearly tripled since 2019. The emissions Google reports fell 2.9 per cent, to 2.8 million tonnes.

GOOGLE, 2025 / MILLION TONNES CO2E, SCOPE 2



Both figures are in Alphabet's own 2026 environmental report and both are audited. The top one is the grid's average emissions for the electricity Google took, which accountants call location-based. The bottom one lets Google count clean power it paid for elsewhere against power it used here, which they call market-based.

How much reported progress is real

Across 115 companies with science-based targets, reported emissions fell 30.7 per cent between 2015 and 2019. With reductions claimed through certificates removed, the fall was 9.9. The authors estimate that 42 per cent of committed reductions will never happen in the real world. Bjørn and colleagues, Nature Climate Change, 2022.

Neither figure is false. One asks what the grids a company plugged into emitted. The other asks what the company paid for, and allows clean power bought elsewhere, at other times of day, to offset the coal it ran on at three in the morning. Google's own 2026 report says the method "doesn't tell the whole story", and that a company can claim to be powered by 100 per cent renewable energy through a global match of energy purchased in distant regions. The gap opens because a data centre never switches off. Princeton's ZERO Lab modelled a buyer matching 100 per cent of its annual consumption with renewables and found it actually running on about 62 per cent clean power on the large eastern American grid.

Microsoft ran the experiment on itself. It stopped buying unbundled certificates in February 2025 and its reported emissions rose tenfold, while the emissions of the grids it uses rose 20 per cent. Almost all of the jump was paper, and Microsoft said so at the time. Underneath the accounting, absolute emissions are rising at all four companies building this infrastructure: Alphabet up 18 per cent, Microsoft 27, Amazon 16, Meta 9, each in its latest reported year.

That last set of figures is the one to hold onto. Whatever the reporting method says, these four put more carbon into the air in their most recent reported year than in the year before, and have done so for several years running. The global picture in the previous section is reassuring in a way this one is not, and both are measurements of the same world.

The connection queue.

If clean electricity is being built fast enough and emissions still rise, the constraint sits elsewhere. Three things bind, and they compound.

The first is time. The median American generation project that connected in 2025 had waited 61 months in the queue; in 2008 the figure was 22. Of everything that applied to connect between 2000 and 2020, thirteen per cent has connected and three quarters withdrew. Gas in the queue rose 86 per cent last year while solar and wind fell, because a gas plant can be built next to its customer and does not have to wait for anyone.

The missing number

A UK household on a medium tariff uses about 2,500 kilowatt-hours of electricity a year. Nobody has published what data-centre connection costs do to that bill. Of everything absent from this report, it is the figure most worth someone's time.

The second is what runs in the meantime. Global Energy Monitor counts 189 gigawatts of American gas plants in development whose main customer is a data centre. Most will never be built, as GEM itself notes, but 52 gigawatts are actually under construction, about 17 of it tied to data centres.

The third is speculation. Great Britain's queue for new demand connections rose from 41 gigawatts to 125 by June 2025, and around 73 of that is data centres, across roughly 315 applications. Peak demand for the entire country is 45 gigawatts. Almost none of the 315 will be built, but each one sits in the same queue as the wind farms and delays them.

Two facts cut the other way and belong here. American power emissions fell 2.7 per cent in the first seven months of 2026, with coal down ten per cent. And the 2025 coal rebound that drew so much coverage was largely a gas-price story, with spot prices up 56 per cent from a record low, rather than a data-centre story.

Who pays.

Delay has a price, and the answer to who pays it comes from regulators and auditors rather than campaigners, which makes it harder to dismiss.

PJM's independent market monitor calculates that adding data-centre load to the demand forecast raised capacity costs to everyone else on the grid by about 23 billion dollars across three years, and states that the high prices in the 2027/28 auction were "primarily the result of existing and forecast data center load". Virginia's own legislative auditor found the state's data-centre tax exemption cost 929 million dollars in a single year.

Ireland is the clearest case anywhere, and it is forty minutes from Britain by air. Data centres took 23 per cent of all metered electricity there in 2025, up from five per cent a decade earlier, while every home in the country combined took 28. In Slough, data centres took 65 per cent of all electricity drawn from the grid in 2024, double the 2020 share.

The carbon has also moved upstream, where it is rarely counted. Capital goods, mostly chips and buildings, are over a third of Microsoft's whole footprint. TSMC, which makes nearly every chip that matters, marked its own 2025 emissions-per-unit target Unachieved. The target was a ten per cent cut. The figure rose.

AI inside the energy system.

Whether the technology can help with the problem it adds to is a fair question. The record splits into one thing that works and one that has not yet been shown to.

Forecasting works, and it is in service. Europe's weather centre has run AI forecasts since February 2025 at about a thousandth of the energy of the physics models. In Britain the grid operator's solar forecast, built with a small non-profit called Open Climate Fix, cut its average error from 650 megawatts to 233, and it runs in the control room in Wokingham every day. Better forecasts mean less gas kept spinning in reserve against a cloudy afternoon.

Discovery has not shown the same. The new-materials models published in Nature in 2025 are real, but an independent review found scant evidence of anything useful emerging from them, and one team conceded that its system had rediscovered a decade of existing chemistry.

A paper that argues the other way

The IEA's 2025 scenario has existing AI tools cutting about five per cent of energy emissions by 2035. A paper from August 2026 that models AI as a general productivity booster finds a net increase of 0.5 to 1.8 billion tonnes a year unless clean energy gains outrun fossil gains four or five to one. Both are models rather than measurements.

Then there is OpenAI's mathematics. On 8 September 2026 OpenAI published a proof for a version of the Navier-Stokes equations, the mathematics of how fluids move and one of the seven Millennium Prize problems. It is serious work, machine-checked across 641,000 lines, and routinely misdescribed. It settles two of the four statements the prize was set on, the two that allow an outside force to push the fluid; the two that ask whether a fluid left alone can break down by itself remain open. It came from an unreleased internal model, it built on years of work by two Spanish mathematicians, and the Clay Institute has so far said only that it "contemplates" the announcement.

For this report the point is narrower. Terence Tao wrote days before the announcement that a result either way would not change how weather or climate is modelled, because computational fluid dynamics is already a mature field, and OpenAI makes no energy claim about it. The line often drawn, that AI can now do real science and should therefore be pointed at renewables, does not run through this result. The more defensible ask is that research effort go into making the models themselves cheaper to run, where the gains have been large and measured.

Generated video.

The last of the four doubts is about people rather than about grids: that using AI for entertainment, generating pictures and clips for amusement, is a waste. It divides into a question about energy and a question about harm, and the two answers point in different directions.

On energy, the doubt is aimed at the wrong output. Text barely registers. A hundred prompts a day, every day for a year, comes to 8.8 to 12.4 kilowatt-hours, about half a per cent of a British household's electricity, or roughly twelve miles in a petrol car. Every consumer message sent to ChatGPT worldwide in a year comes to around 0.05 per cent of global data-centre electricity. It is aimed at the wrong test as well, because the seriousness of the user has no bearing on the cost. An idle question costs a quarter of a watt-hour; a solicitor putting a hundred-thousand-word contract through Claude costs about forty, for a perfectly good reason. What sets the cost is the output asked for and the size of the model answering it.

The only company that counts

No video platform publishes what share of its uploads or views is AI-generated. YouTube reports 200 billion daily Shorts views and no breakdown. TikTok reports three billion AI labels and no denominator. Deezer, a mid-sized French music service, does count: AI tracks passed half of daily uploads there in 2026 while drawing one to three per cent of listening.

Video is where the figures become large. A short clip on an efficient model measures about three watt-hours; a longer one on a heavy model measures over three thousand. Nearly three orders of magnitude of difference, turning on model, length and settings that nobody discloses. Anyone concerned about the energy in casual AI use should be looking at video and not at chat.

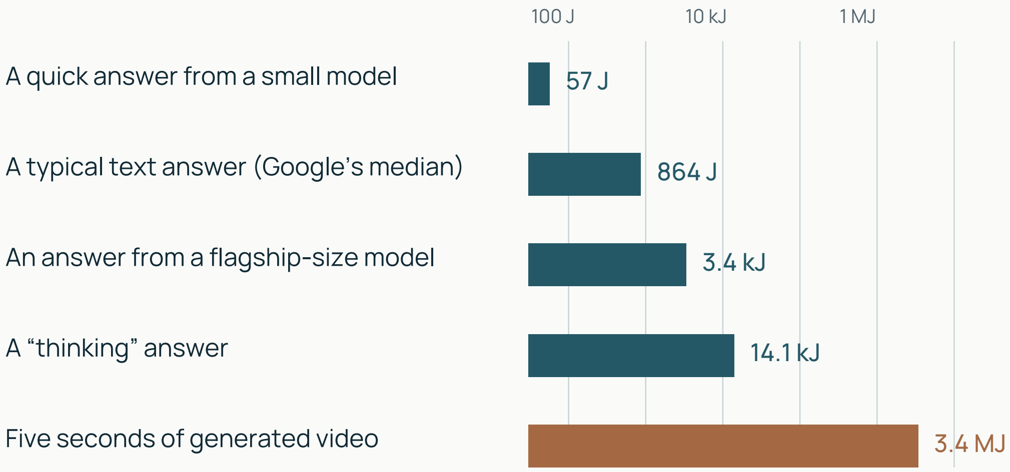
On harm, the electricity is the smaller of the two problems with generated video. A near-perfect video generator available to anyone puts convincing footage of events that never happened within reach of anyone with a grudge or a scam, and the quality is improving faster than the means of detecting it. The people most exposed are the ones least able to check: an elderly relative taking a call, a small firm paying an invoice, anyone shown a clip of a public figure saying something they did not say. That is a question about the product rather than about its power draw, and it receives a fraction of the attention the energy question does.

The harm asserted most often, though, is to attention: that “slop” is degrading how people think. There the honest answer is that nobody knows, because the study has not been run. There is no research on AI-generated video and cognition at all, in either direction. The wider short-form video literature is thinner than its headlines: the largest meta-analysis, published in 2025 across 71 studies and 98,000 people, does associate heavy use with worse attention, but nearly all of it is snapshot surveys, half measure self-described “addiction” rather than actual viewing, and the authors state that they cannot tell which way the causation runs. The one experiment that changed how long people watched found nothing. It is worth knowing that “brain rot” was Oxford’s word of the year in 2024, that Oxford’s own definition says “supposed deterioration”, and that the phrase belongs to Thoreau, who used it in 1854 to complain about people reading too many newspapers.

What this means for a business.

The energy case for an individual changing habits is weak. The case for a business changing how it uses these models is strong, because the spread in cost is enormous and a business chooses at volume.

MEASURED ENERGY PER OUTPUT / JOULES, LOG SCALE



Each gridline is ten times the one before it. The small and flagship-size model bars are measurements of open models by the ML.Energy team at Michigan, reported in MIT Technology Review in May 2025; nobody is allowed to measure ChatGPT or Claude directly, so the open models stand in for them. The typical answer is Google’s own median for a Gemini text prompt. The thinking answer is from a Microsoft and University of Oviedo paper in Joule this year. The video figure was measured for MIT Technology Review. One caution I owe you: nearly every public measurement of video energy comes from one research group, and nobody has repeated it.

Route work to the smallest model that does the job. ChatGPT, Claude and Gemini each offer a fast model and a flagship, and the flagship costs tens of times more per answer for work it does no better: sorting, extracting, summarising, first drafts, which is most of what an office asks. A system that sends routine work to a small model and reserves the large one for hard cases is cheaper and faster with no loss of quality, and it is how the better production systems are already built. The “thinking” mode costs about thirteen times a plain answer and belongs on hard problems rather than on by default. Video deserves the longest pause. The argument here is about cost rather than about the environment, though it happens to point the same way.

A business that wants to act on the environmental side specifically has one real lever, and it is procurement. Ask a cloud or AI supplier for the location-based emissions of the region serving the account, and for the hours the work actually ran in. Most will not answer that today. Being asked often enough is what changes it, and a buyer is in a better position to ask than a reader is.

The second doubt, about data, has a clearer answer than most people expect. Using AI is lawful. Putting a client’s personal or confidential information into a tool without a lawful basis is the breach, and it would be a breach with a filing cabinet. What matters is the account, because consumer and business accounts run on different terms. Anthropic lets consumer users opt in to training and retains their data for five years if they do, with business and API traffic excluded. Google’s consumer Gemini pages say plainly not to enter confidential information. Client work on a personal login is the most common mistake and the cheapest to fix. One rule has also moved in a business’s favour: the old GDPR restriction on automated decisions was replaced in February 2026, and decisions with real consequences are now permitted with safeguards unless they touch health, ethnicity or the other protected categories.

Ownership is the risk most often held the wrong way round. The major copyright cases have largely gone the AI companies’ way, so the danger in using these tools is smaller than the headlines suggest. What cannot be assumed is ownership of what comes out. The vendors assign rights “if any”, and in the United States there is no copyright without a human author. Anything that will carry commercial weight is worth drawing by hand.

The third doubt, about people, has the least comfortable answer. AI helps most where the user knows least: new staff gain the most and experts gain little. Dell’Acqua and colleagues found in 2026 that consultants using AI inside its competence did more work, faster, and outside its competence were nineteen per cent less likely to be right, with the tool equally confident in both cases. Stanford’s payroll data, revised in August 2026, shows 22 to 25 year olds in AI-exposed jobs running nineteen per cent below trend, concentrated where the tool replaces people instead of helping them. A business deciding whether to hire a junior this year should weigh that.

One last correction, to a line often repeated: that AI lacks curiosity. Curiosity-driven exploration is a standard technique in the field. What is true is narrower. These systems have no goals or stakes of their own, so they will not say that the wrong question is being asked, because nothing about the answer matters to them.

Open questions.

Several things in this report could not be settled. Nobody has published a study on whether the long-term clean power contracts these companies sign are genuinely additional, and the accounting section leans on that answer. Nobody publishes how data-centre electricity divides between AI and everything else the buildings do, what data centres add to a household bill, or what the water draw is at any given site. No vendor will say what a single answer from a frontier model costs to produce. No platform publishes what share of a video feed is synthetic. And the study on AI-generated video and attention has not been run, so anyone claiming to know the answer is guessing, in whichever direction.

Sources.

Everything below is the original document, grouped by the section it feeds. Where a figure is a company's own report, it says so, because a company measuring itself is still the only measurement there is. Nothing about AI itself is older than 2025, because the technology has changed too much for earlier figures to hold. Estimates and forecasts are labelled beside the figure in the text.

A teaspoon of water

Sam Altman, interviewed on Sources with Alex Heath, podcast, 1 September 2026.

PolitiFact, fact-check of the almond comparison, 4 September 2026.

International Energy Agency, Energy and AI, April 2025: where data-centre water is used, on site and at the power station.

What one answer takes

OpenAI, GPT-6 Astra system card, dated 3 September 2026 and revised 9 September.

Epoch AI, How much energy does ChatGPT use?, February 2025: the operations-per-word rule and the assumptions behind it.

US Census Bureau, International Database world population clock, read 14 September 2026.

Training or use

International Energy Agency, Energy and AI, April 2025: total energy of frontier training runs since 2020.

Elsworth and colleagues (Google), Measuring the environmental impact of delivering AI at Google scale, arXiv 2508.15734, August 2025. A company measuring itself.

Oviedo and colleagues (Microsoft and University of Oviedo), Energy Use of AI Inference, Efficiency Pathways, and Test-Time Scaling, Joule 10(8), 2026.

What a data centre is for

International Energy Agency, Energy and AI, April 2025: total data-centre electricity in 2024, its share of world electricity, and the composition of data-centre workloads.

No source exists for a clean split between AI and non-AI data-centre electricity. Companies report the total, where they report anything.

Where the new clean power went

International Energy Agency, Global Energy Review 2026, 20 April 2026.

Ember, Global Electricity Review 2026, 21 April 2026.

Ember, Monthly Electricity Data, release covering January to July 2026. Provisional.

International Energy Agency, Electricity 2026, 6 February 2026: renewable additions against demand growth to 2030.

International Energy Agency, Energy and AI, April 2025: data centres' share of demand growth to 2030, and the supply mix for Chinese data centres.

Two ways of counting emissions

Alphabet, 2026 Environmental Report, covering calendar 2025. Location-based and market-based Scope 2. A company measuring itself.

Microsoft, 2025 Environmental Sustainability Report, covering fiscal 2025. A company measuring itself.

Amazon, 2025 Sustainability Report. Meta, 2025 Sustainability Report, covering calendar 2024. Companies measuring themselves.

Xu, Manocha, Patankar and Jenkins, Princeton ZERO Lab, working paper on hourly carbon-free electricity matching. A model of 2030, not a measurement.

Bjørn, Lloyd, Brander and Matthews, Nature Climate Change 12(6), 2022: 115 companies with science-based targets. About corporate accounting rules, not AI.

The connection queue, and who pays

Lawrence Berkeley National Laboratory, Queued Up, 2026 edition: the interconnection queue.

Global Energy Monitor, Global Oil and Gas Plant Tracker, and its August 2026 briefing on gas plants serving data centres.

Ofgem, Proposed data centre connection reforms, 29 July 2026.

Department for Science, Innovation and Technology, Estimate of Data Centre Capacity: Great Britain 2024, 1 May 2025.

US Energy Information Administration, Electric Power Monthly, August 2026: coal and solar generation to June.

Monitoring Analytics, the PJM Independent Market Monitor, Analysis of the 2027/2028 RPM Base Residual Auction.

Virginia Joint Legislative Audit and Review Commission, report on data centres in Virginia, 2024.

Central Statistics Office Ireland, Data Centres Metered Electricity Consumption 2025, 7 July 2026.

Department for Energy Security and Net Zero, subnational electricity consumption statistics, 30 June 2026: Slough.

TSMC, 2025 Sustainability Report: the emissions-per-unit target marked Unachieved. A company measuring itself.

Ofgem, Review of typical domestic consumption values: decision, 27 May 2026.

AI inside the energy system

European Centre for Medium-Range Weather Forecasts, AI forecasts in operation from 25 February 2025.

National Energy System Operator, solar forecasting with Open Climate Fix. The saving figure is Open Climate Fix's own estimate.

Microsoft Research, MatterGen, Nature, 2025, and the independent assessment of machine-discovered materials that followed it.

OpenAI, papers on finite-time blow-up for the Navier-Stokes and Euler equations, 8 September 2026, with a Lean 4 proof artefact. Self-published, not refereed.

Clay Mathematics Institute, statement, 11 September 2026.

Terence Tao, blog post on the Navier-Stokes regularity problem, 3 September 2026.

Alpine and colleagues, npj Climate Action, 4 August 2026. A model, not a measurement.

Generated video

Chatterji and colleagues, How People Use ChatGPT, NBER Working Paper 34255, 2025: messages per week.

Jegham, Gamazaychikov and Luccioni, Lights, Camera, Carbon, arXiv 2607.04553, July 2026: energy per generated video.

Nguyen and colleagues, Psychological Bulletin 151(9), 2025: the meta-analysis of short-form video and cognition.

Lin and colleagues, CHI 2024 Late-Breaking Work: the one experiment that changed viewing time. About short-form video generally, not AI.

Oxford University Press, Word of the Year 2024.

YouTube blog, 10 August 2026. TikTok Newsroom, 10 July 2026. Deezer newsroom, 21 July 2026.

What this means for a business

ML.Energy, University of Michigan, measurements of open models, reported in MIT Technology Review, 20 May 2025.

Anthropic, consumer terms and privacy policy update, October 2025. Google, Gemini Apps privacy notice, August 2026.

Data (Use and Access) Act 2025, section 80, in force 5 February 2026.

Getty Images v Stability AI [2025] EWHC 2863 (Ch). Bartz v Anthropic, settlement approved 21 July 2026. Thaler v Perlmutter, certiorari denied 2 March 2026.

Dell'Acqua and colleagues, Organization Science, 2026.

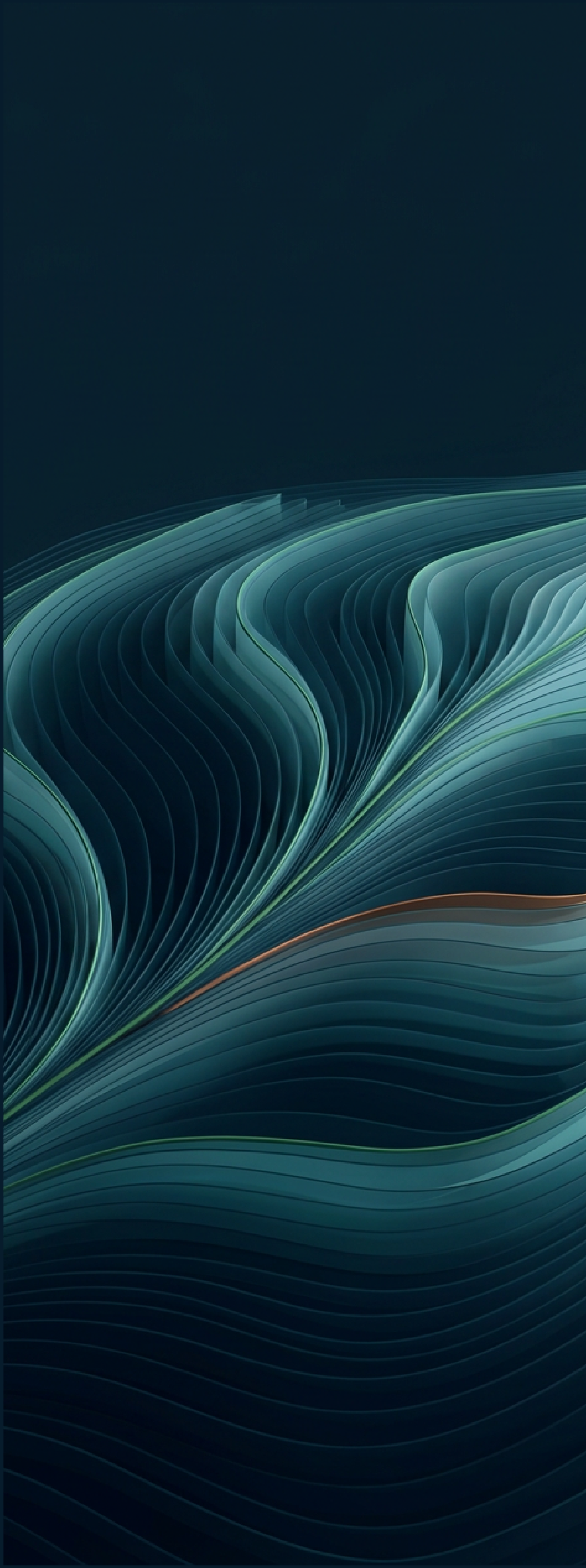
Brynjolfsson, Li and Raymond, Generative AI at Work, Quarterly Journal of Economics, 2025.

Stanford Digital Economy Lab, Canaries in the Coal Mine, revised 12 August 2026.

What should change

Commission Delegated Regulation (EU) 2024/1364, on energy and water reporting by data centres.

What would settle this.



The gaps in this report are not gaps in what can be known. Almost every figure a reader would want already exists inside a company that has chosen not to publish it, and twice over the law requires the measurement without requiring anyone to see it.

Google and Microsoft already print the grid figure alongside the contracted one, which is the whole of what is being asked for here. Amazon, the largest emitter of the four, prints neither that nor a total electricity figure, which makes the biggest of them the hardest to check. Every data centre in the European Union above 500 kilowatts already reports its energy and water to a regulator, and the regulation then keeps each return confidential, so the site-level water draw that would tell a town whether to worry is collected and withheld. One company in the world publishes a fleet-wide figure for the energy in a single answer, and that figure is a median with no mean beside it.

The useful version of all this is smaller than a campaign. Print the location-based figure next to the market-based one. Publish the mean as well as the median, and say what the thinking mode multiplies it by. Open the returns that are already being filed. Give connection priority to sites that bring their own clean power or accept being switched down when the grid is tight. A report of this kind written in three years should not have to estimate any of it.

Every measurement above is already being taken.