

AI and Competition Law: The Data Moat Is a Mirage

Why data is not the barrier to entry in AI markets

Every argument about AI and market power runs, sooner or later, on a single mechanism. A firm with a better model attracts more users. More users generate more data. More data buys a better model still. And the lead, once won, widens by itself.

Call it the flywheel. It is the engine beneath every prediction that AI markets will tip towards one or two winners and stay that way, and it is invoked in nearly every call for pre-emptive intervention, usually without being examined. This article examines it, and the examination does not go the way the worry wants.

(A note for new readers. This series argues that AI dissolves competition law's silent assumptions rather than breaking its rules. The [opening article](#) set out the map, the [previous article](#) took the AI stack apart, and this one tests the engine.)

Why the Flywheel Is Believable

The flywheel's appeal is that it fits the recent past. The platform giants of the previous technological era compounded their advantages through data and network effects, whereby a service becomes more valuable as its user base grows. It is therefore natural to assume that AI would follow the same pattern.

Under this view, the decisive input is data. Whoever holds the most of it trains the best models, and the loop does the rest.

Why Data Is Not a Moat

The trouble is that data, examined as an economic asset, has almost none of the properties a moat requires. Four, in particular, are missing.

Data is non-rival, which means many firms can use the same data at the same time without any of them using it up. A railway line can be occupied by one train at a time. A dataset cannot be occupied at all.

Non-rival is not the same as open, however, and the difference matters. A firm can still keep its data to itself, by contract, by secrecy or by technical control. For a moat, it is the property that counts. So the real question is not whether data can be shared, but whether the data needed to build a frontier model is data anyone can keep to themselves. For the general training material of foundation models, drawn largely from public sources and increasingly supplemented by artificially generated data, the answer is mostly no.

Data shows sharply diminishing returns. Past a certain threshold, adding more of it barely improves a model.

Data depreciates. Stale data loses its value, so a would-be data monopolist has to keep re-acquiring rather than sitting on a hoard.

And data is increasingly substitutable, because synthetic data, meaning data generated artificially rather than collected, and shared public collections now open many routes to a comparable training set.

Whatever holding a great deal of data confers on a firm, it is not the durable exclusivity that the word "moat" implies.

That conclusion is narrower than it sounds, and it should be stated with its limits. Where data comes from a continuing flow of users rather than a public stock, it can be both exclusive and decisive, and courts and regulators on both sides of the Atlantic have acted on exactly that basis.

In the United States, the District Court for the District of Columbia found in August 2024 that Google had unlawfully maintained its monopoly in general search. In its [remedies opinion of 2 September 2025](#) in *United States v Google LLC*, the court ordered Google to share its search index and user interaction data with qualified competitors, on the footing that sharing would deny Google the fruits of its exclusionary conduct.

In the European Union, Article 6(11) of the Digital Markets Act obliges Google, as the only gatekeeper designated for online search, to share ranking, query, click and view data with rival search engines. In July 2026, in [Case](#)

[DMA.100209](#), the European Commission adopted a specification decision setting out how Google must comply, and extended access to providers of AI chatbots from January 2027.

The United Kingdom has taken a narrower approach. Having designated Google with strategic market status in general search and search advertising in October 2025, the Competition and Markets Authority imposed [conduct requirements](#) in June 2026 that include a right for users to move their own search data to authorised third parties. That is portability at the user's direction, not access for rivals to Google's aggregate data.

The same kind of data, an agent's accumulated knowledge of the person it serves, is the source of the agent's power discussed later in this series. The argument here is not that data never confers power. It is that the data needed to build a frontier model is not where the durable barrier lies. The Commission's decision to open search data to AI chatbots fits the same pattern: the contest it protects is the one for answering people's questions, downstream of the model, rather than the building of a frontier model from scratch.

What the Evidence Shows

The empirical record points the same way, and the opposing case deserves to be stated at full strength rather than caricatured.

Through 2024 and 2025 the frontier was not tipping. It was fiercely contested, with several laboratories at or near the top, releases following one another closely, and the lead changing hands.

The sharpest data points arrived from an unexpected direction. DeepSeek showed it could train models rivalling those of the leading laboratories while reportedly spending around five million dollars on computing power. In July 2026, Moonshot AI's Kimi K3 went further, ranking third on a widely followed independent index, behind only two of the leading American laboratories' models, with its weights released publicly.

Both results carry an asterisk that any honest account has to include. In February 2026, Anthropic reported that DeepSeek and Moonshot, among others, had run industrial-scale campaigns to extract its models' capabilities through distillation, which means training a model on the outputs of a stronger one. In September 2026, US agencies made similar allegations against six Chinese firms, which China rejects. If those allegations are right, part of the capability that looked cheap was built on training runs that were anything but, and the five-million-dollar figure measures the cost of following rather than the cost of leading.

That qualification does not rescue the flywheel. It relocates the question. Distillation is itself a route to the frontier that does not run through data, and a leader cannot easily sell access to its model's outputs without also selling some of what those outputs teach. A moat that leaks through the product it exists to protect is not much of a moat.

The Case That AI Is Not a Natural Monopoly

Simon Goldstein and Peter Salib have pressed this evidence into a flat denial that AI is a natural monopoly at all, in an essay published online by the Minnesota Law Review under exactly that title, [AI Is Not a Natural Monopoly](#). Their argument is that capability improves along a predictable curve, so that a model just as good as the leader's, arriving six months later, is cheap to build, and that advances in training technique have made user data, and with it the data-driven network effect, no longer central to improving these systems. On that account, network effects in generative AI are far weaker than they were in the platform era.

I am not going to pretend that case away. As a foundation for durable market power, the data moat is at best unproven and at worst a mirage, and a structural argument resting on it is built on sand. But their case is aimed at data and at training costs. It says much less about the input that remains genuinely scarce, which is where the argument turns next.

What Survives the Concession

Does anything survive? Something does, but it is narrower, and it rests on a different input.

Strip data of its mythology and ask what is genuinely scarce, rival and hard to replicate at the frontier. The answer is the means of production themselves: specialised computing power, and the capital required to command it. A pound of computing spent training one firm's model cannot simultaneously train another's. Frontier training runs, and the data centres beneath them, now cost sums that only a handful of firms and their backers can assemble.

This is a feedback loop too. Capability attracts revenue, and revenue funds the next and larger training run. But its closest analogue is not the social network. It is semiconductor manufacturing, where the moat is made of the sheer size of the entry ticket.

The Strongest Objection to That

A sceptic will immediately offer the obvious reply, and it is strong enough to be met rather than deflected. If capital is the barrier, why is capital flooding in? Combined spending by the large cloud companies is forecast to exceed \$600 billion in 2026. That is not a picture of scarcity. It is a picture of money chasing anyone with a credible frontier story.

Three things have to be separated here, and running them together is what makes the capital argument look stronger than it is.

The first is capital scarcity, and on this the sceptic is simply right. Capital is the most mobile input in the economy. If frontier capability is worth having, funding will find it, and the capex forecasts show exactly that happening.

The second is capital concentration, which is a different claim and a better one. The question is not whether the money exists but whose balance sheet can commit tens of billions a year, sustained across years, against uncertain returns, without having to raise it from anyone. That list is short, and it does not lengthen quickly.

The third cuts against my own argument and should be conceded plainly. Computing power, unlike a semiconductor plant, can be rented by the hour. A chip factory takes a decade to build and cannot be borrowed. Cloud computing converts what would otherwise be a capital barrier into a monthly operating expense, and that is the opposite of a moat.

The answer to the third point is only partial, and I will not dress it up. Renting frontier-scale computing at a sustained scale still runs through the same small group of cloud providers. So the rentability that dissolves the capital barrier delivers you instead to the bottleneck the previous article identified. The moat moves. It does not disappear.

Even then, the surviving version is conditional. If the computing power needed to reach the frontier keeps falling, which is what would happen if efficient training of the kind DeepSeek claimed generalises, or if distillation keeps carrying capability downstream, the capital moat erodes from the inside, undermined not by anyone outspending the incumbents but by the falling price of doing what they do. The capital engine is more defensible than the data engine. It is not unconditionally defensible, and the only honest way to hold it is as an open question.

Why the Concession Costs Nothing

Now the twist, which is the reason all of that conceding costs the wider argument nothing at all. Far less rests on the flywheel than the rhetoric of this field implies.

Two claims sit at the centre of this series. The first is that pricing systems can learn to coordinate without any agreement between the firms that run them. The second is that AI agents can be turned against the very customers they represent. Neither claim needs a tipping market in foundation models.

The first needs only ordinary concentration in the markets where pricing algorithms already operate, such as airlines, hotels, fuel and online retail. That concentration predates AI and owes nothing to it.

The second needs only the familiar platform economics of defaults, switching costs and pre-installation, operating one layer above the models entirely.

So you can conclude that the flywheel is a mirage, that AI markets will stay contestable, and that the concentration alarm is the last war misremembered. Both of those problems are still sitting on your desk, unanswered.

That is not a hedge. It is architecture. The strongest claims in this series are deliberately not staked to its most contestable one.

What does inherit the flywheel's fragility should be named just as plainly. The market-structure worries, which are concentration, tipping and the frontier scenarios, weaken exactly as much as the engine does. When this series reaches them, it will say so.

Food for thought. *For in-house counsel:* if a regulator demanded access to your training data tomorrow, what competitive advantage would they actually be holding? *For practitioners:* how much of your structural theory survives if the data moat goes, and have you tested it? *For policymakers:* are your proposed remedies aimed at the input that is actually binding, or at the one that is easiest to legislate about?

Next in this series: the four assumptions competition law never states (an agreement, an author, a consumer, and time), and what happens when they fail.

Join the conversation. If you hold the data-moat line (or think I have conceded too much to the contestability camp), the comments are yours, and this is where I most want the pushback.