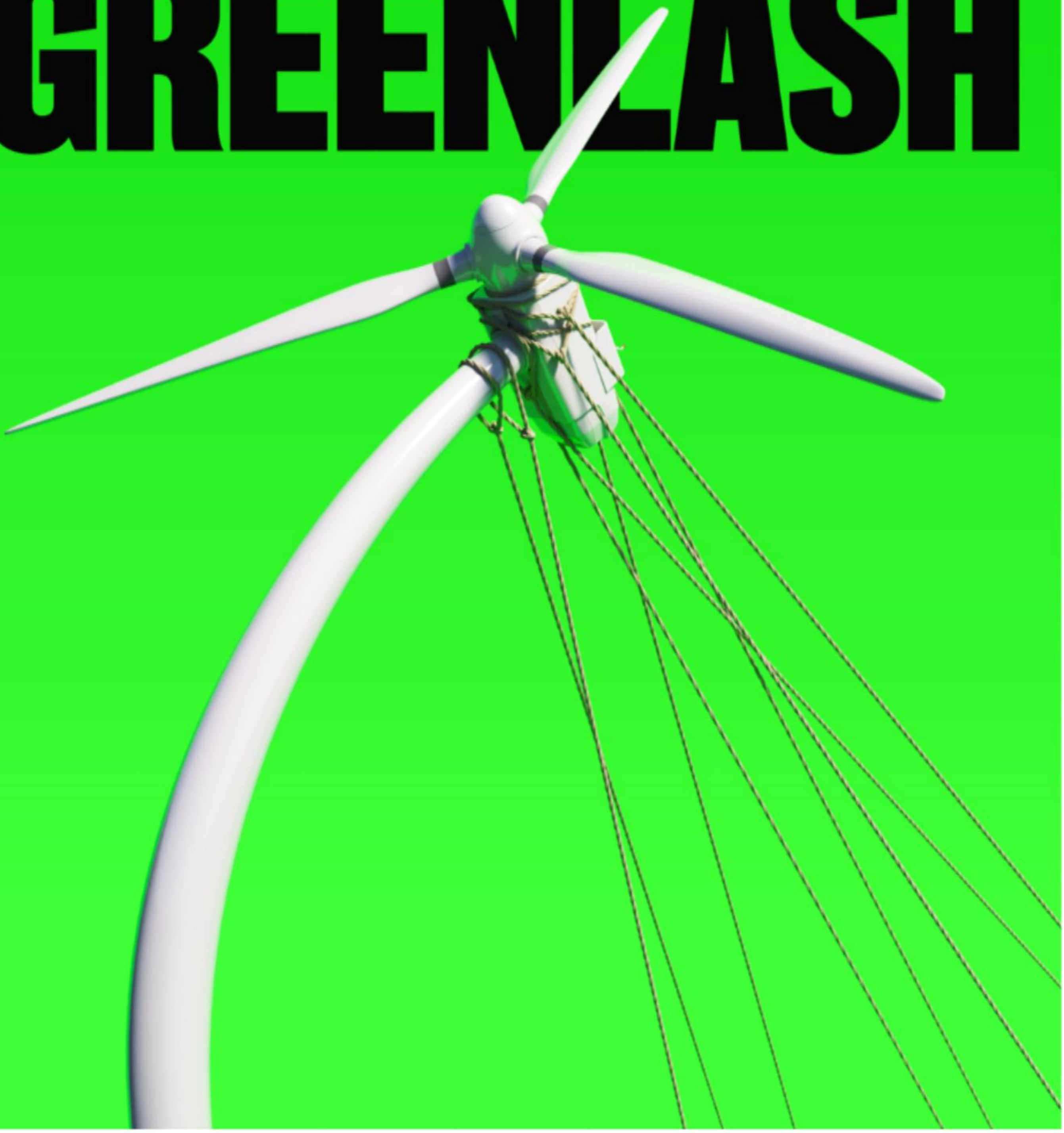
The Economist

AUGUST 2ND–8TH 2025

GREENLASH



Business

Photograph: Getty Images

The European Union reached a preliminary trade deal with America, which gave Donald Trump most of what he wanted. EU exports to America will be subject to a 15% tariff. Mr Trump had threatened 30%. The 15% duty applies to cars made in the EU, a lower rate than Mr Trump's sectoral levy. Meanwhile, the bloc will eliminate tariffs on American-made vehicles and a range of other goods, including aircraft and chipmaking equipment. The EU also committed to buying more American energy and investing more in the United States. Ursula von der Leyen, the president of the European Commission, insisted the deal was "the best we could get".

Other countries scrambled to reach trade deals with America before an August 1st deadline. An agreement with South Korea leaves it with a 15% tariff on its goods. It had faced a 25% duty.

Caught in the trade winds

More companies reported the impacts from tariffs. Mercedes-Benz blamed the duties for a decline in sales in North America and Asia in the first half of 2025, year on year. Taking the EU trade deal into account, it now expects sales for the full year to be "significantly below" last year's. Ford ran up $800m in tariff costs in the second quarter and swung to a net loss. And Procter & Gamble said trade levies would cost it $1bn, and that it would have to raise prices across a range of consumer goods.

The Federal Reserve kept its benchmark interest rate on hold, at a range of between 4.25% and 4.5%. The central bank said that inflation remained "somewhat elevated", but recognised that growth had "moderated", possibly signalling a future rate cut. For the first time since 1993, two governors dissented from their colleagues and voted for a cut.

The Bank of Japan left its key interest rate unchanged, at 0.5%, but raised its forecast for inflation.

Microsoft's net profit soared by 24% in the second quarter, year on year, to $27.2bn, as revenue from its data centres surged. The company announced even more investments in AI. And Meta's stock shot up after it reported solid earnings. "Superintelligence" is going to improve every aspect of the business, said Mark Zuckerberg.

Tesla signed a $16.5bn deal with Samsung for the South Korean company to supply it with chips from its factory in Texas. Samsung will make Tesla's next-generation A16 chips for use in autonomous cars and robots, though production is still years away.

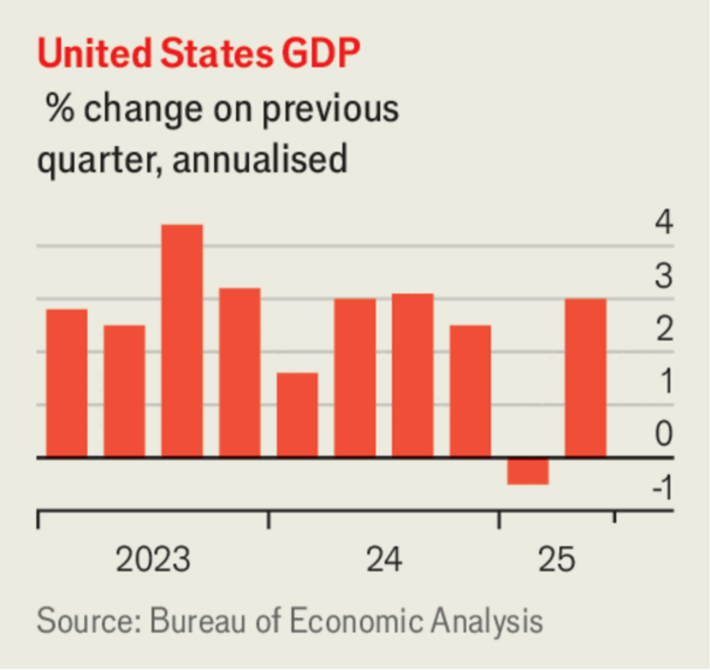


Chart: The Economist

America's economy grew by 3% in annualised terms during the second quarter, a better showing than most economists had expected. A drop in investment was offset by an increase in consumer spending, among other things, said the Bureau of Economic Analysis. Separate data showed the trade deficit in goods narrowing to its lowest level in two years in June, as imports fell sharply.

The IMF raised its forecast of GDP growth for the world economy this year to 3%, from the 2.8% it projected in April. That is partly because America's tariff increases have not been as big as threatened on Mr Trump's "Liberation Day". A weaker dollar has also cushioned the blow for many countries. America's economy is now forecast to grow by 1.9%, up slightly from April's estimate, and China's by 4.8%, up from 4% in April.

Riding the train

America could get its first-ever rail operator to transport goods seamlessly from coast to coast. Union Pacific, which carries freight in the west of the US, agreed to buy Norfolk Southern, which operates in the east, creating a transcontinental company worth $250bn. The deal will attract intense scrutiny from regulators, and could be blown off track by objections from unions and other interested parties.

Novo Nordisk, which makes the Ozempic and Wegovy weight-loss treatments, issued a profit warning for the year amid increasing competition in the American market from compounded versions of the drugs and from Eli Lilly's rival remedies. Its share price sank by almost 30%. The company appointed a new chief executive, Maziar Mike Doustdar, though that did little to allay investors' concerns about its growth prospects. Parvus Asset Management, an activist hedge fund, is reported to be building a stake in the Danish firm.

Renault also got a new chief executive. François Provost was the surprise choice to succeed Luca de Meo, who has left to run Kering, a luxury-goods company.

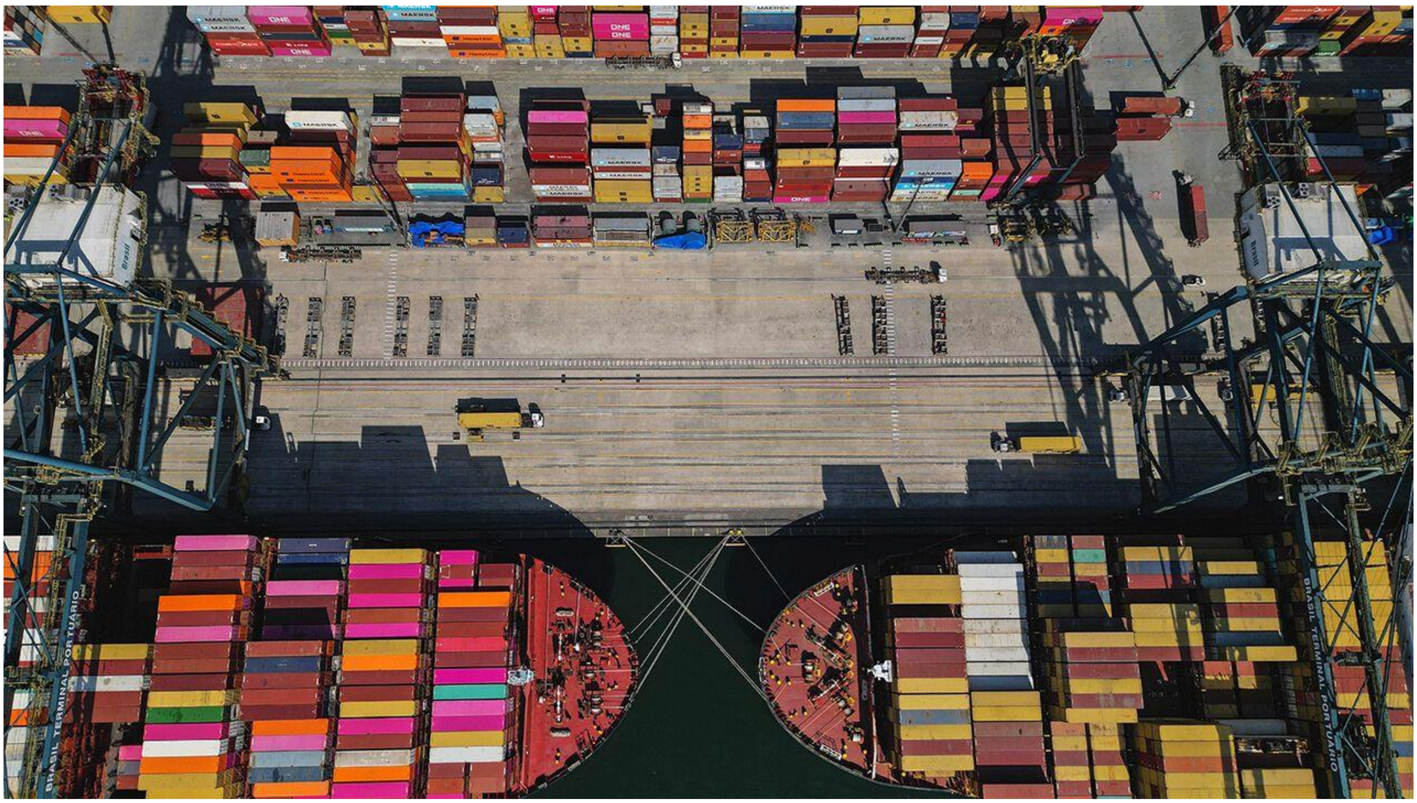
There was some evidence that the efforts by Boeing's chief executive, Kelly Ortberg, to turn round the business were gaining traction. The company's cash burn, the rate at which it spends cash reserves or capital to fund operations, slowed to almost nothing in the second quarter and it recorded a smaller net loss than Wall Street had expected. Boeing's delivery of 280 aircraft in the first six months of 2025 was the most in a first half since 2018.

America's Federal Communications Commission approved Skydance Media's acquisition of Paramount Global, ending the takeover's lengthy legal saga. The transaction is expected to close within weeks.

Finance & economics

Who’s feeling the pain of Trump’s tariffs?

Foreign companies are sharing the load. For now



Photograph: Getty Images

In the bygone age that was 2024, America charged levies averaging just 2% on its imports of goods. In the new era of trade wars, it now has an “effective” tariff of over 16%, the highest since the 1930s (see chart 1). Rates look set to go even higher. On July 31st President Donald Trump signed an executive order that significantly raises tariffs on most of America’s trading partners, with the increases due to go into effect on August 7th. Duties on most products from the European Union and South Korea, which recently struck deals with America, will rise to 15%. India faces a tariff of 25%; South Africa, 30%; Canada, 35%. As we published this, Mr Trump seemed inclined to extend America’s tariff truce with China. But that still leaves the world’s second-largest economy facing levies of around 40% on sales to the world’s largest.

Who pays for these tariffs, in all their infinite variety? Most economists reckon that ordinary Americans will lose out, as prices in shops rise. Mr Trump and his coterie, by contrast, blithely insist that the rest of the world will shoulder the load by cutting their selling prices. So far, the evidence is giving the know-nothings a glimmer of hope.

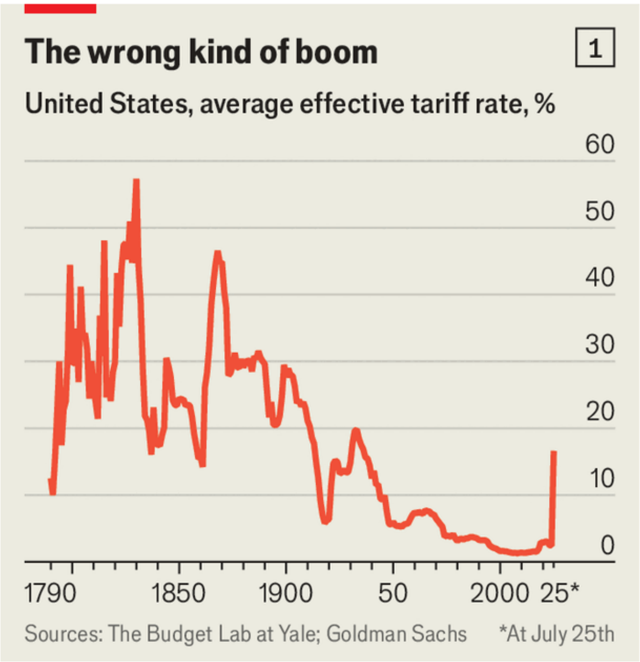


Chart: The Economist

Mr Trump's critics in the economics profession have history and research on their side. Studies show that when a country imposes duties on its imports, its foreign suppliers often keep their prices roughly the same. The tariff is layered on top. So it was during the first Trump administration, which slapped tariffs on China and others. A study from 2019 found "complete pass-through of the tariffs into domestic prices of imported goods".

Some foreign firms are taking a similar stance in response to Mr Trump's new levies. In April Ferrari added up to 10% to the price of its cars. Britain's Ineos said it would charge more for its Grenadier off-roader. Canon, a camera-maker, has warned dealers to brace for price increases.

But the broader pattern is more benign. There is, for example, surprisingly little evidence so far of tariff "pass-through" into inflation. In June America's "core" consumer prices (ie, excluding food and energy) rose by 0.2% on the previous month, below the consensus estimate of 0.3%. Economists have found some evidence of tariff-induced price rises—in car parts, for instance—but they have had to look harder for it than they had expected.

What explains these surprising results? American firms, not consumers, may be paying for the trade war by accepting lower profits, suggests research by Deutsche Bank. Some firms also boosted inventories before the tariffs were implemented, allowing them to avoid raising their prices for now.

America's foreign suppliers may also be sharing more of the load than they did in Mr Trump's first term. Nintendo, a Japanese electronics firm, is keeping the American price of the Switch 2 games console at $449.99. Many Chinese manufacturers seem prepared to follow Nintendo and absorb duties: Fuling, a supplier of cutlery, says its clients expect it to shoulder "part of the increased tariff costs". TIRTIR, a South Korean beauty brand popular with American Gen Zers, has signalled that it can absorb most of the tariffs. Games Workshop, a British manufacturer of war games, also seems resigned to taking the hit, warning investors that tariffs could reduce annual profits by £12m ($16m).

"We found tentative evidence that Korean auto exporters are shouldering the cost of higher US tariffs, at least for now," wrote Kim Jin-Wook of Citigroup, a bank, in a recent note. The Bank of Japan tracks the prices of the country's car exports to America. In yen terms, they have fallen by 26% in the past year. Some of that decline may reflect exchange-rate movements. An unchanged dollar price brings in fewer yen when the American currency is weak. But that only raises another question: why are Japan's carmakers not raising their dollar prices more vigorously in response?

Chart: The Economist

More comprehensive data point in a similar direction. The Economist assembled a series on export prices from a number of America's largest trading partners, including Canada, Germany and South Korea. In the past exporters in these countries have been perfectly willing to raise prices: during the inflationary surge of 2021-22, they increased them by more than 15% over a 12-month span (see chart 2). Yet in the past year the average local-currency price of their exports has fallen by 3.6%. Nothing of the sort happened during Mr Trump's first trade war.

Some economists have noted a disconnect between what foreigners report and what American importers say they are paying. For instance, it is hard to find much evidence of plunging prices for Japanese car imports. Economists at Citi speculate that the time it takes to ship a foreign product to an American port may explain the puzzle. It "implies a lag between falling export prices and when US import-price data would capture the decline", they say.

Why might foreign suppliers be so forgiving? Some bosses worry more than before about the American consumer. With high inflation a recent memory, people already think that everything is too expensive. They have little tolerance for paying even higher prices. The opposite may be true of the foreign companies themselves. They are in a good financial position to withstand the tariffs. Aggregate margins of listed companies in emerging markets have become fatter over the past decade, increasing by over two percentage points. European firms have enjoyed similar gains. These companies can afford to take a small hit to profits, at least for now.

Before long America's economy is likely to feel the pain of the trade war more acutely. Although some Chinese firms may have lowered their prices, these cuts are not nearly deep enough to offset the huge rise in tariffs they now face, points out Deutsche Bank's research. In addition, foreign companies that have borne the costs until now may not be able to bear them for ever—especially if tariff rates keep ratcheting up. The president loves defying his adversaries, in the economics profession and beyond. But he is always his own worst enemy.

A fresh retail-trading frenzy is reshaping financial markets

Blame apps and DORKs, not stimmies

Photograph: Alamy

Investors love an acronym. In recent months they have embraced the TACO (Trump Always Chickens Out) trade. They once swooned over the FAANG stocks (Facebook, Apple, Amazon, Netflix and Google). During Europe's sovereign-debt crisis of the 2010s traders fretted over the PIIGS (Portugal, Italy, Ireland, Greece and Spain). A good memory for acronyms takes a financial historian a long way.

The DORK stocks might be less familiar. They include Krispy Kreme (ticker: DNUT), Opendoor, Rocket Companies and Kohl's. The firms—a bakery chain, estate agent, mortgage provider and old-school retailer—have market valuations ranging between $650m and $31bn. One thing that binds them is the lack of love shown by hedge funds, which have been betting aggressively against all of them. Another is the abundance of enthusiasm shown by retail investors, who are snapping up the shares in the hope of squeezing the short-sellers and driving up the price. During July the price of Opendoor alone rose by over 500% before falling back. A new meme-stock frenzy has begun.

It mirrors the mania that sent the share prices of GameStop and AMC, a cinema chain, rocketing four years ago. Then, as now, the furious trading had little to do with financial performance. Speculative activity was once blamed on government stimulus cheques and low interest rates. But that story now looks less convincing. Perhaps the exuberance instead reflects changes in the investment technology available to retail investors.

Today's mania goes beyond meme stocks. Research by analysts at Goldman Sachs, a bank, suggests that speculative trading (in penny stocks, unprofitable firms and companies with the loftiest valuations) has climbed to levels seen only twice before: during the previous boom that peaked in 2021 and back when the dotcom bubble inflated in the late 1990s. Today's trading remains well below those peaks, but far above what was previously considered normal.

Chart: The Economist

Other signs of fervour abound. Transactions in zero-day options, contracts favoured by day-traders that expire the same day, have surged in recent years. According to Cboe Global Markets, an exchange, a record 2.1m changed hands in the second quarter of 2025, up from 1.4m a year earlier. The exchange thinks retail investors account for at least half of such trading.

In 2020 and 2021 the combination of loose monetary and fiscal policy was often credited for the surge in speculation. Covid-era stimulus cheques (or “stimmies”) delivered cash directly to budding retail traders just as everyone was forced to stay at home all day. Robin Greenwood, Toomas Laarits and Jeffrey Wurgler, three academics, tested the theory in 2023. They found that the handouts in both Hong Kong and America led to an abnormal boost in the price of stocks popular with retail investors.

That cannot be the explanation today. The last stimmies landed on American doorsteps four years ago. Monetary policy, meanwhile, has been tightened: the Federal Reserve’s balance-sheet has shrunk and ten-year Treasury yields are above 4%. Researchers at the San Francisco Fed reckon Americans’ pandemic-era savings were exhausted more than a year ago.

But even without the help of stimulus money, retail investors with a zeal for speculation continue to transform the markets. Such traders now account for about 20% of total American trading volume. That is down from around 24% at the 2021 peak, but well above the 10-16% of the 2010s.

The change has been driven not by government handouts, but by technology. App-based platforms have given individuals easy access to leverage and a wide range of securities to choose from. It is no surprise that the share price of Robinhood, among the largest of the new online brokers, has risen by 169% this year. Investing now appeals to a greater range of Americans. According to JPMorgan Chase, more of the bank’s customers are transferring money to an investment account. The share of high-income customers doing so doubled between 2015 and 2023. For low-income customers, the share quadrupled.

Beyond greater speculation, the long-term consequences of a more retail-heavy market are not yet clear. Some investors fear it is already having a negative effect. Cliff Asness, founder of AQR, a quant fund, thinks it is perhaps the biggest contributor to declining market efficiency. “Technology, gamified 24/7 trading on your phone, and social media in particular are the biggest culprits,” he wrote in 2024.

The big test will come during the next serious downturn. When the S&P 500 swooned earlier this year, falling by more than 20%, retail traders dashed to buy the dip, and a rapid recovery followed. But the drop was a modest one by historical standards. In a deeper slump, paired with a recession, would the DORK-buyers and their brethren be so brave? Sooner or later, investors will find out.

The trade deal with America shows the limits of the EU's power

The bloc opts for prudence over defiance

Photograph: Getty Images

"To be free, you need to be feared," said Emmanuel Macron, the French president, on July 30th. "We were not feared enough." He was speaking three days after Ursula von der Leyen, the European Commission's boss, had agreed a one-sided trade deal with President Donald Trump at his Turnberry golf course in Scotland. The agreement entails higher tariffs on European goods without any retaliation in kind. This humbling asymmetry was quickly condemned by the continent's politicians and press. But the European Union (EU) had good reasons for its faint-heartedness. It had to make enough concessions to keep Mr Trump engaged in Europe, while limiting the damage to its own economy. For now, it seems to have accomplished that.

The ugliest part of the deal is a tariff of 15% on European exports to America. That is more than nine times higher than the rate in place before Mr Trump returned to office. But it could have been worse: earlier in July Mr Trump had threatened a tariff of 30%. The agreed rate is also good news for Europe's carmakers, which had previously faced a 25% levy. Indeed, the agreement may have the perverse consequence of favouring European car exporters over rival manufacturers in America, where they must pay hefty duties on parts and steel.

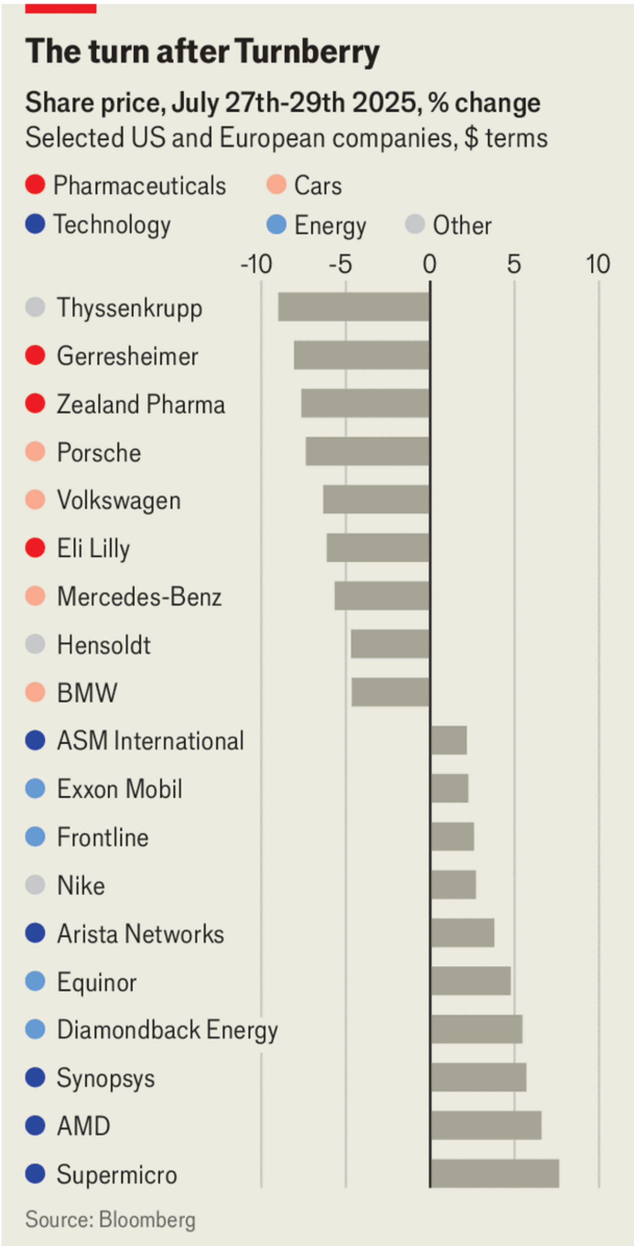


Chart: The Economist

The EU's other big concession was cutting its own tariffs on American cars and other industrial goods to zero, as well as giving some farm and fishery products better access to the European market. The liberalisation affects about a quarter of America's exports to the EU, including cars and footwear; the rest already enter tariff-free.

America has also made a concession of sorts that will help its own businesses. A list of European goods deemed “strategic” will face the old, low tariffs that prevailed before Mr Trump’s second term in office. These items include chipmaking equipment, natural resources and certain chemicals. The news lifted the share prices of European makers of these favoured products (see chart). The list could expand over time, giving Mr Trump a quiet way to ease the pain of his tariff madness.

The EU secured a few, subtler wins. It kept Mr Trump’s hands off its value-added tax, which the American president hates, because he wrongly thinks it disadvantages his country’s exporters. The deal said nothing about the EU’s right to tax and regulate digital services, many of which are of American origin. Nor did the deal dictate how Europe deals with China. The two sides did promise to work together to remove various non-tariff barriers to their trade. They also committed to ring-fence the transatlantic metals market against the threat of “overcapacity” (read: China). But these promises are hardly new; the EU had worked on similar initiatives with the Biden administration.

The economic impact on Europe is unclear. Goldman Sachs reckons that the hit could be a hefty 0.4% of GDP. Others are more sanguine. The calculations depend partly on how America’s other deals affect Europe’s relative position. The damage must also be set against the benefits of the deal. It has reduced uncertainty, which should help businesses make plans. The reduced tariffs on American goods could help bring down inflation, which would lead to lower interest rates. The combined effect of all this could be zero.

The talks were also about more than just tariffs. The EU sees the deal as the starting point from which to improve relations with America. Especially on defence, Europe hopes to keep America engaged. In that spirit, Mr Trump said the EU promised to buy “vast amounts” of American weaponry. The EU denied this: it does not consider weapons part of the deal.

The handshake in Scotland is not the end of the saga. The two sides must first come to a common interpretation of what was agreed. On metals, they seem to differ on whether old tariffs or old quotas rule. On pharmaceuticals, the language is clearer that the 15% tariffs will apply. But Mr Trump remains exercised by drug prices. An investigation into American pharma imports is under way and could result in tariffs on drugs made abroad. Should America make new, more onerous demands, the EU may set aside caution and retaliate against the bully that does not fear it enough. In this negotiation, it has accepted the limits of its power. In the next round it might be Mr Trump’s turn to learn his.

The deeper reason for banking's retreat

Why bankers no longer play golf at 3pm

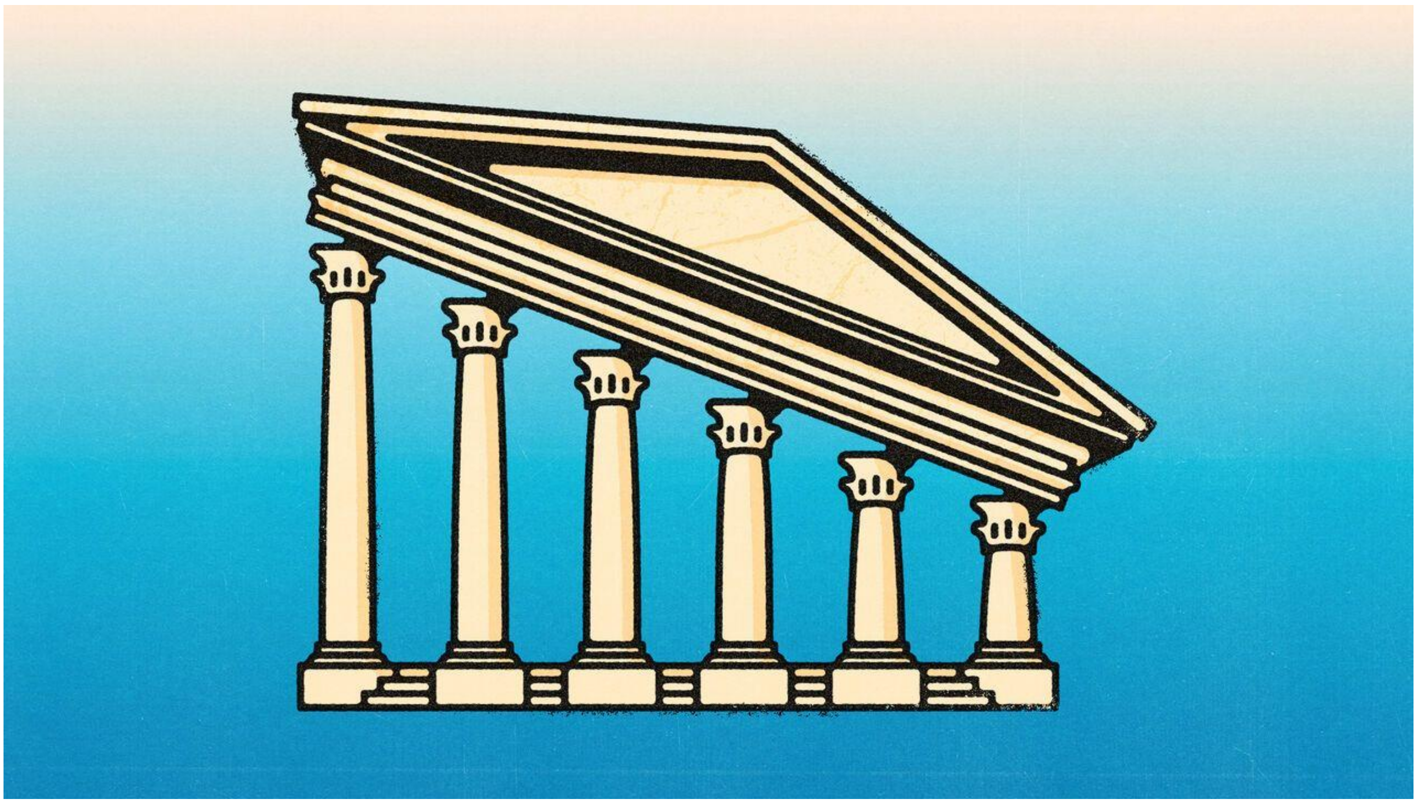


Illustration: Alberto Miranda

In an earnings call on July 15th, Jamie Dimon, the boss of JPMorgan Chase, made a familiar complaint. He rattled off a litany of burdensome, overlapping regulations: "SLR, G-SIFI, CCAR, Basel III, FSRT". He then called on regulators to draw "a deep breath", step back and take stock. Reform was necessary, he said, to "create more liquidity, more loans and a safer system."

Many people share his frustration. They argue that large commercial banks have been hobbled by regulatory red tape in the aftermath of the global financial crisis of 2007-09. As a regrettable but predictable consequence, these giants have pulled back from lending. That has left a vacuum increasingly filled by fintech upstarts and private credit—funds lent by asset managers, not banks.
This story is not entirely wrong. But it misses a longer arc in the history of credit. Since the 1960s direct lending by banks to firms and households has steadily declined. This shift began decades before

the financial crisis and reflects deeper forces than overzealous regulation. Much of what is now blamed on recent policies is simply part of this slower-moving retreat.

To understand this transformation, recall the traditional banking model, popularised in movies like "It's a Wonderful Life". Banks accepted deposits, promising to redeem them on demand, with interest. They lent to firms and households in the form of mortgages, small-business loans and consumer credit. The bank's job was clear: safeguard deposits, evaluate creditworthiness and monitor borrowers. The path to profits was straightforward, too. Banks earned the spread between what they paid their depositors and what they earned on loans. People would joke about the "3-6-3 rule": borrow at 3%, lend at 6%, hit the golf course by 3pm.

Tee-bonds

The mid-afternoon golfers once ruled the world of credit. In 1974 nearly 55% of private lending in America was held on banks' balance-sheets as direct loans. But that world had disappeared well before the financial crisis struck. By the beginning of this century, this form of credit accounted for less than 35% of private lending, where its share remains to this day (see chart 1).

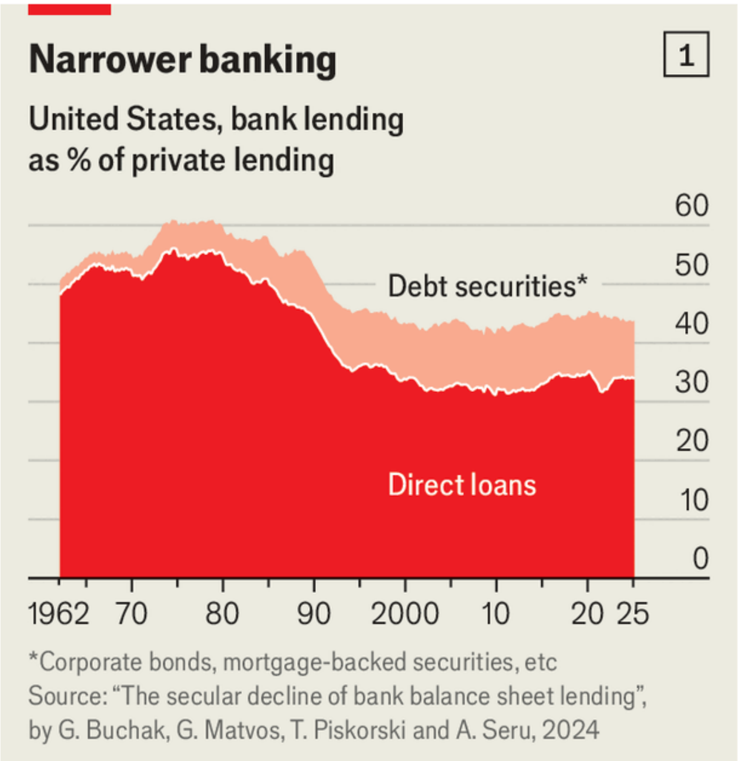


Chart: The Economist

In place of the traditional bank loan is a growing stack of debt-linked securities. These include mortgage-backed bonds, corporate paper and loans to private-credit firms. Banks still collect deposits from households and most still originate loans. But instead of holding these credits on their balance-sheet, they sell them on, while buying slices of securities at market prices from other lenders.

In doing so, they have gradually ceded the role of credit provision to others. Even as banks' size in the economy has grown—their assets have risen from 60% of GDP in 1960 to 94% today—their contribution to lending has stayed flat, at around half (see chart 2).

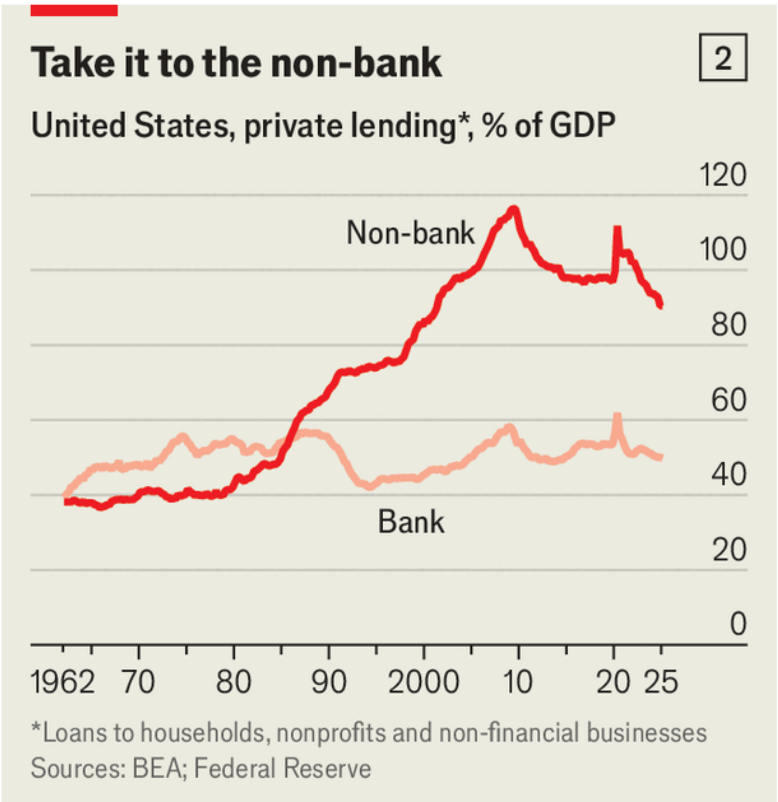


Chart: The Economist

Households and firms instead borrow from a litany of other lenders and guarantors who package and sell the debt to banks. Big companies have increasingly switched to public bond markets for debt financing. Small businesses have turned to fintech and online lenders. Buy-out firms have turned to private credit. Households still rely on banks for mortgages, but these sit on the banks' balance-sheets only after being repackaged and securitised by government agencies.

Why the change? The kind of regulatory impositions highlighted by Mr Dimon have certainly played a part. Recent policy, for example, explains why banks have surrendered much of private-equity lending to other institutions, according to research by Sergey Chernenko of Purdue University and his co-authors. Regulators require banks to hold more capital when making direct loans to firms than they do when buying securities backed by similar loans.

But the shift has deeper origins, as documented in research by Greg Buchak of Stanford University and others. Technological developments, ranging from securitisation software to FICO scores, led to the development of new financial instruments and alternative securities markets. Before 1980 the corporate-bond market was less mature; Fannie Mae, a government-sponsored enterprise, had yet to sell any mortgage-backed securities; and no one needed to understand collateralised loan obligations (which slice up loans and apportion them to investors according to their risk appetites). Today such institutions are entrenched parts of the financial system. Savers, too, began to act differently. They gravitated away from bank deposits towards pensions and money-market funds that channelled their savings into credit markets.

To the characters from “It’s a Wonderful Life”, today’s banks would be a mix of the familiar and the strange. Now, as then, banks use deposits to hold debt-related assets. And now, as then, creditworthiness matters. Losses still hit the balance-sheet whether a household defaults on a mortgage the bank made itself or on one buried in a securitised pool.

But today’s banks are less dominant. And it is not just that their lending is taking different, indirect forms. Nina Boyarchenko and Leonardo Elias of the New York Fed, who trace credit flows through layers of intermediaries, find that net bank lending to households has fallen by about a quarter over the past 50 years.

Is the bankers’ retreat a bad thing? One inconvenient consequence is also faintly ironic. In the old days policymakers could keep the financial system stable by policing a finite number of familiar, marble-floored institutions. Now, as banks play a smaller role and credit easily migrates elsewhere, policymakers must cast a wider net in their pursuit of financial stability. A phenomenon often blamed on regulation may have complicated regulators’ lives.

More happily, banks have become less fragile—less dependent on the unstable alchemy of funding long, opaque, illiquid loans with short, runnable deposits. Rather than holding entire portfolios of mortgages, banks can now hold the safer slices of securitised debt, which get paid first when things go wrong. The riskier slices can be pushed to investors with longer horizons (such as insurance companies) or greater risk appetite (such as hedge funds). Some research suggests that credit booms led by non-bank lenders tend to be less destabilising than other expansions. They are associated with a lower risk of GDP crashes than bank-led booms, according to Ms Boyarchenko and Mr Elias.

The transformation of banking also sheds light on the industry's defining characteristics. Academics have long debated whether banking's edge lies on the liability side of the business (in issuing sticky, low-cost deposits) or on the asset side (in originating, monitoring and collecting loans). The fact that banks have grown so large, while ceding much of their lending business to others, could help resolve that debate. It suggests the secret of their success lies in courting depositors, not marshalling borrowers.

The good news for bankers is that a modern version of the 3-6-3 rule still applies. Banks can still borrow cheaply. Indeed, they now pay depositors closer to 0.5% rather than the 3% of old. They lend, or rather invest, in a broad mix of securities yielding anywhere from 1% to 10%. And as for the 3pm tee time? That, too, has shifted: Stanford researchers say the most popular midweek golf slot is now 4pm.

Tecnology

China has top-flight AI models. But it is struggling to run them

Trump's U-turn on chip-export controls could be a boon

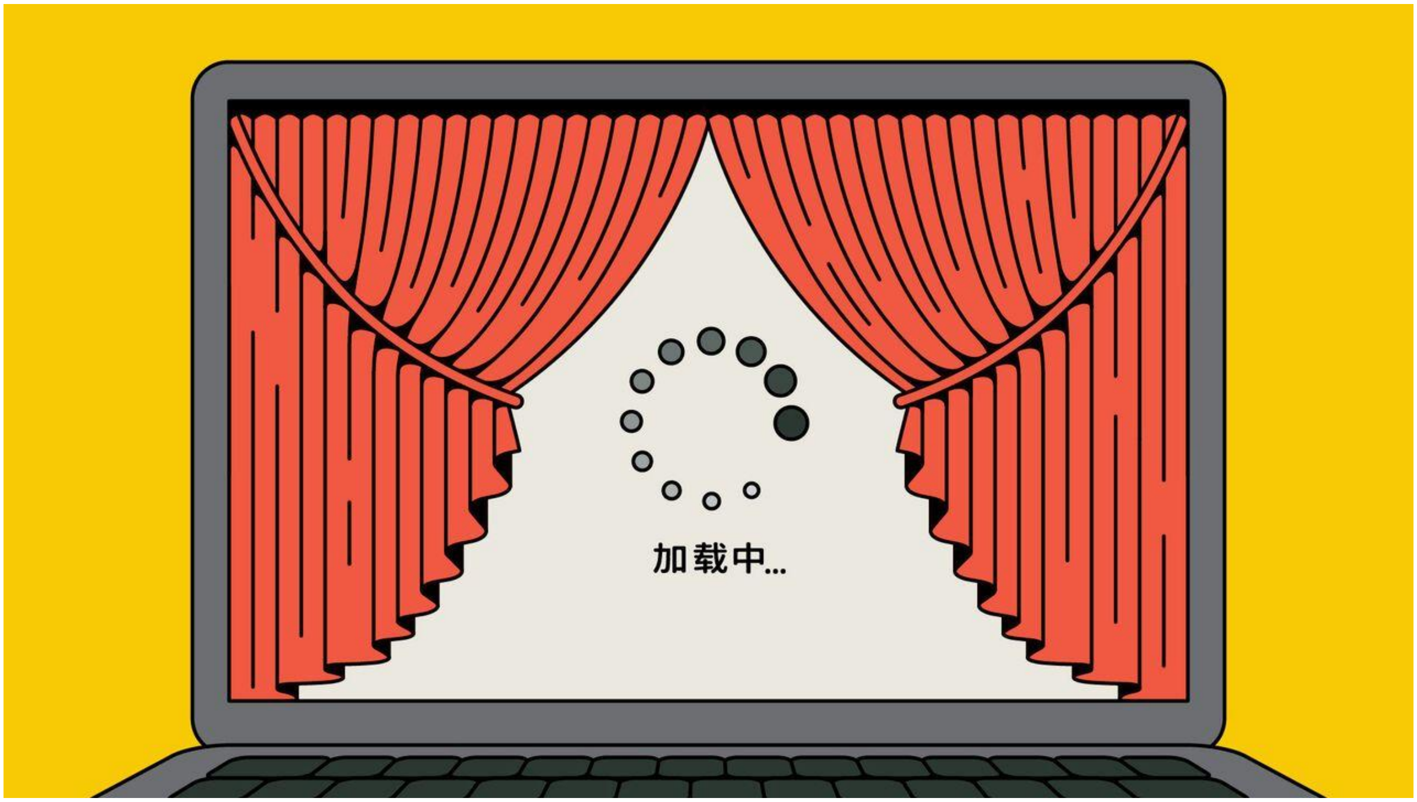


Illustration: Fortunate Joaquin

Six months ago DeepSeek, a Chinese artificial-intelligence (AI) firm, wowed the world with the v3 model and its successors. For the first time, a country other than America—and one that America had cut off from the supply of top-of-the-range semiconductor chips—was producing open-source models that rivalled those designed in Silicon Valley.

Despite the restrictions, Chinese firms kept training world-beating AI models—Kimi K2, unveiled in July by Moonshot AI, a Beijing-based lab founded by an alumnus of Google and Meta, rose straight to the top of the global leaderboards. With more parameters, as the connections between a model's artificial neurons are called, than any open-source equivalent, Kimi K2 outperformed its Western rivals ChatGPT 4.1 on tests of coding ability and Claude 4 Opus on tests of science knowledge.

But for models to really impress, they need to be used. This is where chip restrictions have bitten the hardest. Shortages have affected the data centres AI labs need to run their systems once trained. Slowdowns, usage limits and dropped connections are becoming common. “We’ve heard your feedback—Kimi K2 is SLOOOOOOOOOOOOW,” Moonshot posted on X a few days after the launch. DeepSeek, meanwhile, has delayed the launch of its latest AI model to avoid similar performance issues, according to a report from the Information. And so both companies were given cause to celebrate two weeks ago, when the White House reversed its latest export controls, once again allowing Nvidia to sell its H20 chips in China. Making these available to tech companies there will remove the hurdles currently slowing their growth.

China is fertile ground for an AI boom: the country has millions of science and engineering graduates, spare grid capacity, the political will to build data centres as fast as concrete can be poured, and access to all the West’s public data sources and more of its own. It lacks a home-grown source of computing power, however, a fundamental constraint that has so far shaped the development of its industry.

In the past few months Chinese firms have found many ways to work round American restrictions. Banned chips worth $1bn have entered the country since April and domestic companies, such as Huawei, have developed chips to match Nvidia’s top-end offering in some respects (though at smaller volumes). A relentless focus on efficiency has also led to breakthroughs.

Limited access to chips also explains another feature of the Chinese AI sector that has baffled outsiders: the devotion to open-source releases. DeepSeek v3 and Kimi K2 are both available through third-party hosting services such as Hugging Face, based in New York, as well as to download and run on users’ own hardware. That helps ensure that, even if the company lacks the computing power to serve customers directly, support for its models is still available elsewhere. And the open-source releases serve as an end-run round hardware bans: if DeepSeek cannot easily acquire Nvidia chips, Hugging Face can.

Not all Chinese firms have been equally affected by the restrictions. On Friday Alibaba released the latest model in its Qwen3 family, an open-source reasoning model called Qwen3-235B-A22B-Thinking-2507. The release brings Qwen, and Chinese AI in general, level with not just the best open-source AI models, but the best AI models full stop.

Alibaba's system is around a quarter the size of K2, requiring commensurately less computing power to run, and, unlike DeepSeek and Moonshot, Alibaba has substantial cloud infrastructure behind it to keep the models working. Making models faster and more efficient to use is clearly the new game in the Chinese AI sector: on Monday another lab, Z.ai, released two models, called GLM-4.5 and 4.5 Air, explicitly touting their speed and efficiency.

But the canny workarounds and impressive models can stretch a resource constraint only so far. And since April, one limitation has bitten harder than any others: the loss of Nvidia's H20 chips.

Successful AI companies must be able to do two things: train models and then run them, a process known as inference. The best-funded Chinese labs have continued to launch training runs of comparable scale to their Western peers. But inference has proved trickier. Whereas training data centres need monolithic clusters of top-end chips, inference is best performed by chips that balance power, energy efficiency and the ability to move data at speed. Until April, the H20 was the chip of choice.

Worse, though a training run is an upfront expense that can be recouped as revenue over the lifetime of the model, a company that loses money during inference has no opportunity to make it up. That means access to chips for inference, not training, is the bottleneck limiting the growth of China's AI industry.

In response, the Trump administration has sent mixed signals. Its AI action plan, published in early July, doubled down on some chip controls, emphasising that denying adversaries access to "advanced AI compute" is a matter of both geostrategic competition and national security, and calling for novel approaches to enforcing export controls. At the same time, it has lifted the ban on H20 exports, arguing that it would be better for Chinese AI to rely on American companies for all their technology needs, including inference, than to develop an equivalent domestic capacity.

In the short term, such an easing will be cold comfort to China. Nvidia's own supply constraints mean it will be unable to meet the country's demand for chips until the last quarter of the year at the earliest. That means models which lean on efficient output and the ability to run on phones and laptops directly will continue to be prioritised for now. But if American exports pick up once more, then China's AI sector could, at long last, start 2026 much less constrained.

The Economist: *https://www.economist.com*