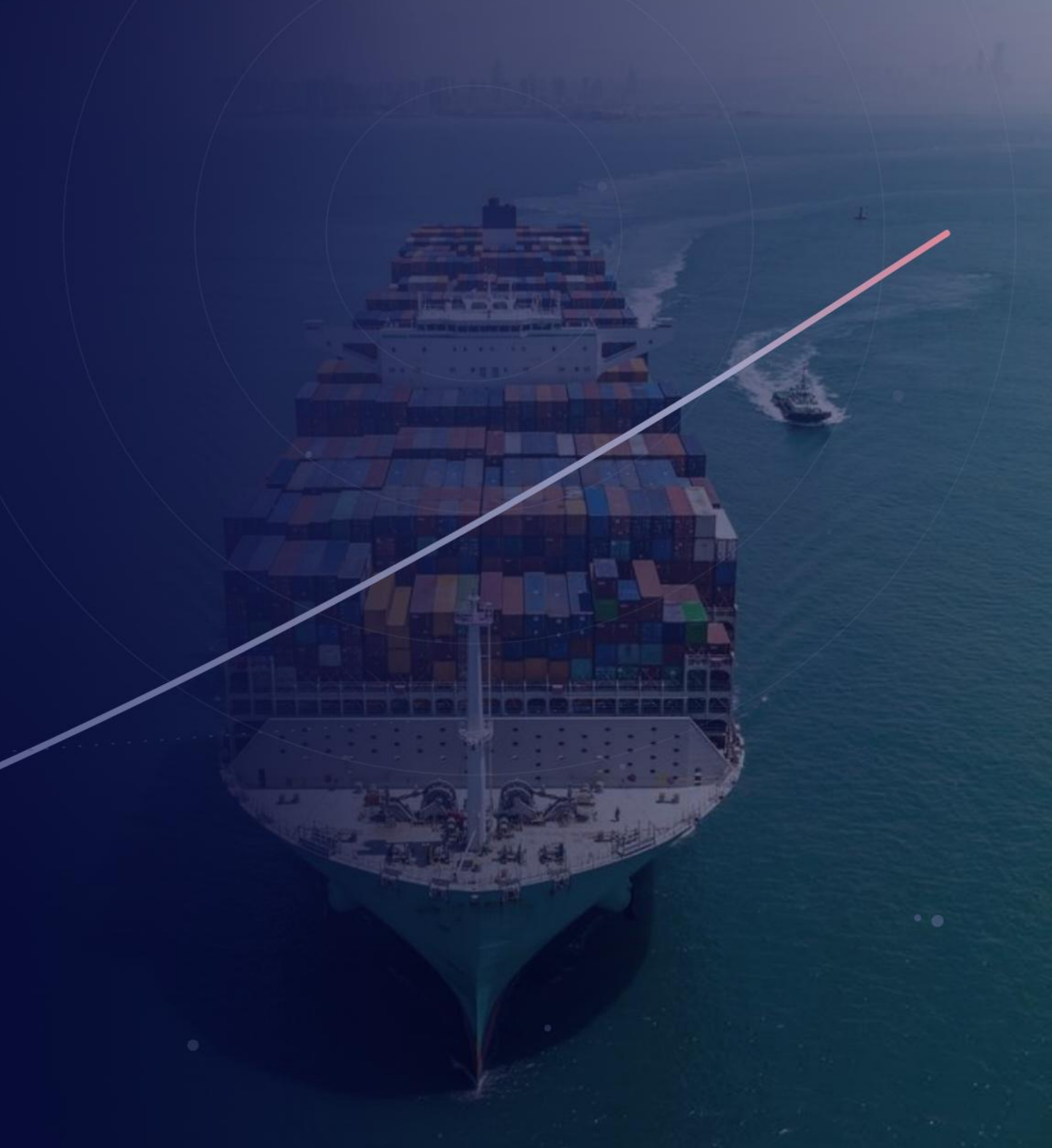
MOODY'S

THE JEVONS CENTURY

# Make thinking cheap enough, *and we find more to think about.*

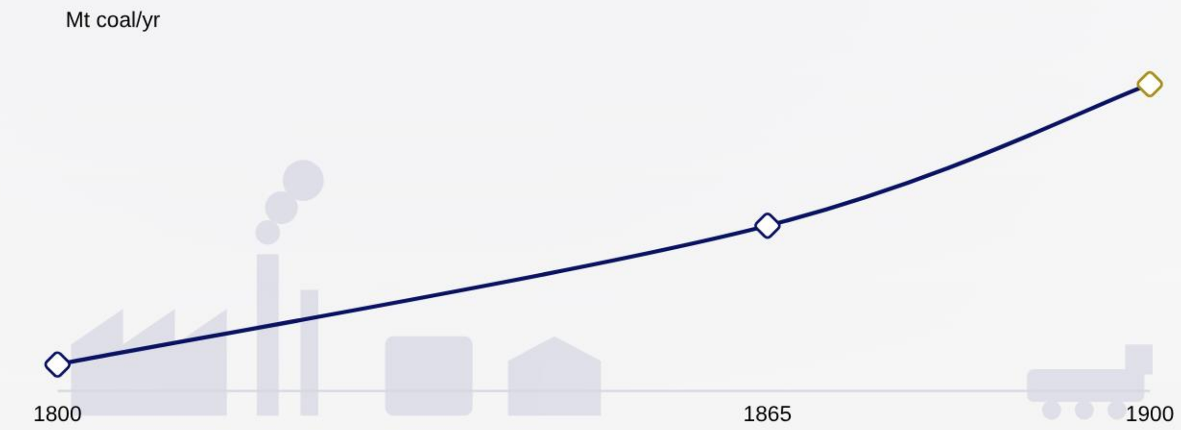


DATA FROZEN JULY 23, 2026

# Engines got efficient. *Britain burned more coal.*

By 1865 Britain feared running out of coal, and engineers offered reassurance: steam engines kept getting more efficient, so the same work would burn less fuel. William Stanley Jevons, in *The Coal Question*, said the comfort was misplaced. Efficiency had made coal cheaper to put to work, and cheaper meant it turned up everywhere — factories, railways, ships, streetlights. Each engine sipped where its grandfather had gulped, about ~15x less fuel for the same work, and the country burned more coal than ever.

> *“The very contrary is the truth.” Efficiency did not spare the coal. It found new uses for it.*

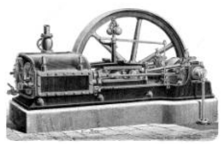



1800
● CARBON BRIEF · UK DESNZ/WARDE · 1800 **14.9** Mt

1865
● CARBON BRIEF · UK DESNZ/WARDE · 1865 **93** Mt

1900
● CARBON BRIEF · UK DESNZ/WARDE · 1900 **172.6** Mt

● KANDER/WARDE; UNIV. OF HOUSTON, ENGINES OF OUR INGENUITY · C.1770–1900



WATT'S SEPARATE CONDENSER, 1770S: THE EFFICIENCY LEAP THAT MADE COAL CHEAPER TO USE

Scrub the timeline 1900 · 172.6 Mt

● 1 · ENGINES GET FAR MORE EFFICIENT → ● 2 · USEFUL POWER GETS CHEAPER → ● 3 · NEW USES EXPLODE · COAL USE CLIMBS

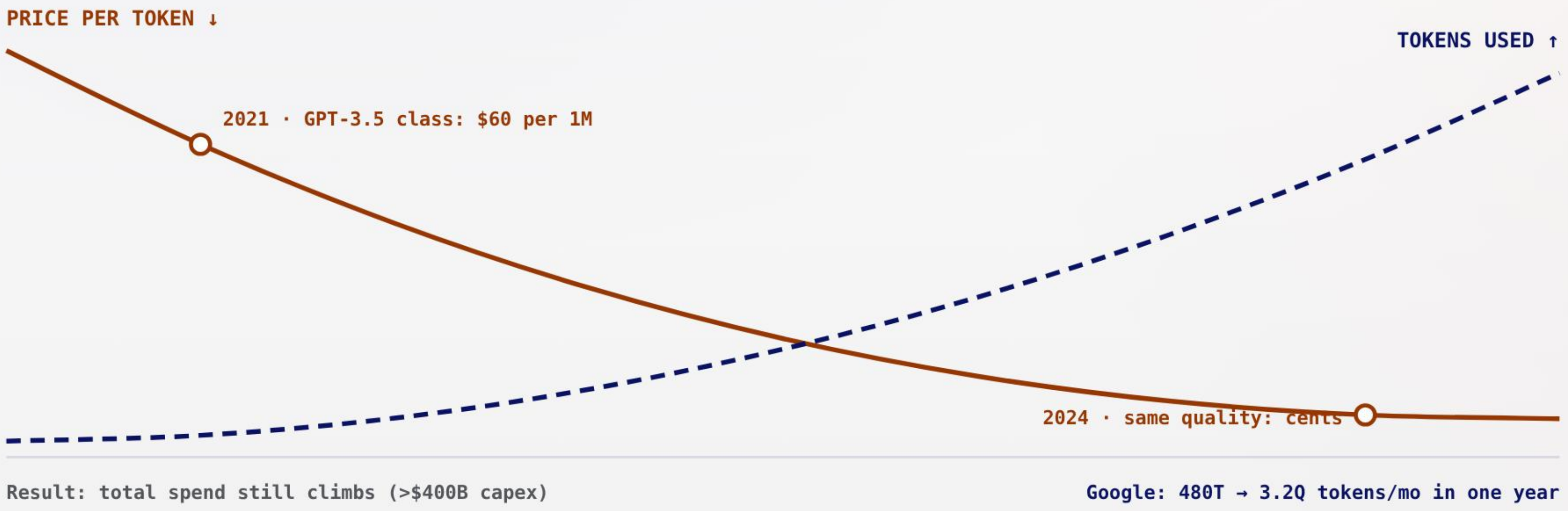
# The same curve, *running again*

Compute is rerunning the coal curve. Three views of it here: prices fall while usage soars, a price floor forms, and each task gets cheaper.

A · Price vs usage | B · Tokens vs tasks | C · Price floor

[ and ] switch panels

Two lines, moving in opposite directions. The price of a token keeps falling, and the number of tokens people use climbs much faster. Multiply the two and total spending still goes up.




• EPOCH AI · A16Z LLMFLATION · GOOGLE I/O 2026 · IEA

# The same curve, *running again*

Compute is rerunning the coal curve. Three views of it here: prices fall while usage soars, a price floor forms, and each task gets cheaper.

A · Price vs usage | **B · Tokens vs tasks** | C · Price floor

[ and ] switch panels

CHEAPER MODEL · GPT-5.5

67%

solved · **$7.23** per task

GPT-5.5 · DeepSWE v1.1 pass@1 / avg cost · fewest output tokens of any model

• DATACURVE DEEPSWE V1.1 LEADERBOARD, DEEPSWE.DATACURVE.AI · JUL 25, 2026

VS

BOARD LEADER · CLAUDE OPUS 5

74.0%

solved · **$11.84** per task

Claude Opus 5 · leads the board · released Jul 24 2026

• DATACURVE DEEPSWE V1.1 LEADERBOARD, DEEPSWE.DATACURVE.AI · JUL 25, 2026

WHAT THEY WERE ASKED

113 original tasks inside 91 live open-source codebases, across five languages. Written from scratch for the benchmark, which keeps the answers out of training data. A typical solution touches around 668 lines across seven files.

WHAT COUNTS AS SOLVED

One attempt per task, run alone in a sandbox with no internet. A hand-written verifier checks whether the software behaves correctly, not whether the code matches the author's own fix. Running out of context or time counts as a miss.

WHAT THE PRICE IS

The average API cost of that one attempt. Every model runs on the same agent harness, so the spread reflects the model itself.

**Mind the error bars.** The full suite is rerun four times, and pass@1 lands within roughly ±4–6 points. On those margins 67% against 74% is a real gap, and 67% against 68.5% is a tie. • DATACURVE DEEPSWE METHODOLOGY, DEEPSWE.DATACURVE.AI/BLOG/DEEPSWE · MAY 26, 2026

AND CHEAPER STILL **Kimi K3** is the best value on the whole board: **68.5% @ $4.65/task**, within six points of the leader at under half the cost.

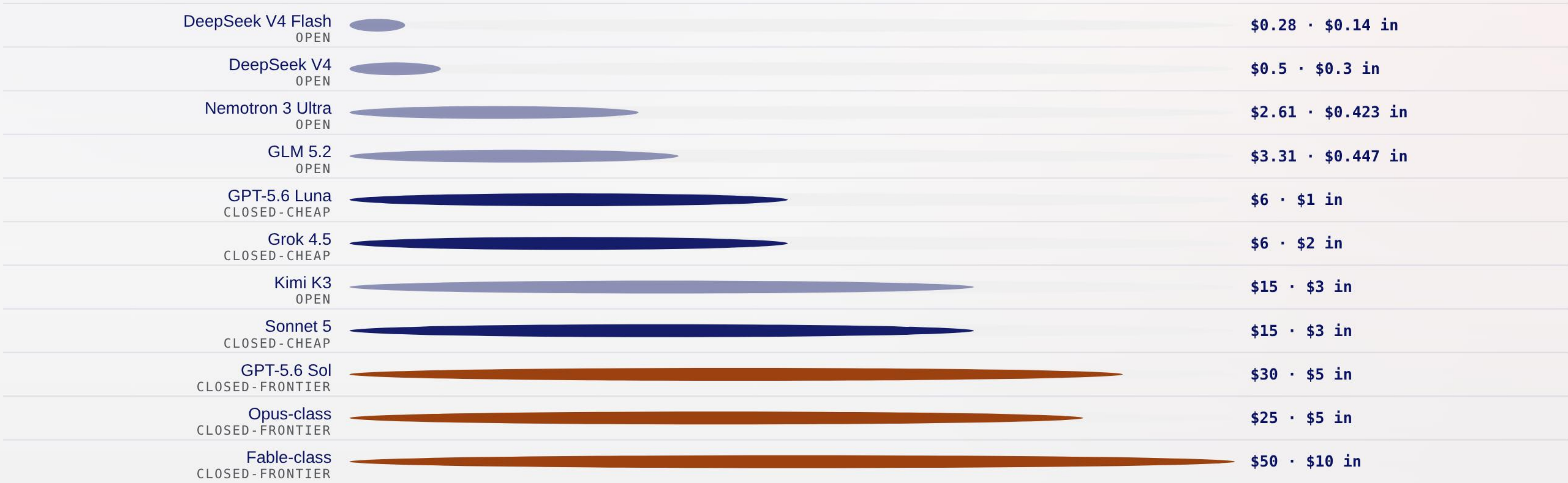
• DATACURVE DEEPSWE V1.1 LEADERBOARD, DEEPSWE.DATACURVE.AI · JUL 25, 2026

# The same curve, *running again*

Compute is rerunning the coal curve. Three views of it here: prices fall while usage soars, a price floor forms, and each task gets cheaper.

A · Price vs usage | B · Tokens vs tasks | **C · Price floor** [ and ] switch panels

$ per 1M tokens (output), log scale · lighter = open-weight · navy = closed-cheap · amber = closed-frontier




+ Show data table

# Constraint one: *electricity*

Cheap tokens still have two inputs that stay scarce: the electricity that makes them and the people who direct them. Start at the grid, where the multiplication shows up as load. Data-centre electricity is set to climb **485 → ~950 TWh** by 2030, and about **~50%** of new US electricity demand comes from data centres alone. Click a node to see what it carries.

Play the chain

Steps through the chain, one link at a time


Energy
**485 → ~950 TWh**

→

Compute
**3.4x/yr**

→

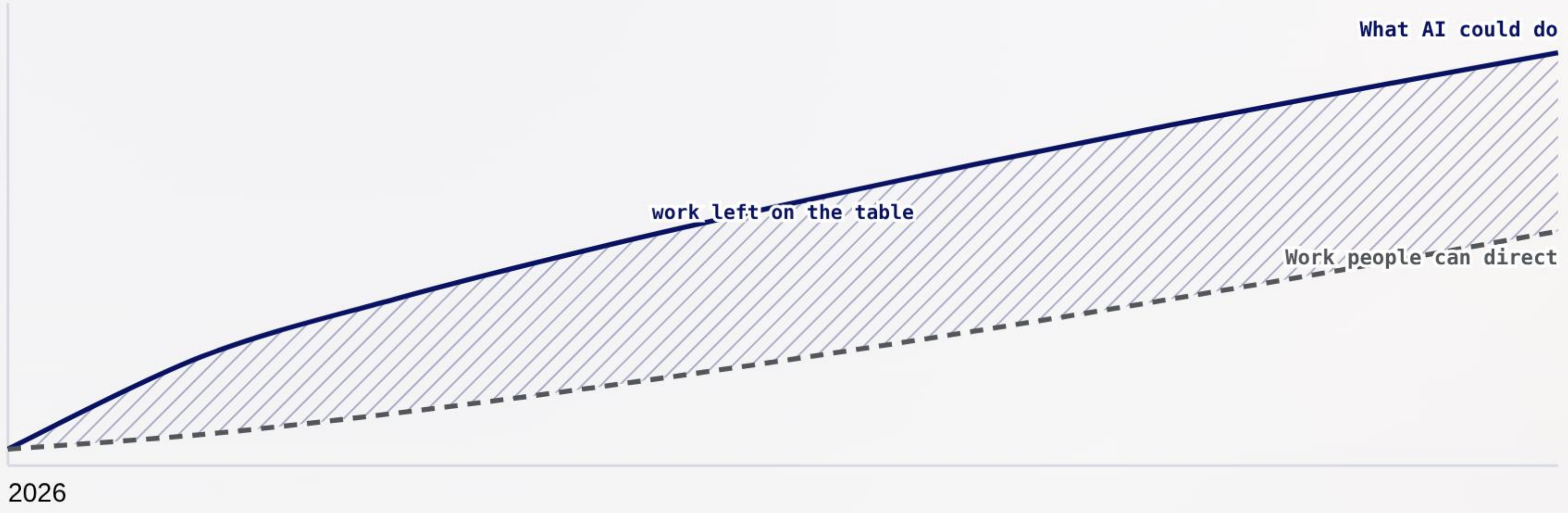
Tokens
**480T → 3.2Q /mo**

→

Work

# Constraint two: *people*

For the first time, AI skills top employers' list of hardest-to-find capabilities — and **72%** of employers still report difficulty filling roles. Cheap models created more work than there are people to direct it.




Play 2023 → 2026

or scrub the years

Top line: what AI could do each year. Bottom line: how much of that companies could actually put to use, capped by the number of people who can direct it. The striped gap is work left on the table.

● MANPOWERGROUP 2026 TALENT SHORTAGE SURVEY · 2026 ● JOBSPIKR 2026 · 2026 ● IDC · 2026

SO WHAT Cheap, capable AI needs somewhere to go. The largest pile of unread paper on earth sits in trade finance.

# Four billion pieces of paper move the *world's trade*

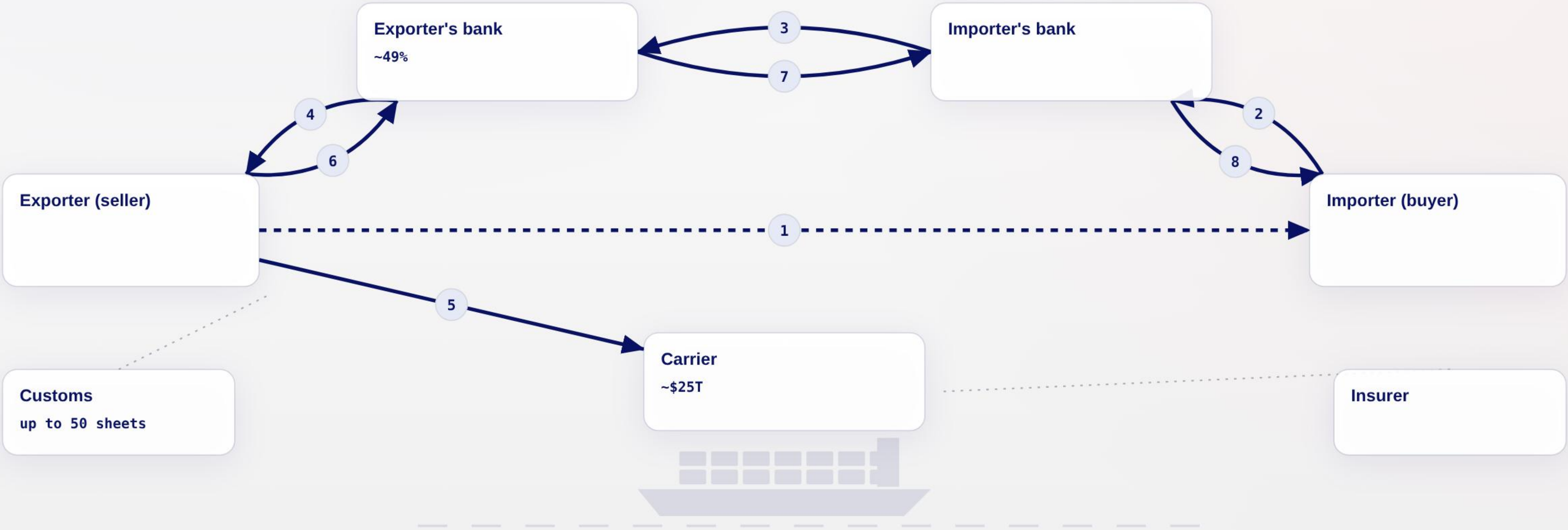
This is that pile. World trade still moves on paper, and a letter of credit is how banks make strangers trust each other. Here it is in its simplest form. One shipment can involve **up to 50 sheets · 30 stakeholders**. Even among ICC-surveyed firms, only **~49%** have touched an electronic bill of lading — and paper still carries nearly all the volume. Press play to follow the money and the documents.

Play the flow

Every one of those hops is paper. Four billion sheets a year.

**4,000,000,000**

paper documents circulating in global trade every year ● ICC · 2024



# Where the risk *hides*

Two numbers, often confused. Fake trade invoices **launder** an estimated **$1.6T–$2.2T**[1] a year. Separately, fraud **steals** over **$500B+**[2] a year. A decade of court cases caught about **~$60B**[3] of the laundering.

**A · The detection gap** B · The losses, sorted [ and ] switch panels

LAUNDERED · EVERY YEAR

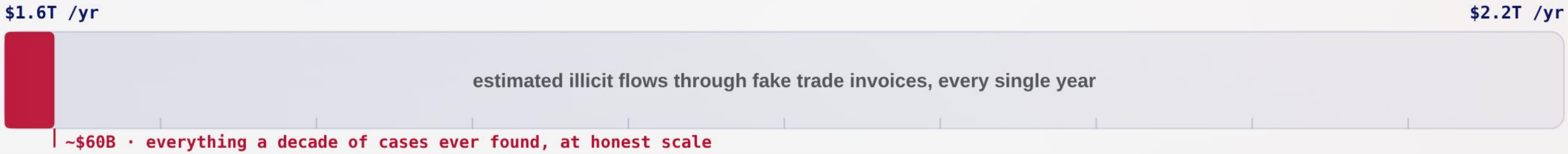
**$1.6T–$2.2T**

moved through fake trade invoices, **each year**

CAUGHT · IN TEN YEARS

**~$60B**

found across **ten years** of court cases




**1** FATF estimates, 2024 **2** Nasdaq Verafin Global Financial Crime Report 2026, 2025 **3** Global Financial Integrity, 2023, 2011–2021

● FATF ESTIMATES · 2024 ● GLOBAL FINANCIAL INTEGRITY, 2023 · 2011–2021

Regulators keep tightening: FATF 5th round (2024) · EU AMLR Reg. 2024/1624 · FinCEN TBML advisories

**Ten years of investigations caught about $60B. Meanwhile, trillions more flow through unchecked every single year.**

# Where the risk *hides*

Two numbers, often confused. Fake trade invoices **launder** an estimated **$1.6T–$2.2T**[1] a year. Separately, fraud **steals** over **$500B+**[2] a year. A decade of court cases caught about **~$60B**[3] of the laundering.

A · The detection gap | **B · The losses, sorted**

[ and ] switch panels

These three numbers get quoted interchangeably, and they should not be. One is money moved, one is money gone, and the last is a single company. Largest to smallest.


And the enforcement response is small against all of it: **$4.6B** in global AML fines in 2024, the year TD Bank alone paid **$3.09B**.

- FATF · GFI · US DHS OIG 2023 · 2023
- NASDAQ VERAFIN GLOBAL FINANCIAL CRIME REPORT 2026 · 2025
- FENERGO, FEB 2025 · 2024
- US DOJ / FINCEN · 2024-10

MOODY'S

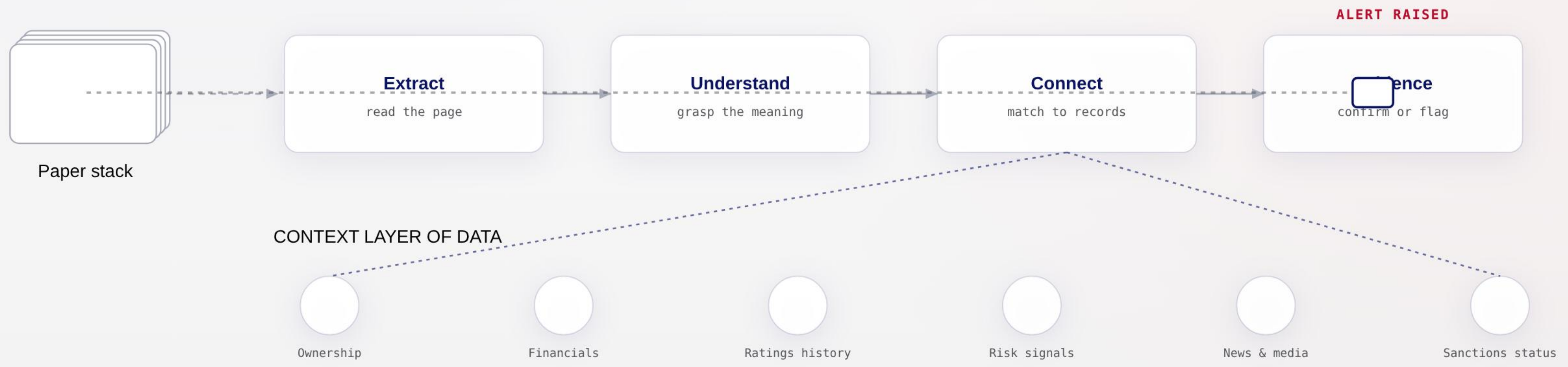
08

# The bridge from physical to digital

A paper stack becomes structured data in four steps: Extract, Understand, Connect, Evidence. What comes out lands in the context layer, where one document can confirm what other sources already suggest.

Send a document across

Press again to skip



MOODY'S

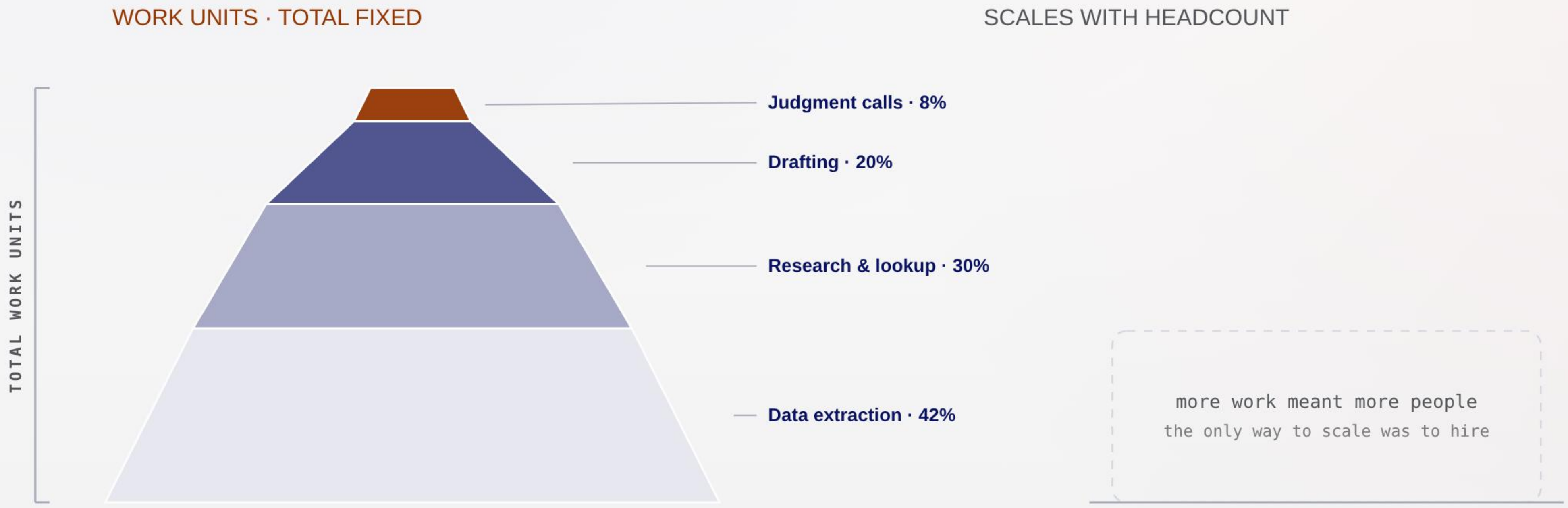
# Same work. *The pyramid flips.*

Once machines read everything, AI absorbs the repetitive base and human time concentrates at the top, where judgment lives. Slide the year to watch it invert.

1995 2026 Year 1995 tokens | tasks

**1995: the pyramid everyone knows. Most human hours sink into the tedious base.**

Illustrative. Before the token era, so no frozen figures apply.




Unit price **n/a** Work volume **n/a** Spend **n/a**

MOODY'S

09

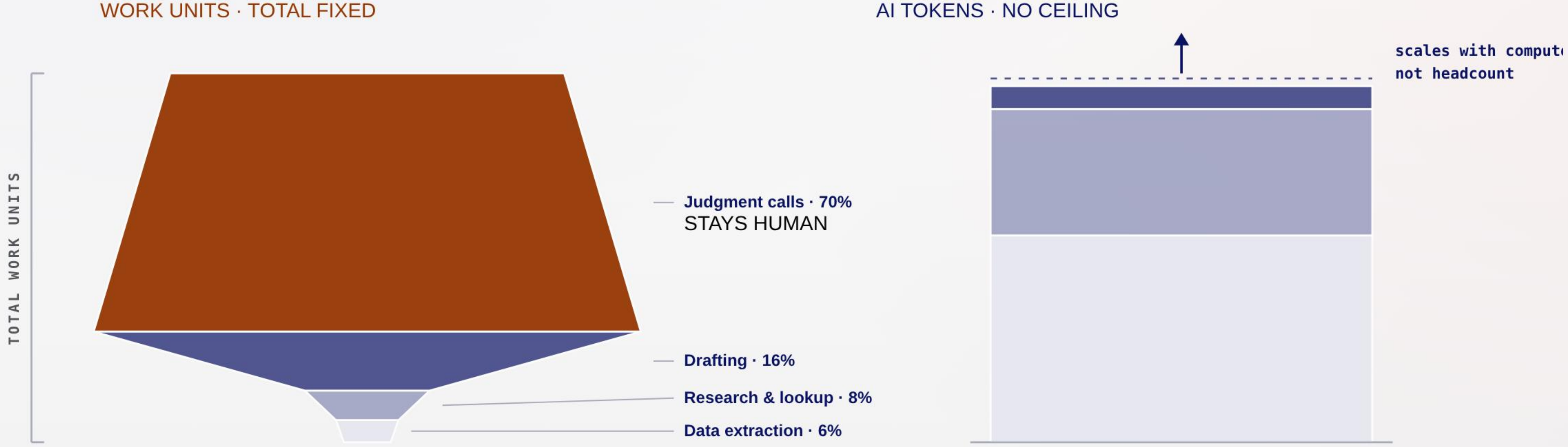
# Same work. *The pyramid flips.*

Once machines read everything, AI absorbs the repetitive base and human time concentrates at the top, where judgment lives. Slide the year to watch it invert.

1995 | **2026** | Year | 2026 | **tokens | tasks**

**2026: same total work, pyramid inverted. Judgment has taken the largest share.**

Hour shares illustrative. The price and spend figures below are sourced (as of July 23, 2026).




Cheapest useful token **$0.28/1M out** How fast prices fall **~50x per year (median)** What companies spend anyway **>$400B**

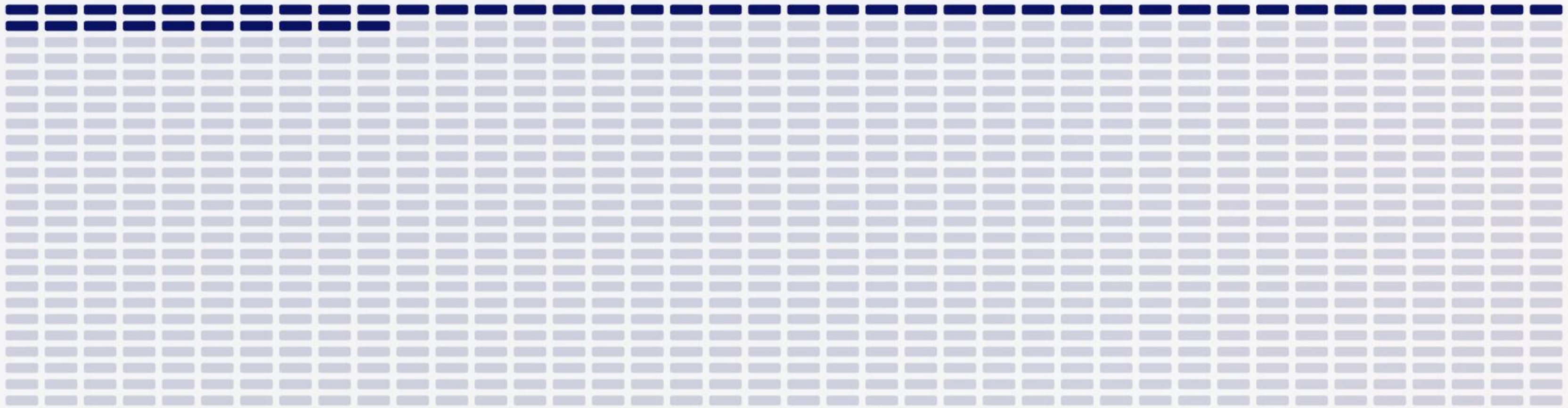
• EPOCH AI · 2025 • GOOGLE I/O 2026 KEYNOTE · 2026-05 • IEA, KEY QUESTIONS ON ENERGY AND AI · 2026 • DEEPSEEK RATE CARD

# From sampling to *coverage*

Reviewers only ever see a small slice of the documents moving through trade and finance. The cost of reading one has fallen far enough to change that. (The 5% is illustrative.)

**Switch to full coverage**

Coverage: 5%



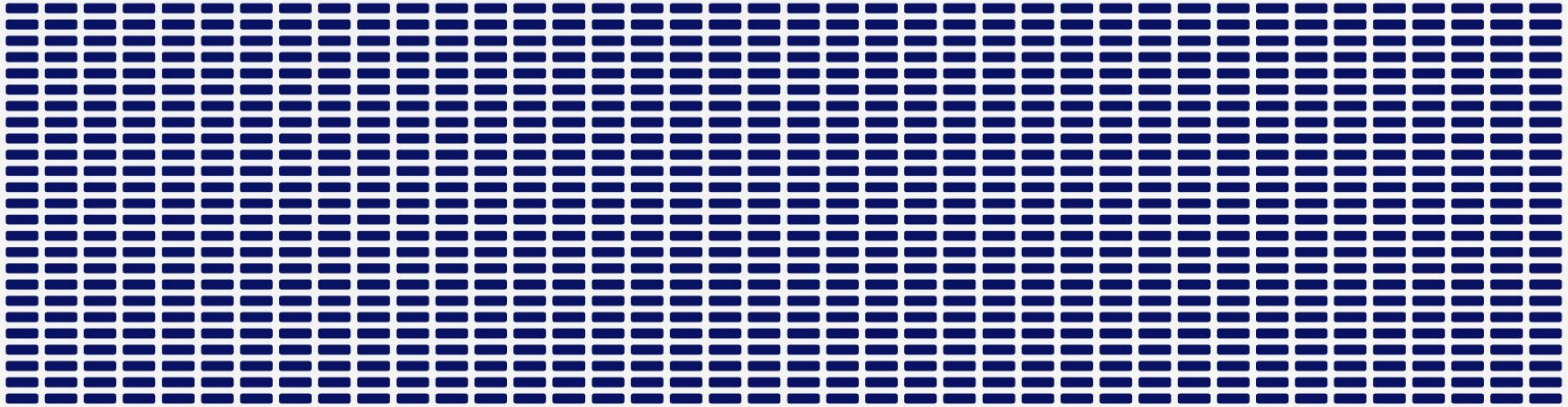
SO WHAT This is the Jevons dividend: check everything, every time.

# From sampling to *coverage*

Reviewers only ever see a small slice of the documents moving through trade and finance. The cost of reading one has fallen far enough to change that. (The 5% is illustrative.)

Reset to sample view

Coverage: 100%



• OWNERSHIP • FINANCIALS • LEGAL FILINGS • SANCTIONS STATUS • NEWS • SHELL COMPANIES • TARIFFS

SO WHAT This is the Jevons dividend: check everything, every time.

# What stays human

Machines are taking over the repeatable parts of knowledge work. Four kinds of hours are left.

**JUDGMENT**

Deciding what a signal means once it shows up.

**CONTEXT**

Knowing which history makes a document plausible.

**EDGE CASES**

The case the pattern has not seen yet.

**ACCOUNTABILITY**

Standing behind a call, even on the bad days.

> *The cheaper the routine work gets, the more you are hiring for the exceptions.*

# The Jevons century, *and the work it leaves us*

Coal taught us the shape of this. The cheap thing gets used far more than anyone forecasts, and the work reorganises around whatever stays scarce. Cognition is the cheap thing now. Judgment is what stays scarce.

> *The reading is solved. The deciding is ours.*

**MOODY'S**
*the context layer*

WHAT JUST ENDED

The old scarcity. Nobody could read it all, so we sampled. Hidden risk lived on the remainder.

WHAT REPLACES IT

Reading everything costs less than missing one thing. The scarce hour moves from finding to deciding.

WHAT TO DO ABOUT IT

Rebuild the job around that hour. Hire for judgment. Treat the model as a colleague you supervise.

+ All figures in this deck are dated and sourced (as of July 23, 2026). Show all 35 sources