

MetisLLM

Case study · eighteen months of a research flow

What the platform produced from
962 research reports — back-tested.

Eighteen months of one institution's inbound sell-side research, resolved into standing views with dated stance histories — then replayed against what the market actually did. Banks are anonymised throughout; outcomes are from public FRED series anyone can re-pull.

Not investment advice. Episodes selected with outcomes known — method and limits inside.

WHAT WE RAN, AND HOW TO READ IT

Point the platform at a past date.
Read its output. Then look at what happened.

- 01 Freeze the corpus
Only documents published before the as-of date are visible to the platform.
- 02 Read what it produced
The standing views, splits and stance histories it assembled — as of that date.
- 03 Only then, the outcome
Public FRED series anyone can re-pull. No private data, no proprietary marks.

Everything that follows is platform output — **not curation**.
The strongest episodes are ones where the street was *wrong*.

962

REPORTS

26,368

PAGES

257.5k

FACTS

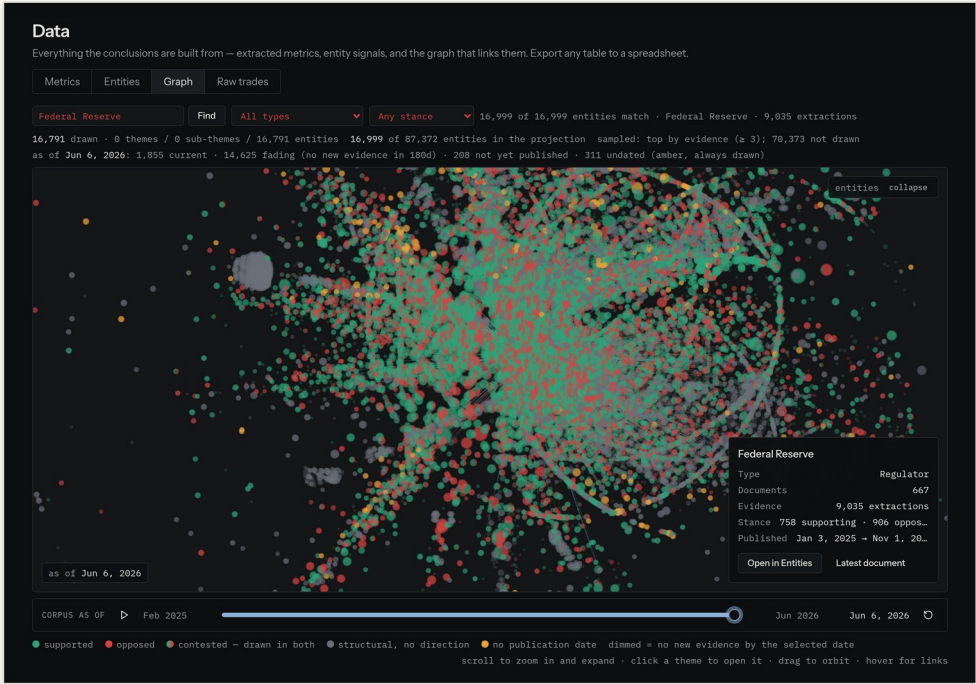
18 mo

WINDOW

Corpus: one institution's inbound sell-side research, 20 publishing institutions, anonymised. Episodes selected with outcomes known.

THE CORPUS, RESOLVED · BEFORE ANY EPISODE, THIS IS THE STANDING OUTPUT

One dated, signed map of everything the street had said.



The evidence graph, entity level, as of Jun 6 2026. Every node is a topic the street wrote about, signed by the balance of evidence and dated off publication timestamps. One click on any entity shows its documents, extractions and support/oppose split.

Counts as of the Jun 2026 capture. Desk names anonymised in this document; entities shown are public topics, not sources.

Athena AIC

156

STANDING VIEWS

82

CONSENSUS

19

LIVE DISAGREEMENTS

7

ASSET CLASSES

Every view is a group of individual trade calls from named desks — resolved bottom-up, never keyword-matched — with its split, its vintage and its weekly diff. **Rebuilt every time a report lands.**

WHY THE NUMBERS CAN BE TRUSTED · THE PLATFORM DELETES MORE THAN IT SHOWS

Every number traces to a source page — and every deletion is on screen.

A desk restating an already-published print is not forecasting. A UK call is never scored against a US index. Implausible extractions are pruned *and listed*. Consensus carries a vintage. Every claim resolves to a firm, a date and a page in the source document.

MEASURED ON THIS CORPUS

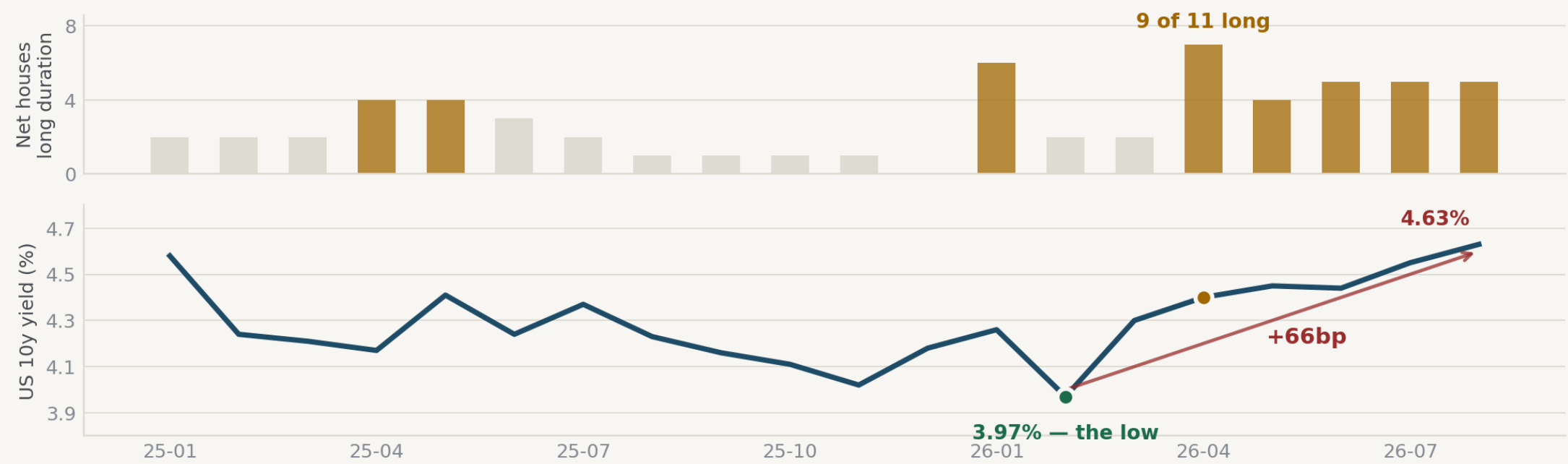
GUARD	BEFORE	AFTER
Points scored as genuine forecasts	886	129
Non-US call scored against a US index	scored	blocked
Chart-derived points on one metric panel	89	87 + 2 named

85% of what looked like forecasting was restatement of prints already published. One guard, turned on once. Deletion is the difference between a number and a fact.

Every pruned point is printed on screen with its source — nothing is silently dropped.

WHAT THE PLATFORM SHOWED • US DURATION • 2,735 CALLS, ONE STANCE PER BANK PER MONTH

The street added to a losing position for six months. Two banks did not.

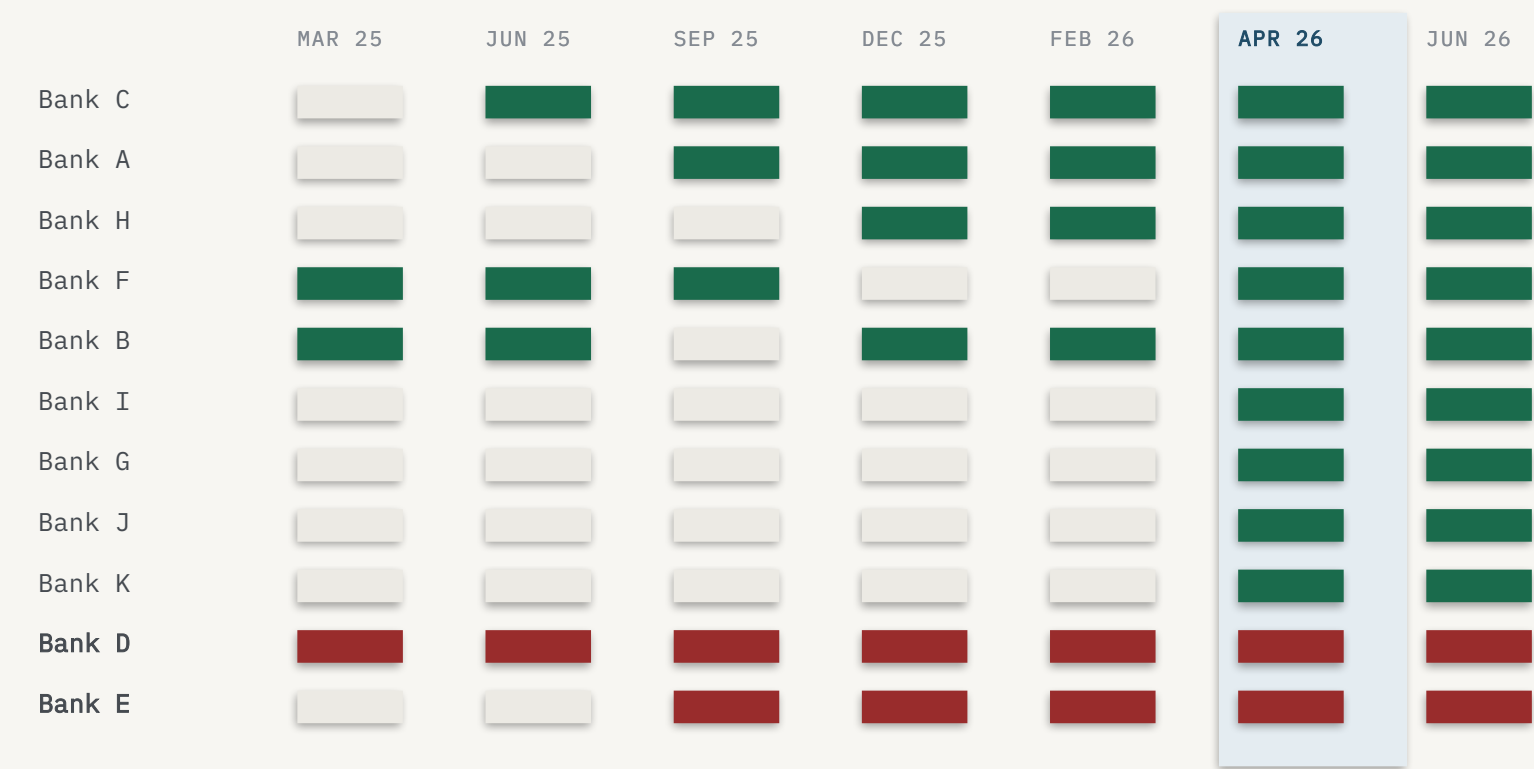


February 2026: four banks long duration, two short — Bank D and Bank E. By April the long side was **nine of eleven**, and yields had already risen 43bp against it. They rose another 23bp. Bank D’s dissent was on the record from January, dated and cited in the platform.

+80bp US 2-year, from the February low

THE SAME EPISODE AS THE PLATFORM STORES IT · ONE STANCE PER BANK PER MONTH

The flip, bank by bank. Not a headcount — a dated record of who moved.



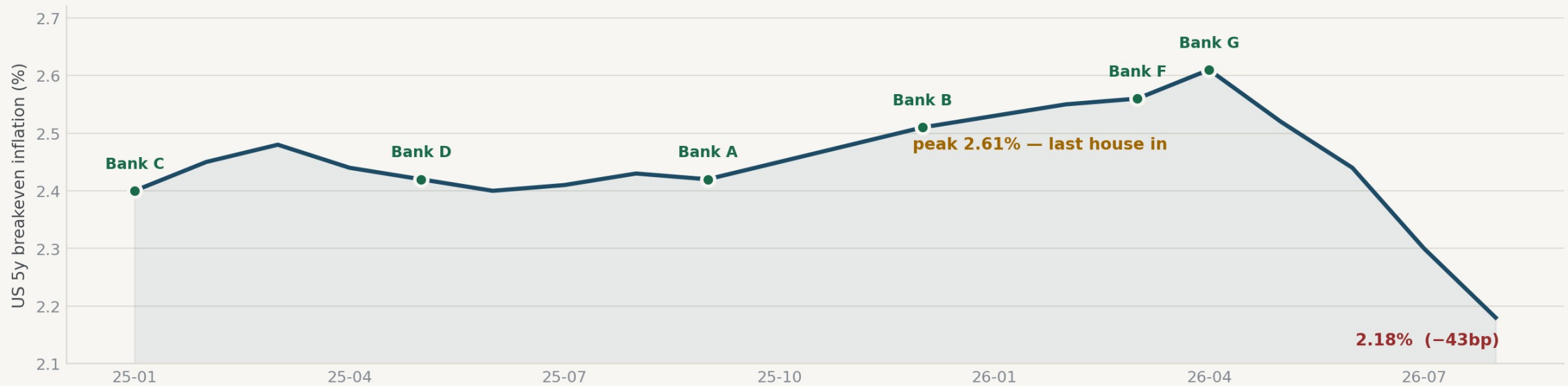
Four to nine, in two months. The Apr 26 column is the crowding — after yields had already turned.

Bank D was short the entire window. Bank E joined it in autumn 2025. The other nine arrived on the long side in waves — five of them in the final eight weeks.

Green long duration · red short · grey no standing stance. One stance per bank per month, condensed to seven dates. 2,735 underlying calls.

WHAT THE PLATFORM SHOWED • US INFLATION BREAKEVENS • 901 POSITIONS

Fifteen months to assemble six banks —
and it topped as the last one arrived.



Bank C opened the view in January 2025. Bank D joined in May, Bank A in September, Bank B in December, Bank F in March, **Bank G last, in late April 2026**. The 5-year breakeven peaked within weeks of that final entry.

Two arguments, not six. The platform deduplicates the reasoning: six banks, two distinct arguments. Knowing which you disagree with is a decision.

WHAT THE PLATFORM DID WITH THE REASONING • DEDUPLICATED, NOT COUNTED

Six banks on the same side. Only two distinct arguments between them.

ARGUMENT ONE

Banks C • A • B • F

The inflation risk premium is too low relative to the nominal term premium — breakevens under-price the distribution of inflation outcomes.

ARGUMENT TWO

Banks D • G

Pressure on central-bank independence biases the term premium higher — and breakevens with it.

Knowing which argument you disagree with is a decision. A headcount of six reads as conviction; two deduplicated arguments read as what they are — a premise you can test.

Arguments paraphrased from the underlying research; no verbatim quotation in this document.

A TABLE NOBODY ELSE HAS • THE BANKS' OWN PUBLISHED CLOSED TRADES

297 closed positions carry the bank's own reported result. The platform read them all.

BANKS' OWN PUBLISHED CLOSED TRADES	TRADES	WIN RATE
Bank A +459bp total	227	61%
Bank B +166bp	29	86%
Bank C +96bp	22	77%

Why this matters. The episodes are what the platform surfaces occasionally. This is what it holds permanently: every bank's own record, on one table.

Bank A's 227-trade ledger wins 61% of the time — but its average win (+8.2bp) is the same size as its average loss (−8.0bp). **Its entire edge is hit rate, not payoff asymmetry.** Assembled by the platform from eighteen months of weekly publications.

Reporting is voluntary and uneven — two banks look complete; one bank's eleven rows are almost certainly an adverse partial sample and are excluded from any claim.

Median holding period across the closed book: 49 days. Results as published by each bank in its own trade-tracking tables; aggregated, not audited.

THE ARITHMETIC, SHOWN SO YOU CAN CHECK IT

One episode, on one position, pays for years of the platform.

\$80,000

PER BP • \$100M 10Y POSITION

66bp

THE MOVE FROM THE FEB 2026 LOW

\$5.3m

ON THAT POSITION

2

EPISODES THIS SIZE IN 18 MONTHS

You would not have caught all of it. You would have seen, in April, that your trade had become **the only trade on the street** while the market had already turned — with the dissenting argument in front of you rather than months later on a call.

Two episodes is the count of large, fully audited cases in one asset class — not the count of useful signals. The everyday output is the Monday diff: who moved, who dissented, what your positions became.

A \$100m 10-year position carries roughly \$80,000 of P&L per basis point at a modified duration near 8. Move measured on public FRED series, Feb–Aug 2026.

THE CLAIM, STATED PRECISELY

The information was legible in real time. No one reading by hand assembles it.

WHAT THE PLATFORM PRODUCED

A dated, sourced map of the street's views — who held what, when they took it, who dissented and why — plus a cross-bank P&L table no one else compiles. All of it citable to its source page at read time.

WHAT WE DO NOT CLAIM

No prediction, and no view into anyone's actual positions or committed capital. Episodes were selected with outcomes known — the claim is legibility, not foresight. Published research is the only input.

See it on your own research flow. A pilot starts with your book: you name the positions you hold and the desks you read, and coverage is reported honestly on day one.