



TRADING LEADERSHIP

Which models make market judgments you can defend?

A prospective test of what matters, what is exposed, and what happens next.

To learn more, reach us at founders@whalebench.com

Background

What we do

Prospective evaluation
for market judgment.

Whalebench freezes public information at a known time, asks every model the same four questions, and seals each answer before the market can reveal it.

Who it is for

For leaders deciding where
models deserve trust.

For CIOs, trading leaders, AI platform owners, and model-risk teams that need inspectable decisions, not another benchmark score.

FRONTIER MODELS EVALUATED



Built by



Mark

Product + markets



Krishna

Research + operations



Ansh

Engineering + systems

Why do you need Market-Native Evaluation?

Most benchmarks ask whether a model knows the answer. A market test asks whether it can form a useful view before the answer exists.

----->

/ 01

Scores hide the cutoff,
so future knowledge is hard to rule out.

/ 02

Summaries avoid a view,
so nothing falsifiable is recorded.

/ 03

Results collapse the trail,
so luck can look like judgment.

The useful record is simple: what the model saw, what it claimed, how certain it was, and what reality later said.

Whalebench / Decision replay

ASK THE RECORD

A

Why did Claude flag this as material?

RECORDED ANSWER

The filing changed forward free-cash-flow guidance, mapped directly to the issuer, and was not present in the prior public record.

EXPOSURE

HORIZON

CONFIDENCE

ISSUER

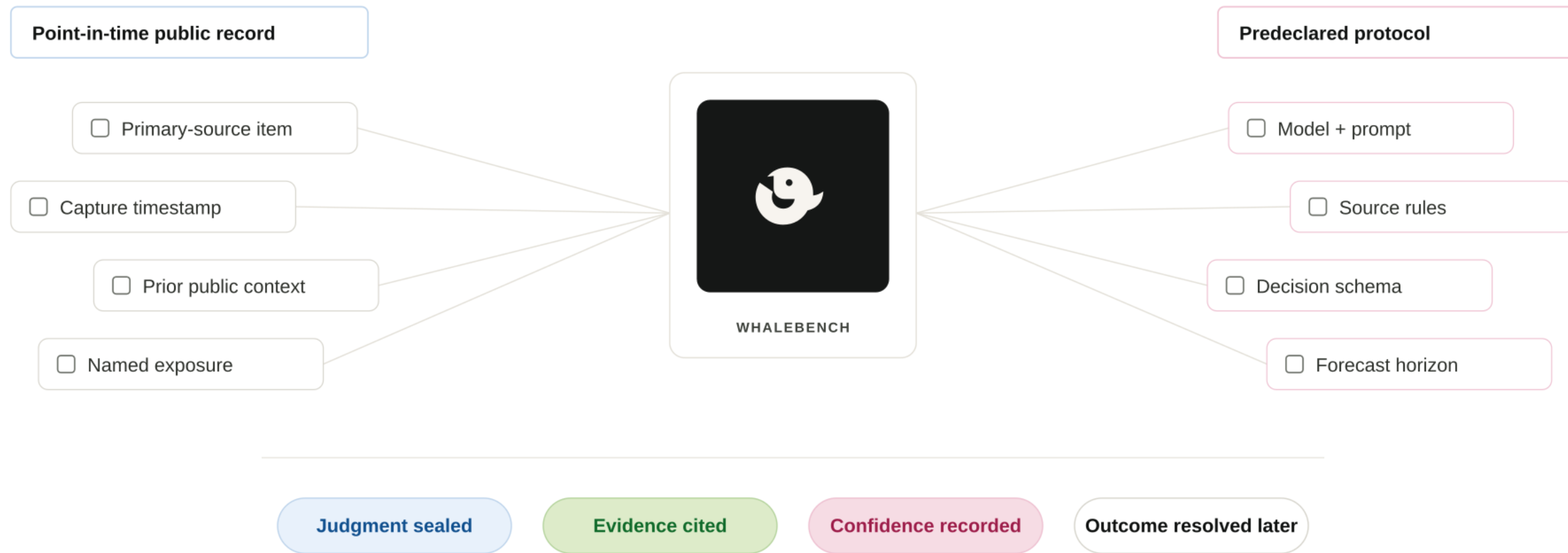
5D

72%

SOURCE TIME · MODEL · PROMPT · SEALED VIEW · OUTCOME

Our Solution

Whalebench turns one frozen public item into four sealed judgments: relevance, novelty, exposure, and a falsifiable market view.



Meet Whalebench, The Record Before The Result

Every claim is tied to the public evidence
available when the model made it.

Watch study →

 Whalebench

STUDY PORTAL

Briefings
Capture
Judgments
Replays
Outcomes

ASK THE RECORD

Show every high-confidence view sealed before its outcome.

ILLUSTRATIVE SEALED JUDGMENTS

 Claude Opus Material · exposure named	POSITIVE	5D	72%
 Mistral Large Material · exposure named	NEGATIVE	3D	65%
 GPT-5.6 Evidence threshold not met	ABSTAIN	—	58%

EVERY JUDGMENT SEALED BEFORE OUTCOME

Source · evidence · exposure · view · confidence

Whalebench's Use Cases

One prospective record supports model selection, governance, and a more disciplined research loop.

Compare judgment

1. *Freeze one public item*
2. *Ask the same four questions*
3. *Score only after the horizon*

See which models identify what matters before the market makes it obvious.

Govern the process

1. *Lock model and prompt*
2. *Preserve every citation*
3. *Separate reasoning from outcome*

Give the desk and control functions one reviewable record.

Test a workflow

1. *Start with one decision*
2. *Run beside the desk*
3. *Publish misses with wins*

Earn trust with prospective evidence before a model enters production.

Prospective Judgment: Whalebench Record vs. Static Evals

 whalebench ✓

Static eval ✗

JUDGMENT / 2026-08-13 / 14:00 UTC

A Claude Opus

POSITIVE · 72%

FROZEN PUBLIC EVIDENCE

The filing raises FCF guidance. The model names the issuer exposure and commits to a five-day positive view.

MODEL	claude-opus-5
HORIZON	5 trading days
PROMPT	sha256: 12f...a9c
OUTCOME	resolve after horizon

Static benchmark

QUESTION

Which answer best describes the filing?

- ☐ A. Revenue increased
- ☒ B. Guidance was raised
- ☐ C. The market is efficient
- ☐ D. None of the above

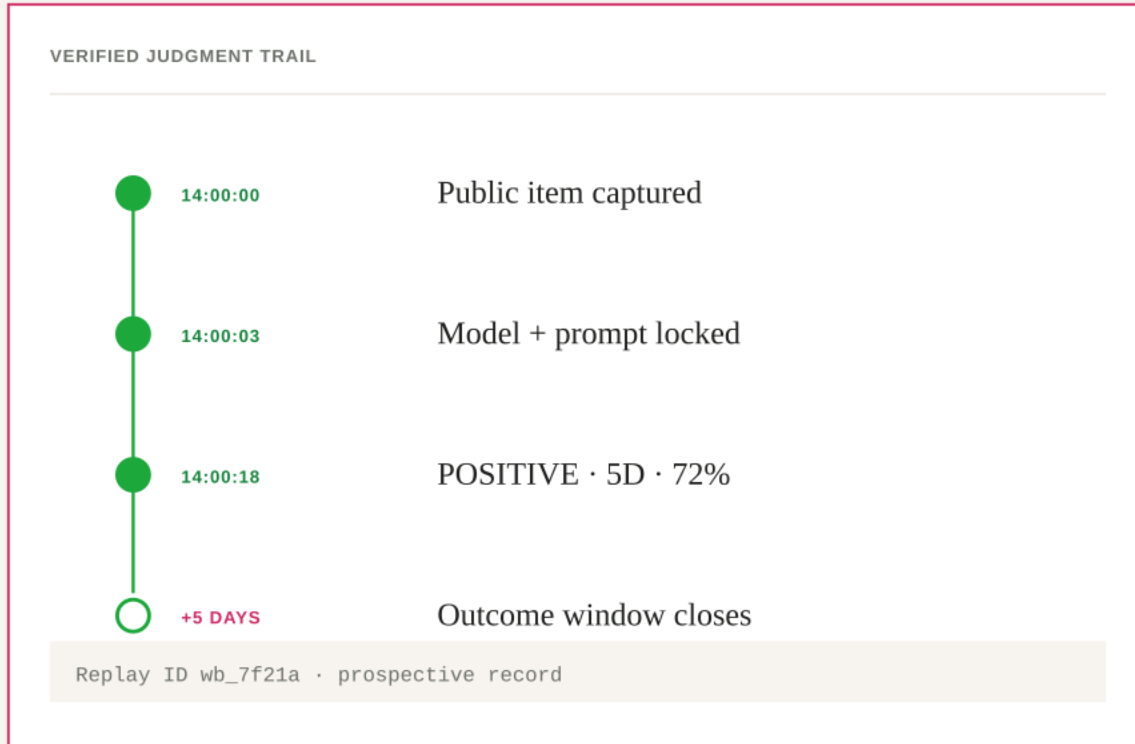
RESULT

Correct

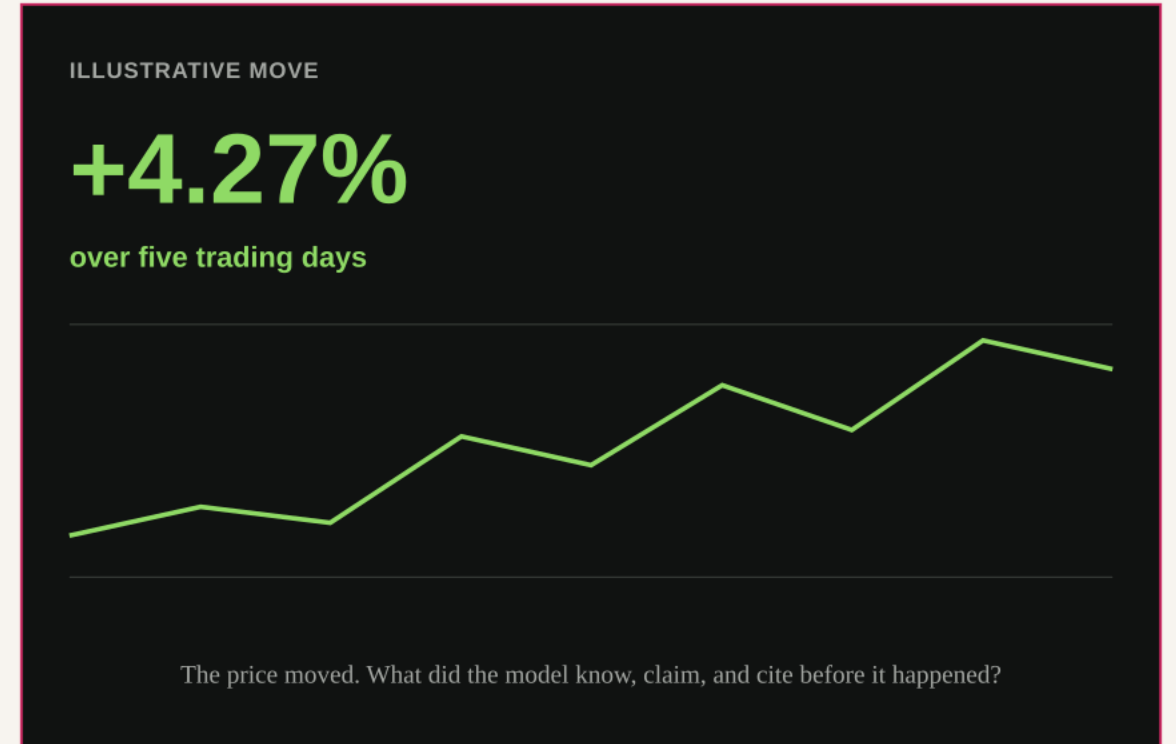
Whalebench seals a falsifiable view while the answer is still unknown.

A static eval can grade recall without testing timing, calibration, or consequence.

Auditability: Sealed Judgment vs. Outcome Screenshot

Outcome screenshot ✗



The source, configuration, evidence, view, confidence, and later outcome stay attached.

A result without its prior decision trail cannot distinguish judgment from luck or selection.

「Thank you.」

TRADING LEADERSHIP

Mark · Krishna · Ansh

founders@whalebench.com

