

whalebench

RL GYM

Train on a record that cannot leak the answer.

Frozen evidence, structured actions, and delayed outcomes for market-agent research.

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

Background

What we do

Prospective market records
become clean episodes.

Whalebench captures what was knowable at a stated time, seals the action, and attaches the outcome only after the horizon closes.

Who it is for

For teams building agents
that must generalize.

For RL research, applied AI, quant research, and platform teams that need leakage-safe evaluation data across changing regimes.

FRONTIER MODELS EVALUATED



Built by



Mark

Product + markets



Krishna

Research + operations



Ansh

Engineering + systems

Why do you need a Market-Native RL Gym?

An agent cannot learn a robust policy from episodes that quietly contain hindsight or change underneath the comparison. --->

/ 01

Cutoffs stay implicit,
so hindsight enters the state.

/ 02

Actions stay unstructured,
so rewards lose meaning.

/ 03

Averages hide the slice,
so regressions survive.

A defensible episode binds one state, one policy version, one action, and one delayed outcome.

Whalebench / Decision replay

ASK THE RECORD

Why did policy v17 abstain?

RECORDED ANSWER

Two required primary sources were absent at the cutoff, so the policy followed its declared evidence threshold and did not act.

ACTION

EPISODE

REWARD

ABSTAIN

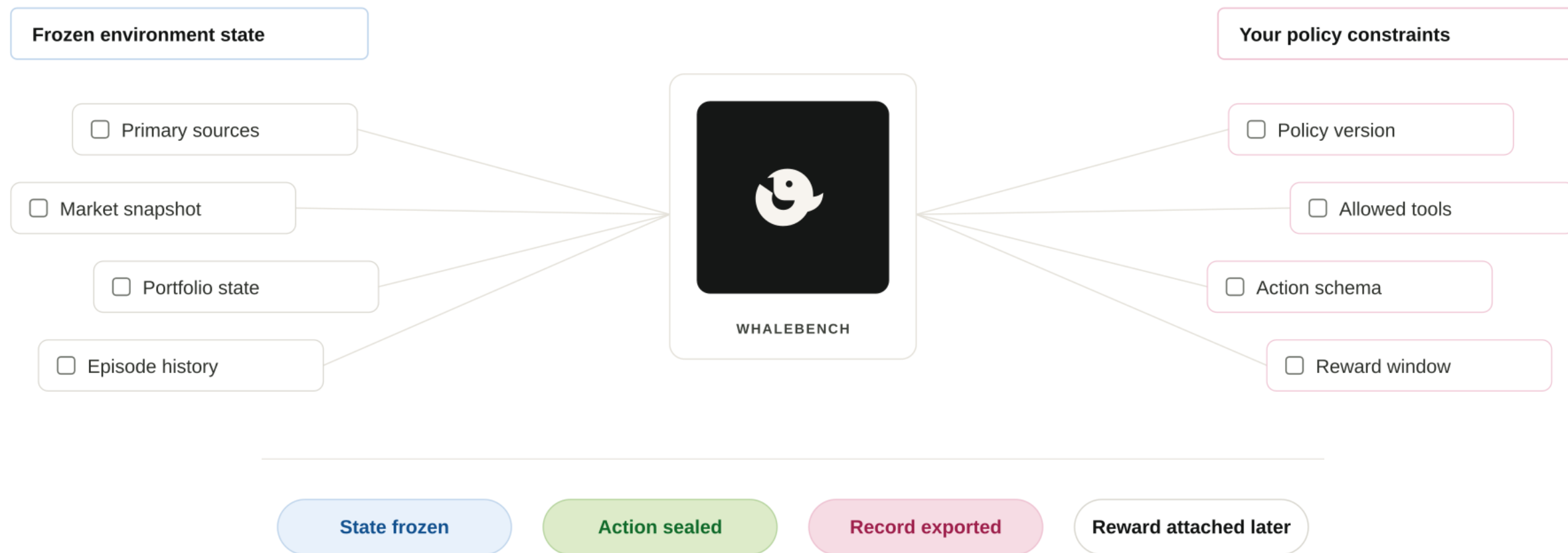
E-441

PENDING

STATE HASH · POLICY · TOOLS · ACTION · REWARD VERSION

Our Solution

Whalebench supplies the prospective record: a frozen observation, a constrained action, and an outcome withheld until it becomes knowable.



Meet Whalebench, Your Prospective Episode Layer

Whalebench creates leakage-safe records.
Your team controls the policy and
training loop.

Inspect... →

 **Whalebench**

Environments

Episodes

Policies

Replays

Rewards

INSPECT THE POLICY

Show every action that changed after policy v17. 

ILLUSTRATIVE HELD-OUT EPISODES

 Policy v17 Frozen state · tools locked	LONG	5D	R +0.42
 Policy v16 Same state · prior version	SHORT	5D	R -0.31
 Baseline Evidence threshold failed	ABSTAIN	—	R 0.00

EVERY EPISODE STARTS BEFORE THE ANSWER

Cutoff · state · policy · action · delayed outcome

Whalebench's RL Use Cases

The same prospective record supports dataset construction, policy evaluation, and failure analysis.

Build clean episodes

1. *Freeze the information cutoff*
2. *Bind one action schema*
3. *Attach outcomes later*

Keep future information outside the observation by construction.

Evaluate policies

1. *Hold states constant*
2. *Compare policy versions*
3. *Stress new regimes*

Measure the policy change without changing the evidence underneath it.

Inspect failures

1. *Find the weakest slice*
2. *Trace action to evidence*
3. *Separate process from luck*

Turn an aggregate reward into a concrete, reviewable miss.

Prospective Episodes: Whalebench Record vs. Static Answers

 whalebench ✓

Static answer set ✗

EPISODE / E-441 / STATE FROZEN

Policy v17

ABSTAIN · VALID

OBSERVATION STATE

Two required primary sources are absent. Policy follows the evidence threshold and does not act.

POLICY	market-policy-v17
STATE	sha256: 7b1...d02
TOOLS	search off · filings only
REWARD	version r3 · pending

Static answer set

QUESTION

Should the agent trade this event?

- ☐ A. Long
- ☐ B. Short
- ☒ C. Abstain
- ☐ D. Ask for more data

RESULT

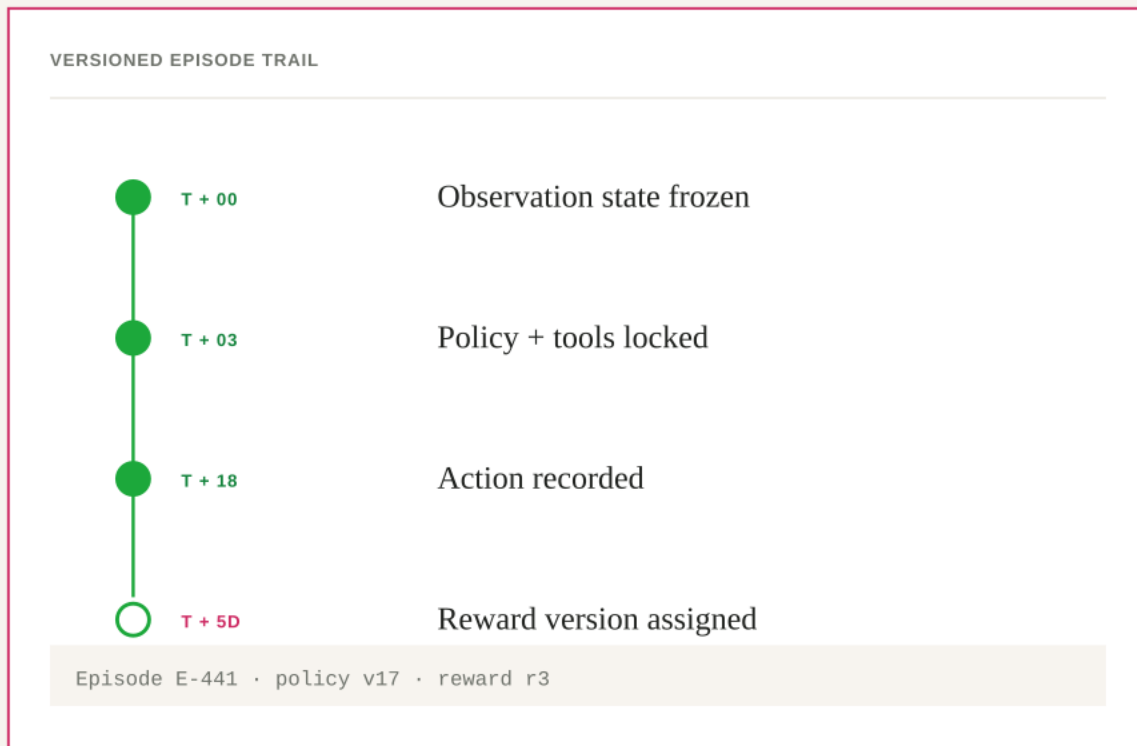
Correct

Whalebench binds the action to its exact cutoff, state, constraints, and delayed outcome.

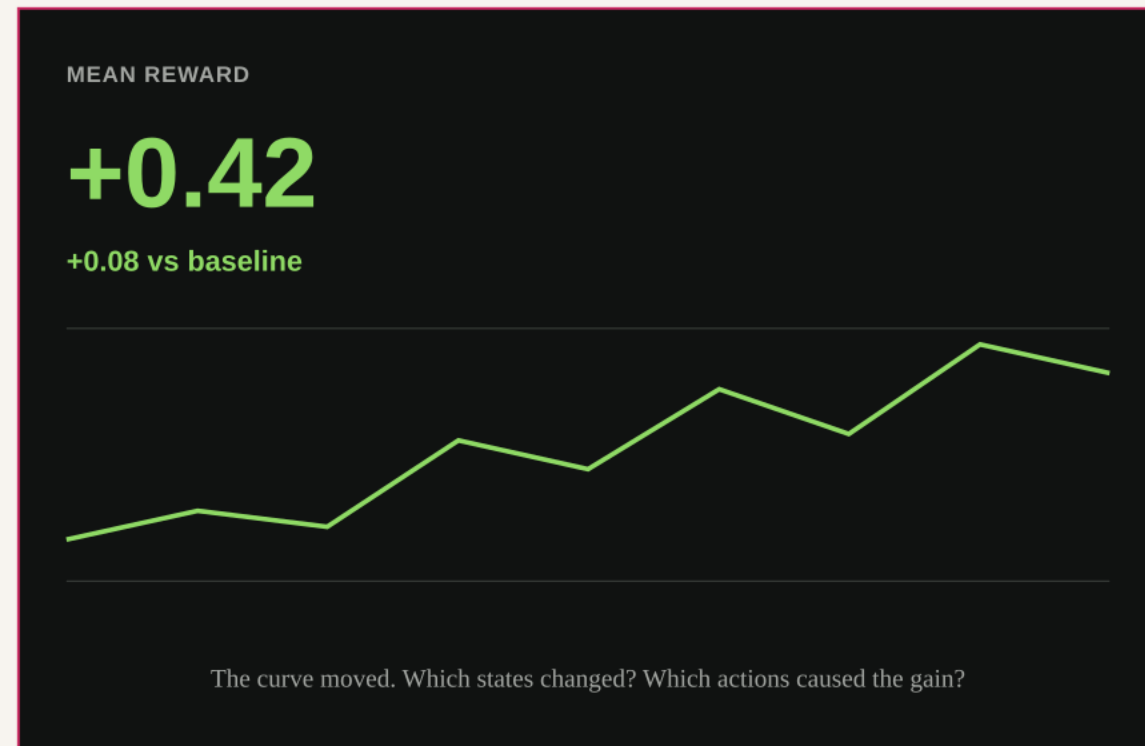
A static answer can be correct without proving that the policy respected the observation boundary.

Diagnosis: Prospective Record vs. Aggregate Reward Charts

whalebench ✓



Reward chart ✗



Every policy change can be compared on the same frozen evidence and declared action boundary.

An aggregate reward can rise while hiding the exact regimes and decisions that regressed.

「Thank you.」

RL GYM

Mark · Krishna · Ansh

founders@whalebench.com

 **whalebench**