

GUEST LECTURE · UC BERKELEY · OCTOBER 15, 2026

From Goldfish to Self-Improving Agents

Continual learning outside the weights

Chen-Yu Lee

Google Cloud AI Research



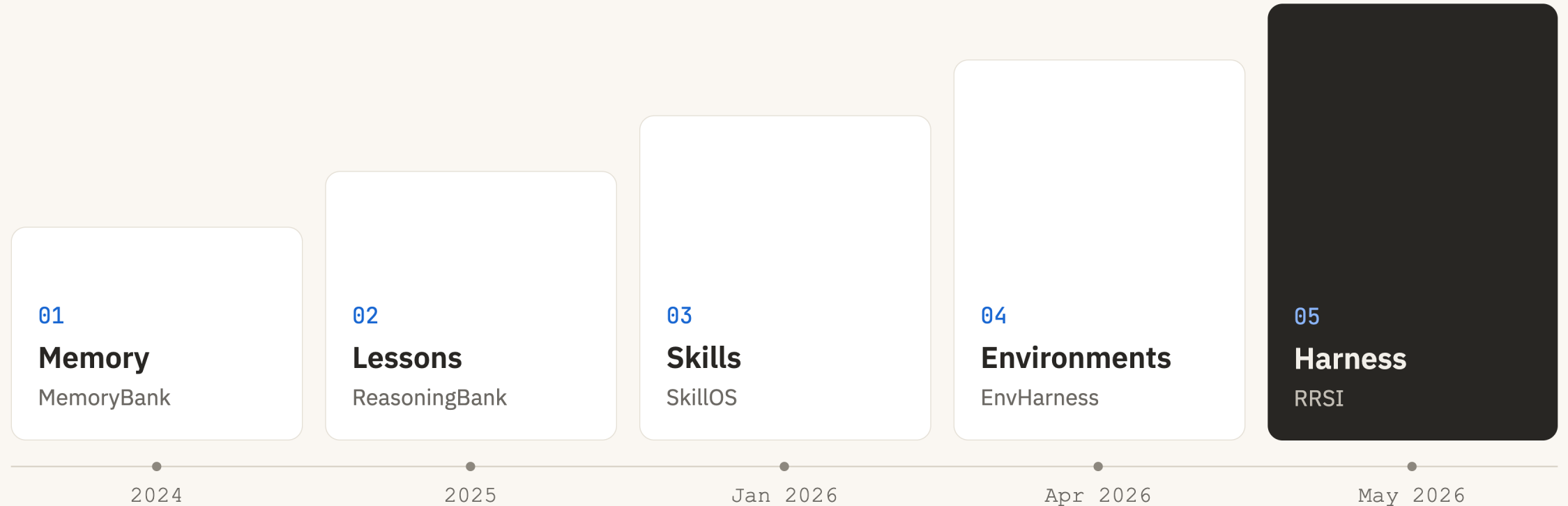
chenyulee.com

Neural nets forget. Deployed agents never learn.

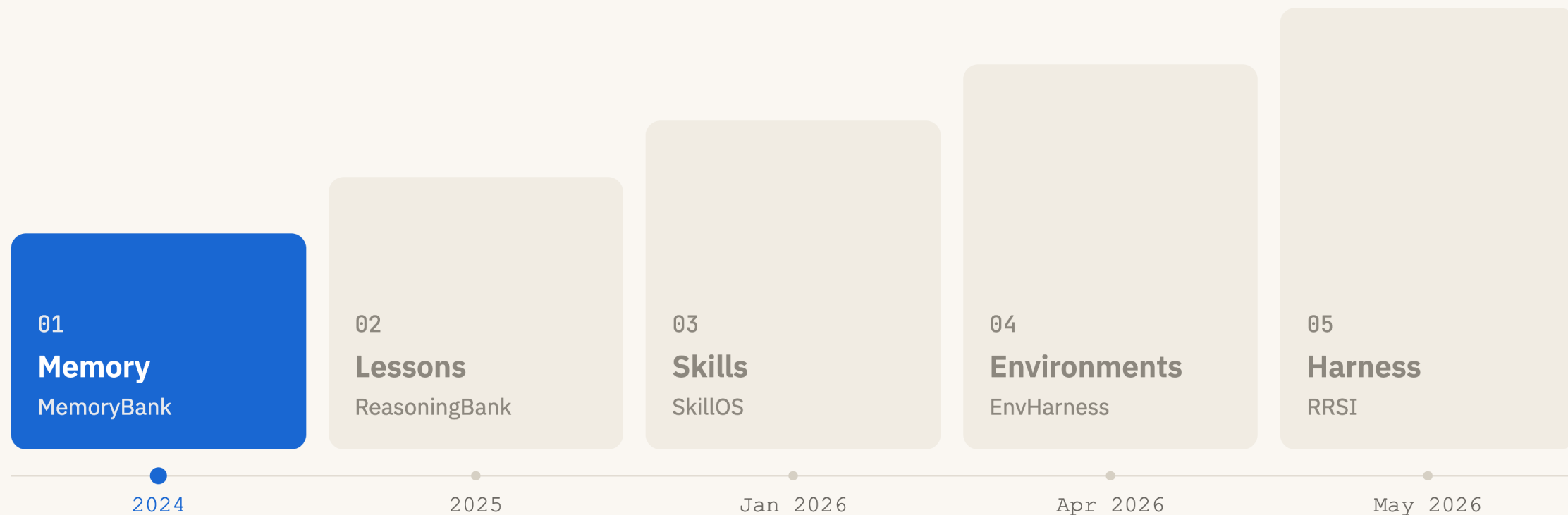
	Classic continual learning	Deployed agents
LEARNING LIVES	In the weights	Nowhere: weights are frozen
THE RISK	New tasks overwrite old ones	Nothing is learned at all
THE QUESTION	How do we keep what we learned?	Where else can learning live?

THE MAP FOR TODAY

Five places learning can live



Five places learning can live



Can an agent remember across sessions at all?

Ever talked to a forgetful agent?

It keeps asking **the same questions**

It doesn't remember **who you are**

Every session **starts from scratch**

I lost my credit card.

Card Services
I can help. What's your account number?

It's 987654321.

Card Services
Thanks. Transferring you to Lost & Found.

— transferred to Lost & Found —

Lost & Found
Hi! How can I help you today?

Sometimes it's more than frustrating

Session 1
(3 days ago)

Patient: I have a persistent cough and a slight fever.

Healthcare agent: Okay. Please tell me about the severity of each symptom.

Session 2
(today)

Patient: I have a headache and the fever is gone.

✓ With memory

Thanks. So, to update: **the fever is gone, a headache started, and the cough persists.** Given that progression, let's consider...

✗ Without memory

Okay. Given the headache, let's consider...
The cough, three days old, is silently dropped.

Why not put every past conversation in context?

Higher **cost**

Every turn re-reads the whole history, and the history only grows.

Worse **performance**

The few facts that matter get buried in everything that doesn't.

Accuracy (%)

No memory



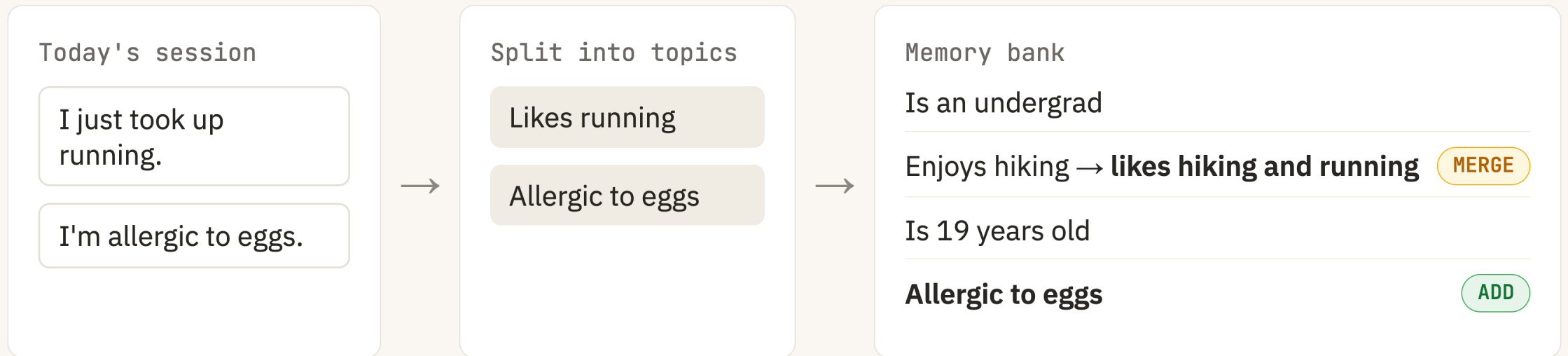
Full context



MemoryBank (ours, no re-ranker)



After each session, organize memory by topic



Memory stays organized by topic, not as raw chat logs. Similar facts are **merged**, new ones are **added**.

Better memory, not more context

Method	MSC		LongMemEval
	METEOR	BERTScore	Accuracy
No history	5.2	10.6	0.0
Long context	14.8	31.9	57.4
MemoryBank (ours)	33.4	57.1	70.4

+13 points over long context on LongMemEval, and **about 2×** the reply quality on MSC.

MemoryBank with GTE retriever; averaged over 3 runs, in %. MSC gain: 2.3× METEOR, 1.8× BERTScore.

From paper to product

Mar 2025

Paper on reflective memory, accepted at ACL 2025

Jul 2025

Launched on Google Cloud as Agent Engine Memory Bank

Today

Core memory layer of Gemini Enterprise Agent Platform

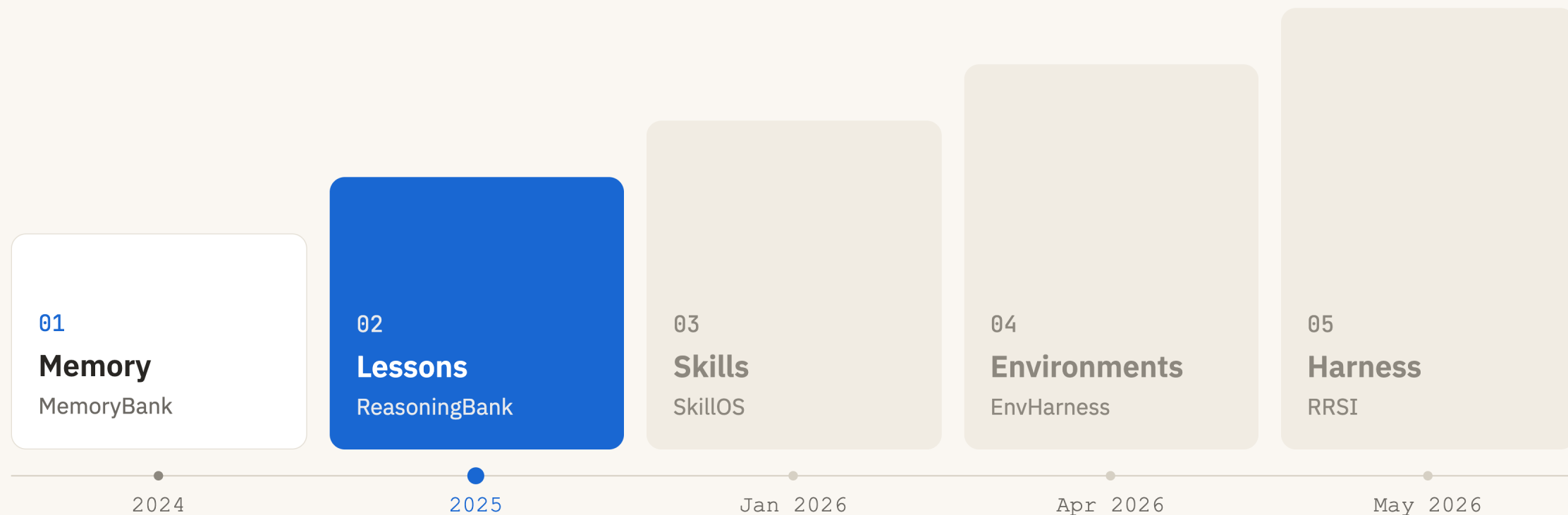
WHAT BROKE

It remembered what happened.
It didn't get any better.

Recalling facts is not the same as learning.

NEXT • ACT 2 • LEARN FROM EXPERIENCE

Five places learning can live



It recalled facts but didn't improve. What if it stored lessons instead?

Store lessons, not transcripts

SAME TASK "When did I make my first purchase on this site?"

Trajectory memory

Every step it took

```
click('188')  
think: on 'My Account'...  
click('1530')  
think: on 'My Orders'...  
click('1614') ...
```

× Long and page-specific: hard to reuse

Workflow memory

The routine that worked

Workflow: get info from a page section

1. click(section)
2. extract info
3. send to user

× Successes only: every failure is thrown away

ReasoningBank

The lesson learned

Title: Navigation strategy

When searching history, detect pagination and check every page, not just the most recent items.

✓ From successes and failures, reusable across tasks

raw

→ abstract, reusable

A shopping agent, day after day

 failed attempt  success  attempt no longer needed

Without ReasoningBank

Every day starts from zero

DAY 1 First order date



7 tries

misses page 2 of orders

DAY 2 Cheapest shoe



gave up

cheapest shoe was on page 3

...

DAY N Orders by month



8 tries

same mistake, again

With ReasoningBank

Lessons carry over to the next task

DAY 1 First order date



7 tries

Lesson: check every page, not just the recent list

DAY 2 Cheapest shoe



5 tries

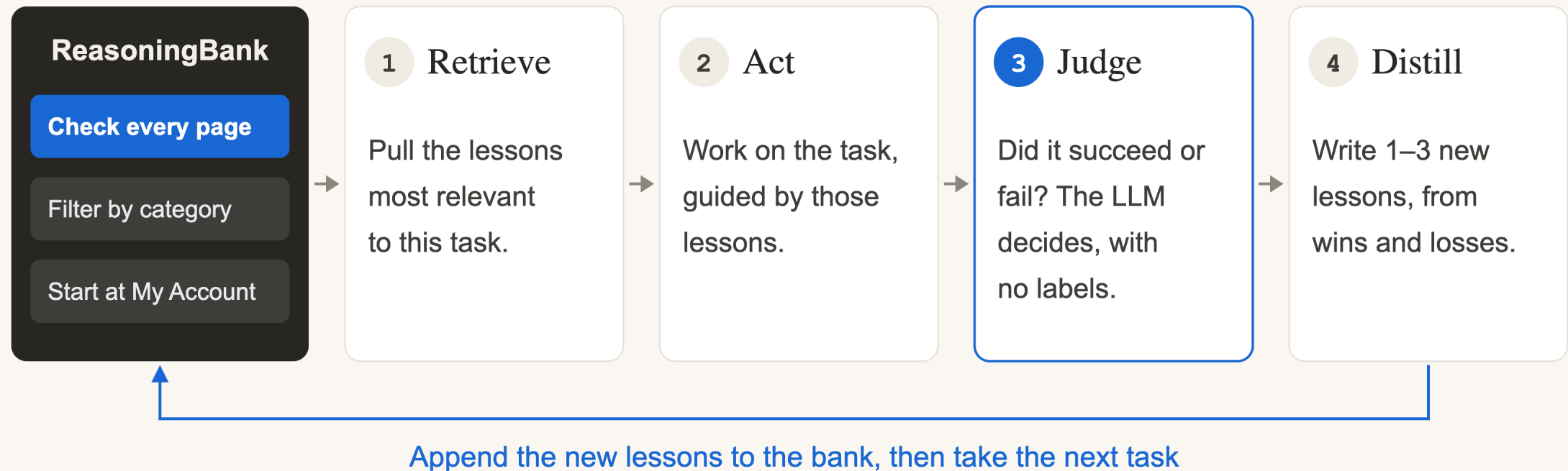
...

DAY N Orders by month



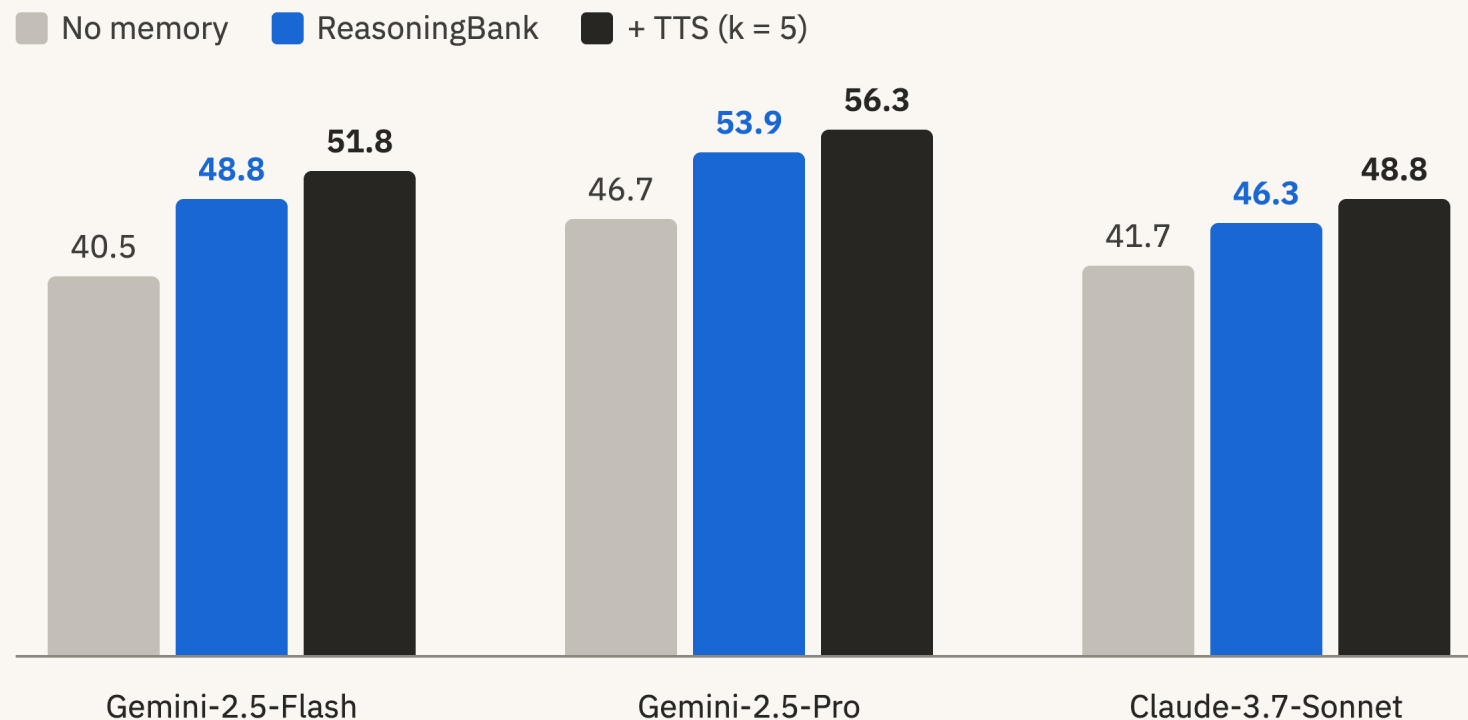
1 try

A closed loop: retrieve, act, judge, distill



REASONINGBANK

Better on every backbone, and with compute



Also

54.0 → 57.4 on SWE-bench
Verified (Gemini-2.5-Pro)

Up to 16% fewer steps

Community feedback

✓ Accepted at ICLR 2026

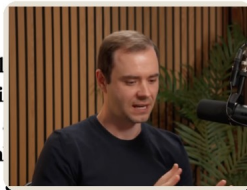


REASONINGBANK: Scaling Agent Self-Evolving with Reasoning Memory

Siru Ouyang^{1*}, Jun Yan^{2†}, I-Hung Hsu², Yanfei Chen², Ke Jiang², Zifeng Wang², Rujun Han², Long T. Le², Samira Daruki², Xiangru Tang³, Vishy Tirumalashetty², George Lee², Mahsan Rofouei⁴, Hangfei Lin⁴, Jiawei Han¹, Chen-Yu Lee^{2†} and Tomas Pfister²

¹University of Illinois Urbana-Champaign, ²Google Cloud AI Research, ³Yale University, ⁴Google Cloud AI

With the growing adoption of large language model agents in persistent real encounter continuous streams of tasks. A key limitation, however, is their accumulated interaction history, forcing them to discard valuable insights propose REASONINGBANK, a novel memory framework that distills generalization from an agent's self-judged successful and failed experiences. At test time, REASONINGBANK memorizes from REASONINGBANK to inform its interaction and then integrates new learnings back



Dwarkesh Podcast • Oct 2025

Andrej Karpathy mentioned ReasoningBank in his interview with Dwarkesh Patel.

Dwarkesh Podcast, Oct 2025, youtube.com/watch?v=36OBX5lQjGc (t=3:44)



New memory framework builds AI agents that can handle the real world's unpredictability

Ben Dickson

4:35 pm, PT, October 8, 2025



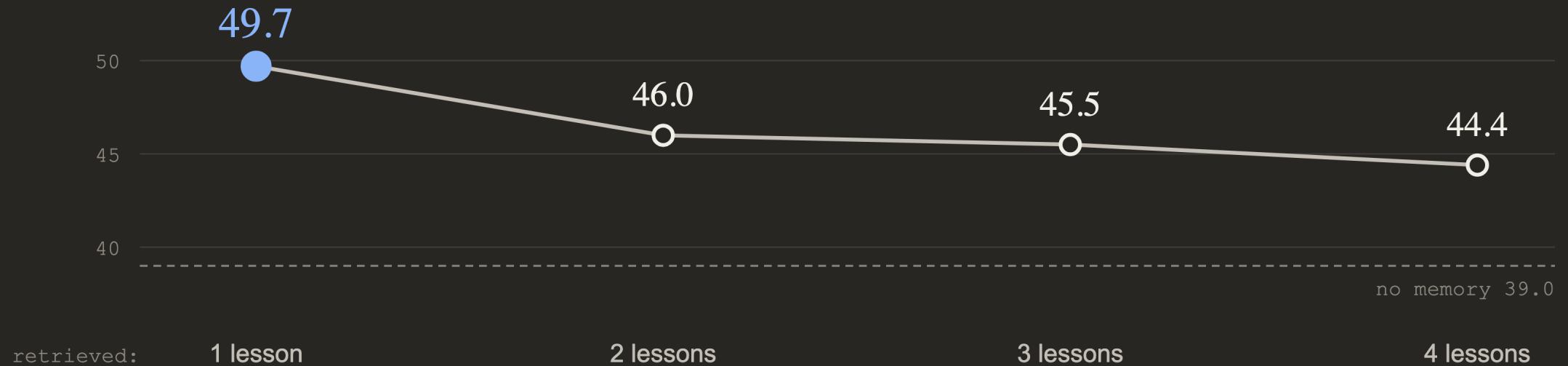
VentureBeat • Oct 2025

Covered ReasoningBank as a way for agents to learn from both their wins and their losses.

venturebeat.com, Ben Dickson, Oct 8, 2025

WHAT BROKE

More memory started to hurt.

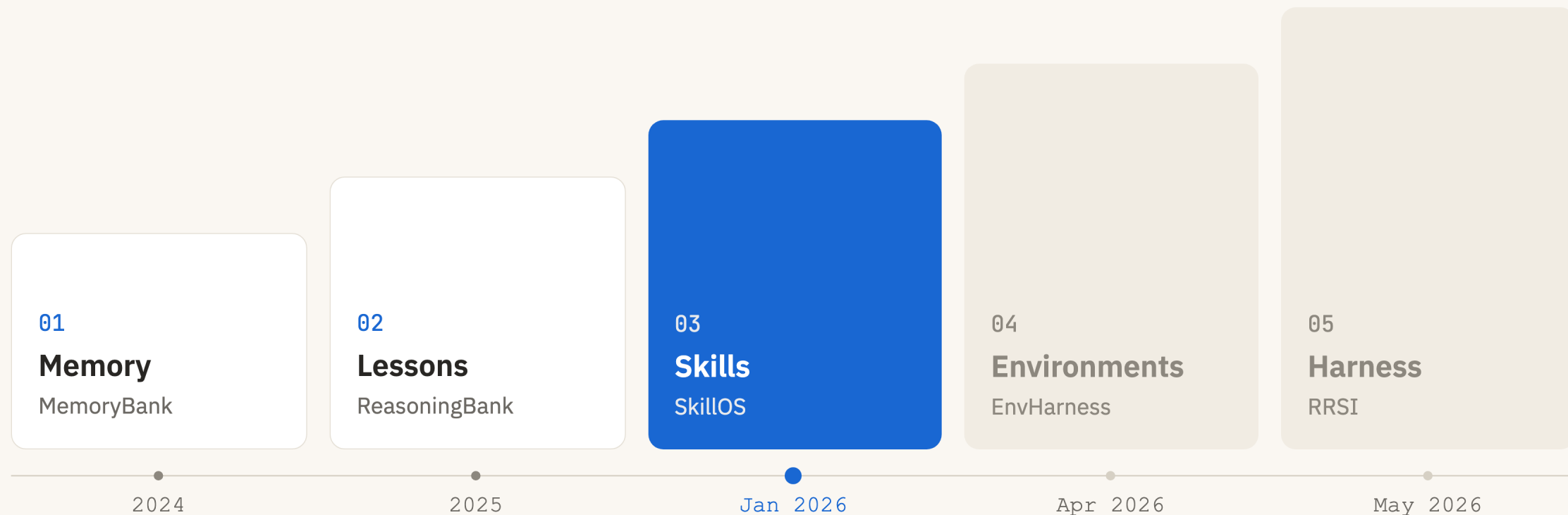


The bank only grows. Nothing decides what is worth keeping.

Success rate, WebArena-Shopping, Gemini-2.5-Flash

NEXT • ACT 3 • CURATE

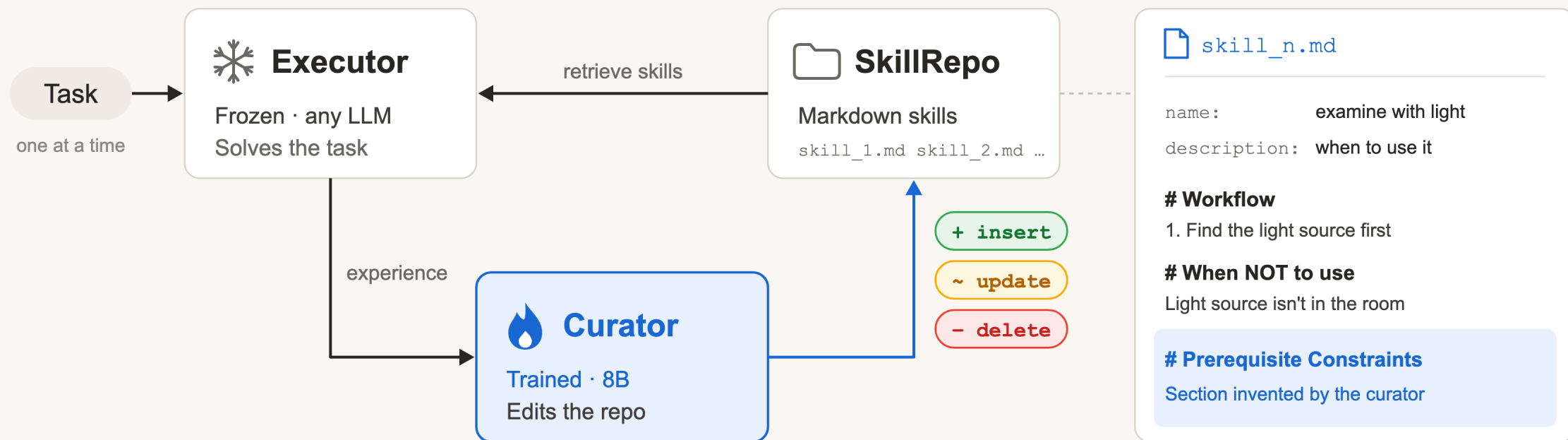
Five places learning can live



More lessons started to hurt. Can the agent learn what to keep?

Train a librarian, not a bigger shelf

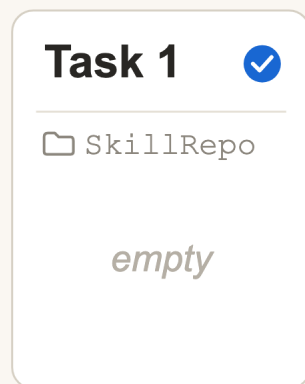
✓ Accepted at NeurIPS 2026



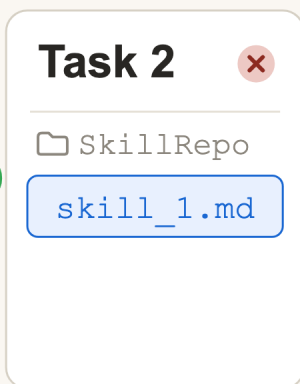
The trick: train on **streams of related tasks**, so a skill is only rewarded if it helps the tasks that come *later*.

How the curator learns

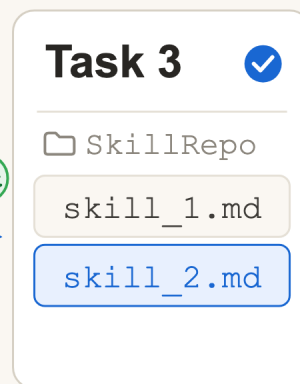
1 Sample ~10 tasks; repo starts empty.



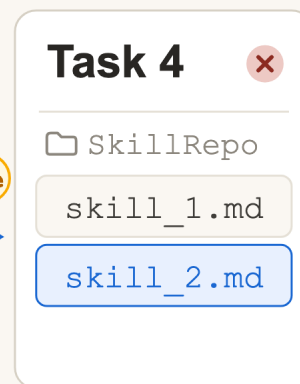
+ insert



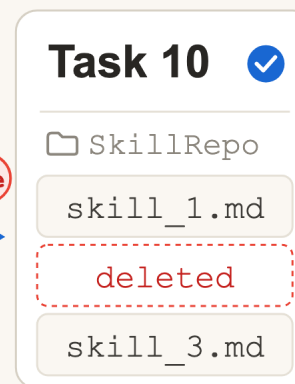
+ insert



~ update



- delete



3 Score the whole episode

Reward = **success rate on Tasks 2...10**

+ helpers: valid edits, skill quality, conciseness

4 Learn by comparison (GRPO)

Replay the same group 8 times.

Above-average curators get reinforced.

SKILLOS

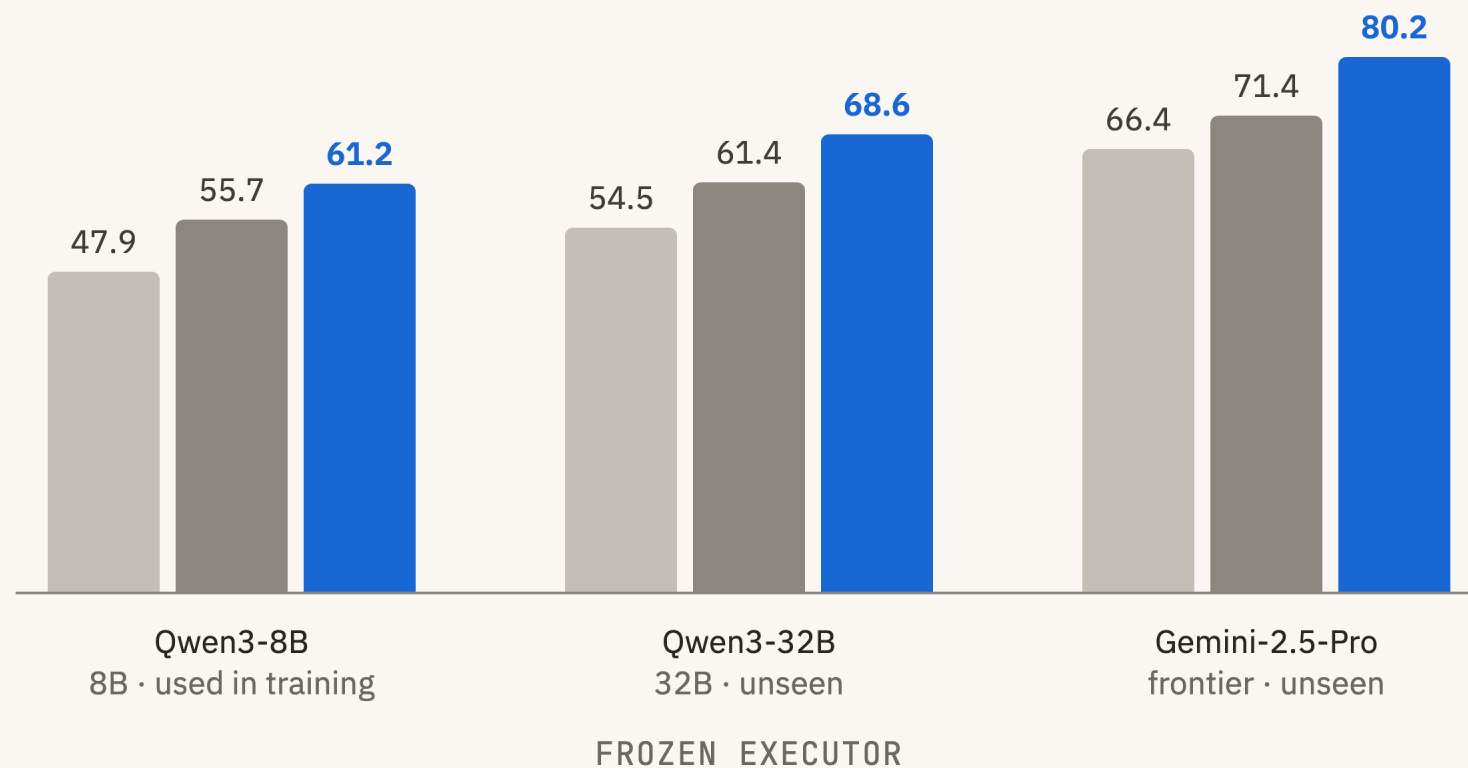
A small curator that transfers

Curator: **Qwen3-8B**, RL-trained once

No memory

ReasoningBank

SkillOS

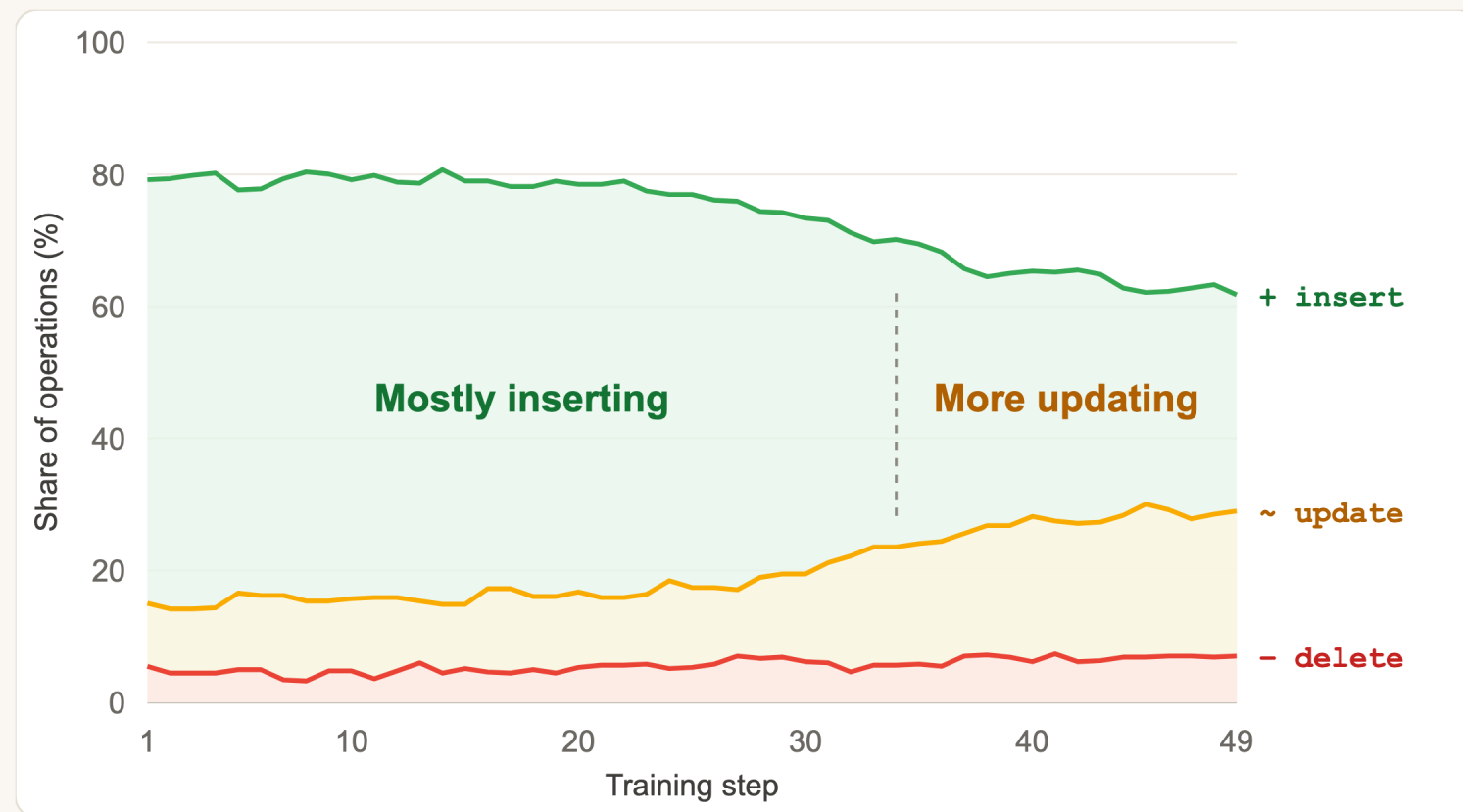


The **8B** RL-trained curator also beats **Gemini-2.5-Pro** asked to curate zero-shot.

Also holds on a newer family: with **Gemini-3.1-Flash-Lite**, **73.1** vs. 66.0 for ReasoningBank.

ALFWorld avg. success rate (%) by executor; curator trained once with Qwen3-8B.

What the curator learned on its own



From adding to editing

Insert dominates early; update roughly doubles as training goes on.

From tips to meta-skills

Task tips give way to verification, fallback planning and strategy change.

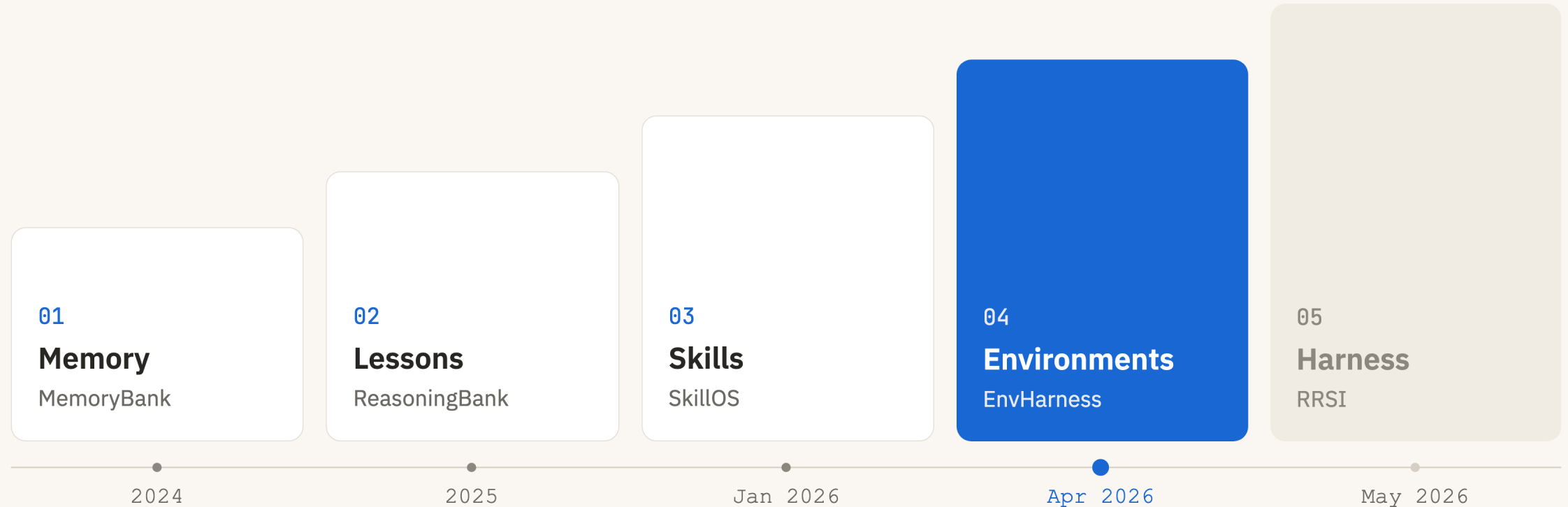
WHAT BROKE

An agent can only learn from the worlds we give it.

Most of those worlds are static.

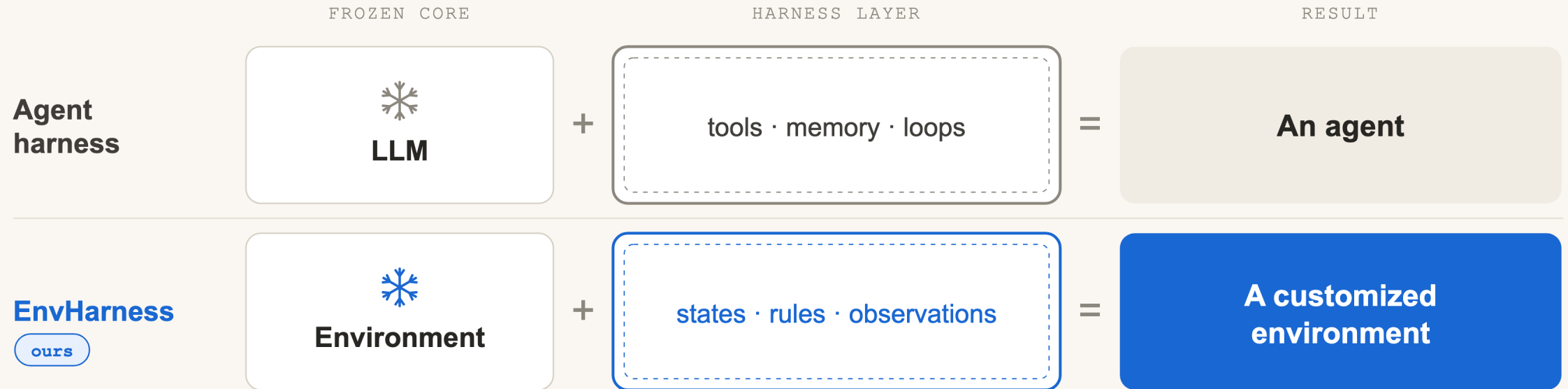
Better curation can't fix a poor curriculum.

Five places learning can live



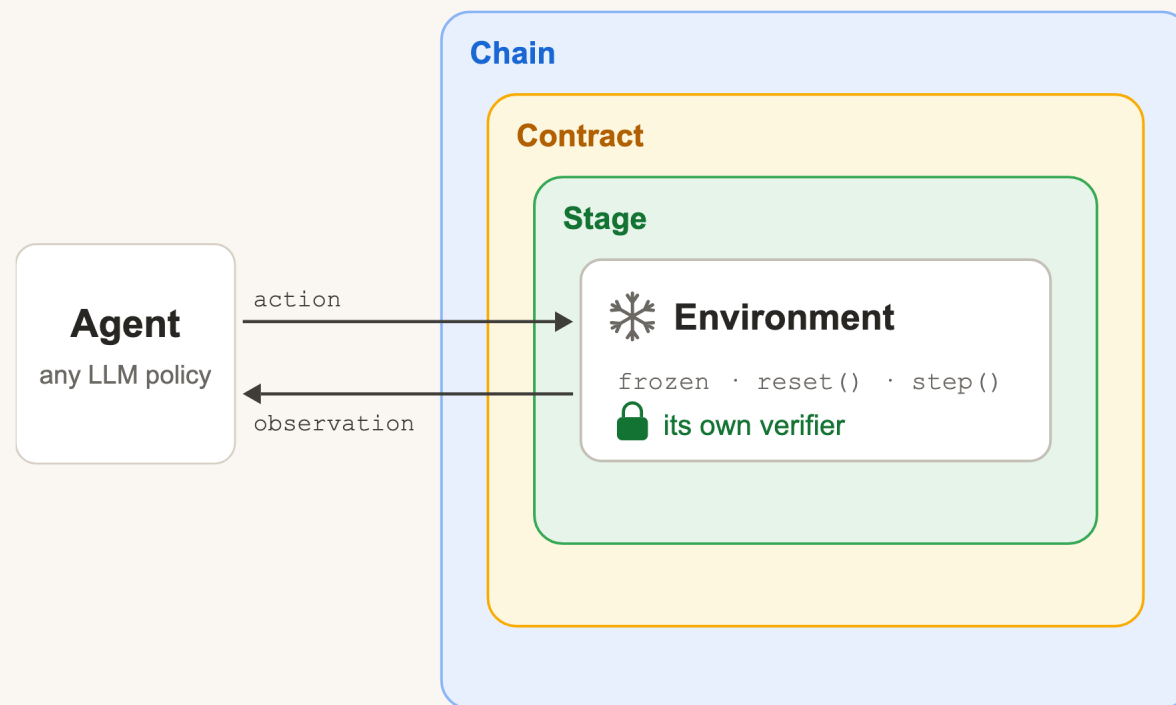
Curation can't fix a static world. Can we reshape the environment?

A harness, but for the environment



Both add power through outside layers, without changing the core. **Same trick, other side of the loop.**

Wrap the world, don't rebuild it



SWE-BENCH EXAMPLE • ONE BUG-FIX TASK

 **Stage** a different starting point

The issue arrives without its stack trace, so the agent must reproduce the bug first.

 **Contract** a new rule

“submit” is rejected until the agent has run the failing test.

 **Chain** a follow-up task

After this fix, solve a second issue in the same repo. Both test suites must pass.

EnvRigger: an agent that builds the practice world

THE AGENT

ENVRIGGER · same LLM as the agent



Too hard → make it easier

The agent never finds the bug in a large repo.

Stage: start with the relevant file already open.
→ it gets to practice the fix itself.

Too easy → make it harder

The agent solves the task every time.

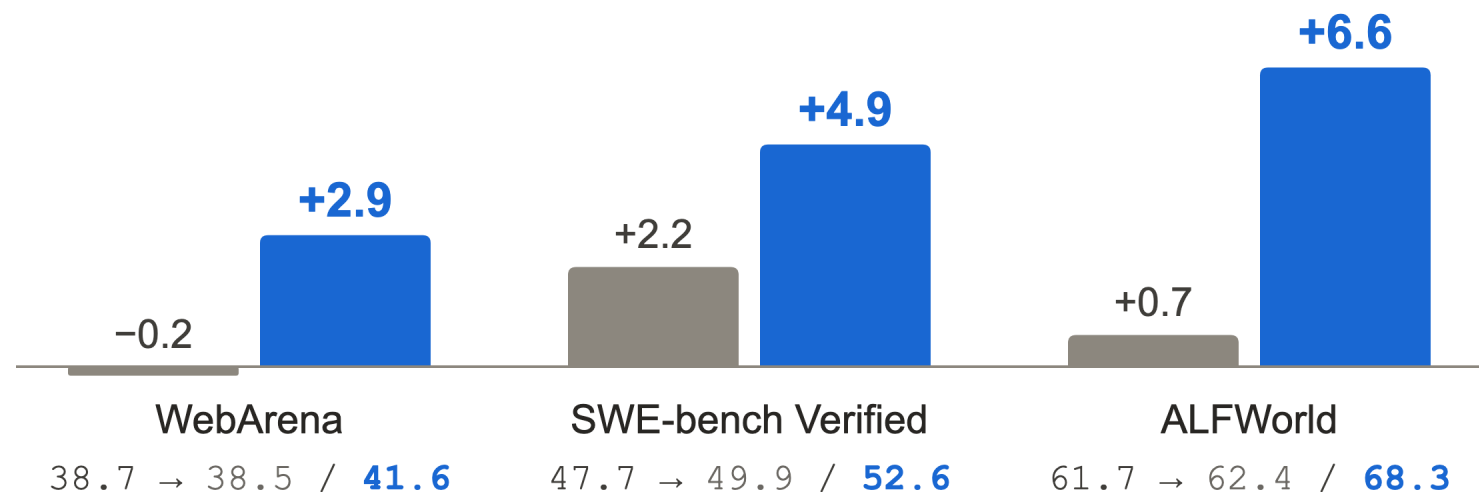
Stage: start from an issue with no stack trace.
→ it has to reproduce the bug first.

Same frozen worlds, better learners

GAIN OVER NO SKILLS (POINTS) · SUCCESS RATE % BELOW

■ Learned in original envs

■ Learned in EnvHarness envs



9.8%

fewer steps

SWE-bench Verified

ENVHARNESS

Community feedback

EnvHarness: Awakening Static Worlds for Agent Learning

▲ Upvote 160

Published on Aug 19 · ★ Submitted by  taesiri on Aug 21 [#1 Paper of the day](#)

#1

Paper of the Day on Hugging Face

Aug 21, 2026 · 160 upvotes

 envharness Public

👁 Watch 5

🍴 Fork 55

★ Star 641

EnvHarness: Awakening Static Worlds for Agent Learning

641

GitHub stars in 8 weeks

Open source · [google-research/envharness](https://github.com/google-research/envharness)

Also featured on the **Google Open Source Blog** and **36Kr**.

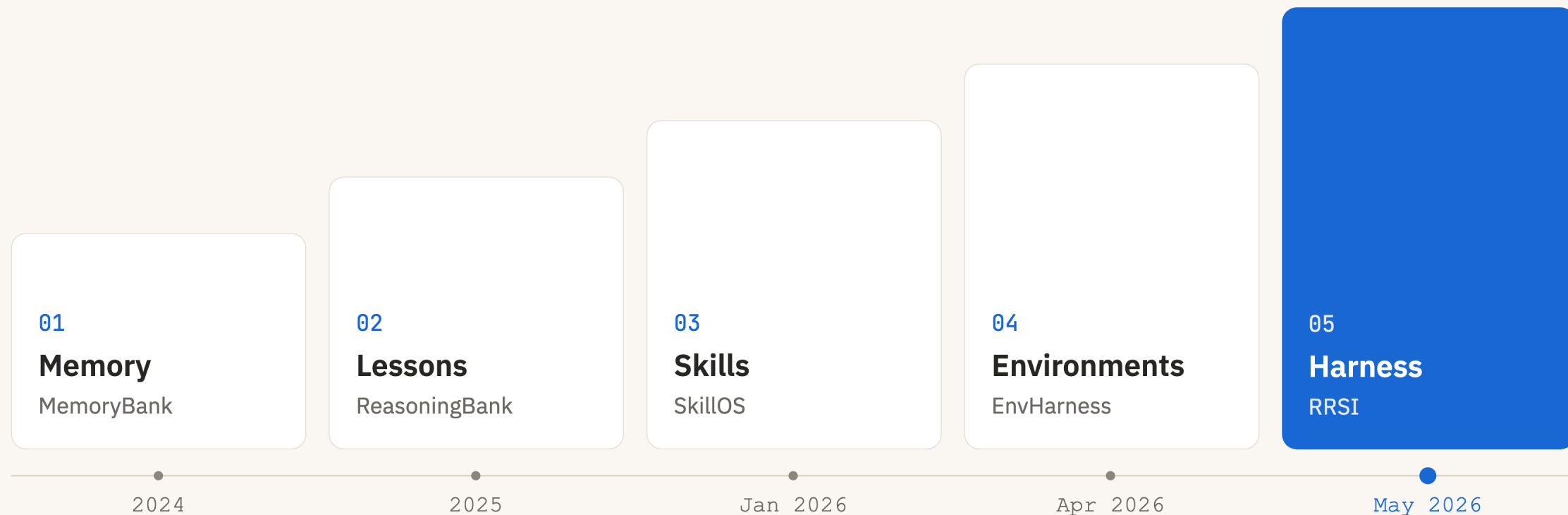
THE TURN

**If a harness can reshape the world,
the agent can start reshaping its own harness.**

Prompts, tools, memory, control flow: all of it is just code.

NEXT • ACT 5 • IMPROVE ITSELF SAFELY

Five places learning can live



If we can reshape the world, can the agent reshape its own harness, safely?

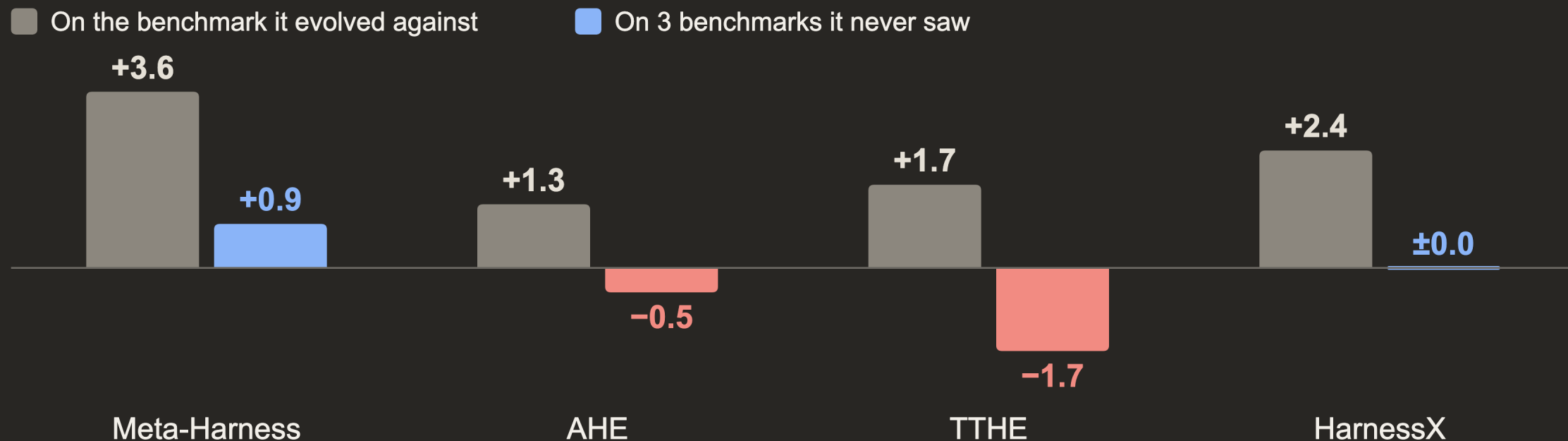
Recursive self-improvement, concretely



WHAT BREAKS

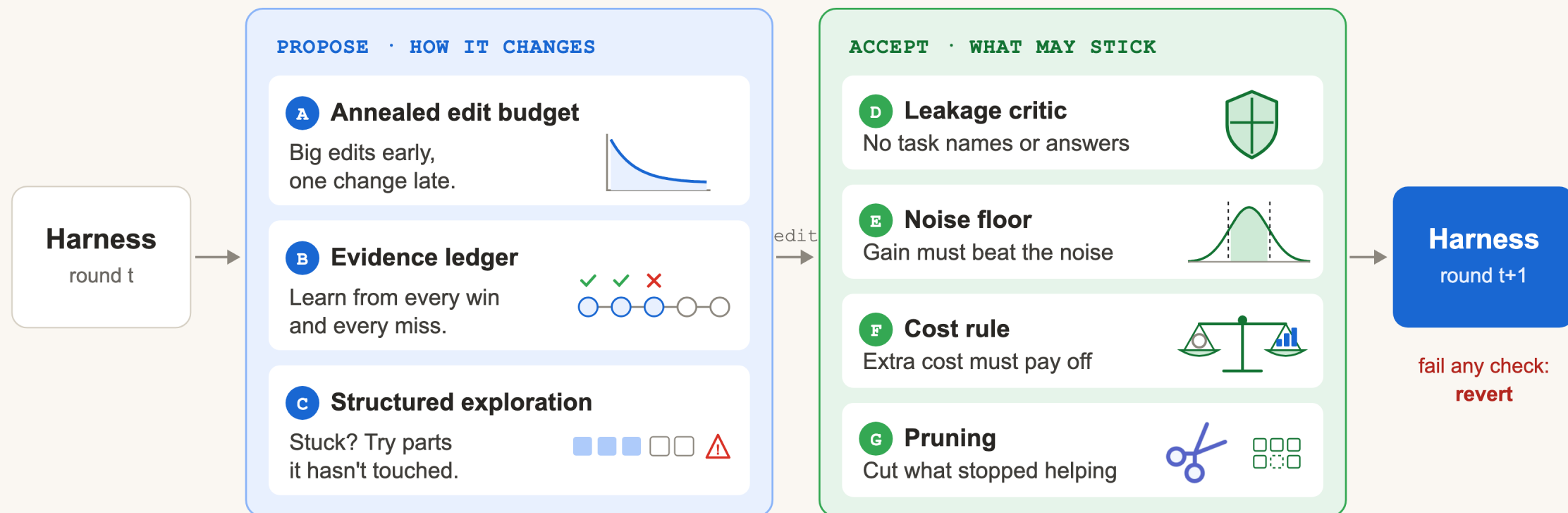
Self-improving harnesses learn the test

AGENTIC WORKSPACE · GAIN OVER THE STARTING HARNESS (POINTS)



Xia et al., “RRSI,” arXiv 2609.24972, 2026 (Table 1). Same tools, judge and policy as the starting harness.

Regularize the search, not the harness



Old regularizers, new place

IN TRAINING WEIGHTS

L0 sparsity

Train/test leakage checks

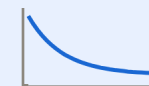
Beat the seed variance

L2 (ridge) penalty

L1 (lasso) shrinkage

IN RRSI, FOR AN AGENT EDITING ITS OWN CODE

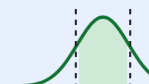
A Annealed edit budget
few changes per round, fewer over time



D Leakage critic
no task names or answers in the code



E Noise floor
a gain must clear run-to-run noise



F Cost rule
added complexity must pay for itself



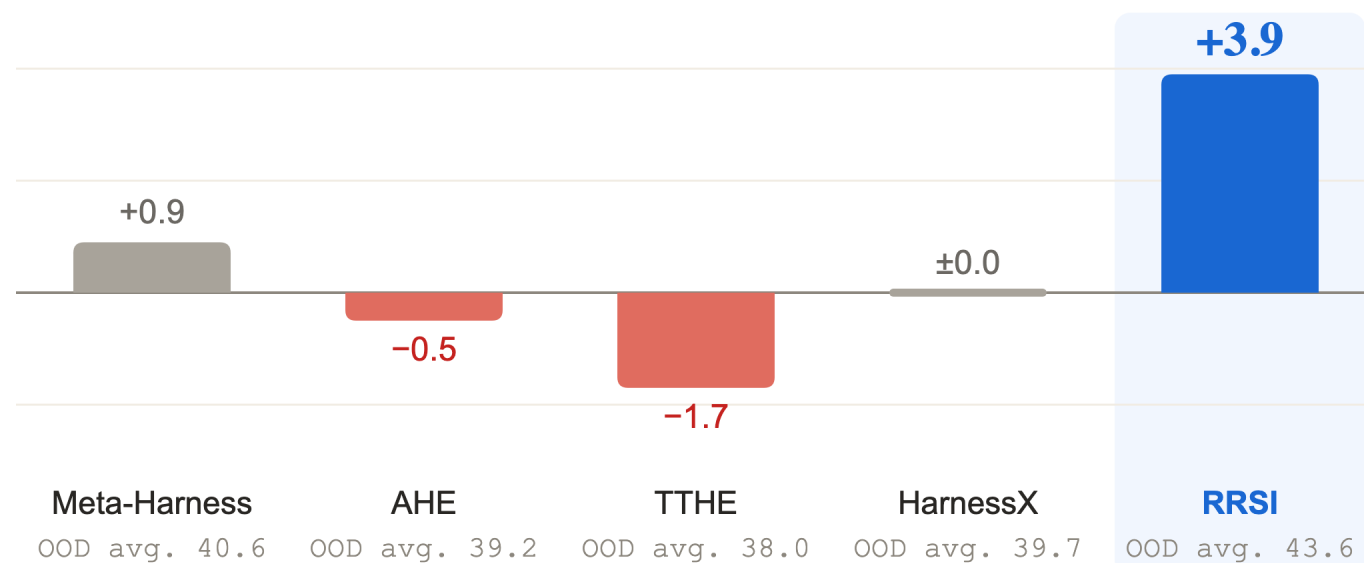
G Pruning
parts that stop helping go to zero



Gains that survive leaving the lab

3 never-seen workspace benchmarks

JobBench, GDPval, APEX-Agents · gain over the starting harness



All 6 held-out benchmarks

RRSI gain over the starting harness

+3.4 points on average
6 of 6 improve

JobBench		+4.7
Frontier-Eng		+4.3
APEX-Agents		+3.7
GDPval		+3.5
Harvey LAB		+2.3
SWE-bench Verified		+1.8

RRSI

Community feedback

 **rrsi** PublicWatch 3Fork 144Starred 1.5k

RRSI: Regularized Recursive Self-Improvement of Agent Harnesses

1.5k **GitHub stars in 3 weeks**
144 forks · open source · google-research/rrsi

RRSI: Regularized Recursive Self-Improvement of Agent Harnesses▲ Upvote 228

Published on Sep 20 · ★ Submitted by  Peng Xia on Sep 22 #2 Paper of the day **Hugging Face**

#2 **Paper of the Day on Hugging Face**
Sep 22, 2026 · 228 upvotes

Every fix revealed the next problem

01 • Memory

MemoryBank

ACL 2025

Organize what happened, and learn what to retrieve.

→ Remembering ≠ improving

02 • Lessons

ReasoningBank

ICLR 2026

Distill strategies from both wins and failures.

→ Memory only grows

03 • Skills

SkilloS

NeurIPS 2026

Train a small curator to keep what helps later.

→ The worlds are static

04 • Environments

EnvHarness

arXiv 2026

Wrap frozen worlds to target the agent's weaknesses.

→ The agent can rewrite itself

05 • Harness

RRSI

arXiv 2026

Regularize self-improvement so gains transfer.

→ Keep it real, bounded, readable

The models are good enough to be useful. The open question is whether they keep getting better after we ship them, and whether we stay in charge while they do.

A QUESTION FOR THIS CLASS

When should experience go back into the weights?

Context is cheap to edit. Weights generalize. Where's the line?

REFERENCES

The five papers behind this talk

MemoryBank

ACL 2025

In Prospect and Retrospect: Reflective Memory Management for Long-term Personalized Dialogue Agents

arxiv.org/abs/2503.08026

ReasoningBank

ICLR 2026

ReasoningBank: Scaling Agent Self-Evolving with Reasoning Memory

arxiv.org/abs/2509.25140

SkillOS

NeurIPS 2026

SkillOS: Learning Skill Curation for Self-Evolving Agents

arxiv.org/abs/2605.06614

EnvHarness

arXiv 2026

EnvHarness: Awakening Static Worlds for Agent Learning

arxiv.org/abs/2608.19880

RRSI

arXiv 2026

RRSI: Regularized Recursive Self-Improvement of Agent Harnesses

arxiv.org/abs/2609.24972

CREDITS

Thank you

This work was led by students and researchers, including

Zhen Tan · Siru Ouyang · Chengsong Huang · Peng Xia · Jun Yan · Zifeng Wang · Rujun Han · Yanfei Chen · I-Hung Hsu, and Google Cloud AI Research